Managing Anger: Enhancing AI-driven Cognitive Reappraisal Through Emotional Validation

Hanqiu Li[1]

Margaret Shih[1]

[1] Anderson School of Management, University of California, Los Angeles, United States

Author note:

Acknowledgements—We gratefully thank Bella Dong for assistance with literature review and manuscript format checking.

Anger is a common emotion that negatively impacts well-being, relationships, and workplace functioning. Although cognitive reappraisal is an effective strategy to regulate anger, deploying it at scale remains challenging. Artificial intelligence (AI) systems have shown promise in delivering reappraisal, yet their effectiveness is reduced when users are aware the support is AI-generated, a phenomenon known as the AI Label Effect. In this pre-registered study (N = 1,208), participants recalled a workplace event that made them angry and interacted with a perceived AI or human partner providing either standard reappraisal or emotionally validating reappraisal. Consistent with prior research, AI-labeled reappraisal was perceived as less effective than human-labeled reappraisal. Critically, embedding emotional validation into AI responses significantly improved perceived effectiveness, narrowing the gap between AI and human interactions. Mediation analyses revealed that emotional validation enhanced perceived empathy, support, expertise, and attachment, which in turn improved outcomes. These findings suggest that emotional validation is a key mechanism to reduce AI aversion in emotional support contexts, offering a scalable strategy to improve AI-mediated anger management.

Introduction

Anger is a prevalent emotion these days. A Pew Research Center Poll in 2017 found that 79% of Americans report feeling angry at least once a day, and 24% say they get really angry at least every day. This widespread emotional state manifests in various ways, from interpersonal tensions to public outbursts. Gallup's *State of the Global Workplace* report, approximately 18% of employees in the United States and Canada reported feeling "angry a lot of the day yesterday" (Gallup, 2023). Similarly, the American Psychological Association's *Work in America* Survey found that 19% of U.S. workers reported experiencing "irritability or anger with coworkers and customers" as a result of work-related stress (American Psychological Association, 2023). These statistics highlight that anger is not a fleeting feeling - it is a deeply rooted issue that affects mental well-being, social interactions, and workplace culture.

Experiencing frequent anger has been empirically linked to a range of both positive and negative interpersonal outcomes. There is research to show that anger can have potential positive effects on some interpersonal outcomes. One potential positive outcome resulting from anger is that it can motivate people to address perceived injustice (Lerner & Tiedens, 2006). Psychological studies have found that anger is a motivating emotion that can also serve as a catalyst for constructive social action. For instance, Van Doorn et al. (2014) found that participants who felt anger were more willing to confront and correct unfair situations. Moral anger highlighting injustice can signal strong leadership capability (Wang et al, 2018). In this way, anger, when expressed constructively and within appropriate social boundaries, can promote interpersonal accountability and drive meaningful improvements within social systems.

Conversely, anger can also hurt interpersonal relationships. Research has found that anger can reduce empathy and perspective taking. Individuals induced to feel anger were less likely to take the other person's point of view than individuals in a neutral mood (Yip & Schweitzer, 2019). Angry people are also more likely to engage in hostile attribution bias, wherein they interpret ambiguous social cues as intentionally harmful, which in turn, can escalate conflicts and interpersonal friction (Wilkowski & Robinson, 2008). These attribution styles may lead individuals to have negative opinions of their interaction partners and to see their interactions as less objective or empathetic. In addition, chronic expressions of anger have been found to be related to aggression, social rejection, and lowered trust between individuals (Finkel et al, 2002). Thus, over time, anger not only strains personal relationships, but can also erode social support networks, and exacerbate feelings of isolation and emotional distress.

In the workplace, heightened anger frequently manifests as workplace incivility or aggression, straining relationships among coworkers. For example, Kim et al. (2021) found that supervisors' uncivil behavior elicited anger in employees, which subsequently led those employees to engage in deviant behaviors that violated significant organizational norms, thereby threatening the well-being of the organization and its members. In a large-scale study involving 3,852 participants, Porat and Paluck (2024) demonstrated that employees who express anger are typically perceived as less competent and of lower status by their peers. This impression may negatively affect workplace relationships and professional advancement. Moreover, the impact of workplace anger may extend beyond the organizational setting. Fritz et al. (2019) found that employees who carried their anger home experienced disrupted sleep, including insomnia, which subsequently led to fatigue and reduced productivity at work the following day.

Considering the many impacts of anger, it is critical for researchers to develop effective tools to manage anger. One promising avenue is using AI in the application of emotion regulation strategies, particularly cognitive reappraisal, to help people better manage their anger.

Cognitive Reappraisal as a Strategy to Reduce Anger

Emotion regulation is defined as the activation of a goal to influence the emotion trajectory (Gross, 1998). One commonly used framework for studying emotion regulation strategies is the process model of emotion regulation (Gross, 1998), which outlines a sequence of stages in the emotion-generative process, from situation selection and modification to attention deployment, cognitive change, and response modulation. Within the domain of cognitive change, one particularly well-studied strategy is reappraisal (Gross, 2015), which refers to altering emotions by changing the way one thinks (McRae et al., 2012). Successful reappraisal influences many aspects of emotional responding, including self-reported negative affect (Gross, 1998), peripheral physiology (Jackson et al., 2000; Ray et al., 2010), and neural indicators of emotional arousal (Hajcak & Nieuwenhuis, 2006; Ochsner et al., 2004; Urry et al., 2006).

Given its efficacy in regulating emotions, cognitive reappraisal has been applied to help individuals manage negative feelings such as anger, frustration, and stress. For example, Berking et al. (2010) found that a structured training program, which included reappraisal, improved police officers' ability to accept and modify negative emotions. In a similar study, Buruck et al. (2016) showed that affect regulation training led to lasting improvements in well-being among elderly care workers. These effects were still present six months after the training. Abbott et al. (2009) conducted a randomized trial with sales managers and found that online resilience training with reappraisal components improved psychological health and work performance. Additionally, Truță (2013) found that training individuals in antecedent-focused strategies such as cognitive reappraisal led to reduced emotional effort, higher job satisfaction, and decreased turnover intentions. Many organizations now offer in-house workshops or online courses focused on reappraisal-based emotion regulation (Tan, 2018).

However, there are practical limitations and barriers to applying cognitive reappraisal in real-world settings. Most emotion regulation programs require trained facilitators, such as therapists, coaches, or workshop leaders, to guide employees through the learning process. This can be time-intensive and costly. Studies by Berking et al. (2010) and Buruck et al. (2016) relied on multi-session, in-person trainings with professional instructors, while Truță (2013) emphasized the need for structured workshops. As Yarker et al. (2022) noted, many workplaces lack the time, staff, or infrastructure to deliver such interventions.

Given these limitations, there is growing interest in developing emotion regulation tools that are scalable, personalized, and accessible on flexible schedules. One potential solution is to use artificial intelligence (AI) to deliver cognitive reappraisal strategies.

AI Conducting Cognitive Reappraisal is Driven by Reasoning Rather than Feeling

Cognitive reappraisal is driven by reasoning, rather than feelings of caring. Reappraisal skills can be learned, improved, and taught with relative ease and success (Cohen et al., 2018; Wang et al., 2021). Due to its simplicity, reappraisal is a good candidate for AI in mental health care (Li et al., 2024). Recent work has begun to explore the use of large language models (LLMs) to deliver cognitive reappraisal. Zhan et al. (2024) introduced the RESORT framework, which guides LLMs to generate targeted reappraisals based on appraisal theory. Sharma et al. (2023) deployed an LLM-based cognitive restructuring tool in a large field study, finding that it reduced emotional intensity for 68% of users and helped 66% overcome negative thoughts. Other studies have framed reappraisal as a style transfer task (Ziems et al., 2022) or developed models that control specific reframing attributes like specificity and empathy (Sharma et al., 2023; Reif et al., 2022).

There is even empirical evidence that advanced AI systems may match or exceed typical human performance in reappraisal. In a study conducted by Li et al. (2024), they trained GPT-4 to perform cognitive reappraisal of negative personal scenarios. Human participants were asked to generate reappraisals for the same scenarios. They used blind evaluation without telling the raters whether the generated reappraisals were from an AI or a human. The evaluation results showed that GPT-4's reframed interpretations were rated as more effective, more empathetic, and more novel on average, compared to human reappraisals. Even when people were given incentives to improve their reappraisal performance, the model still outperformed the majority of humans. These findings showed the capabilities for AI to effectively perform cognitive reappraisal and reduce users' negative emotions. However, there is a potential limitation to the effectiveness of using AI's ability to conduct cognitive reappraisal. When users become aware they are interacting with an AI system, the effectiveness of AI is reduced, which is often known as the AI Label Effect.

AI Label Effect on Emotional Support

Although AI systems can generate high-quality reappraisal responses, their effectiveness is reduced when users are aware that the emotional support was provided by an AI rather than a human. Yin et al. (2024) found that while AI-generated responses helped individuals feel heard, this effect was significantly reduced when responses were labeled as AI-authored. Similarly, Rubin et al. (2025) demonstrated across nine experiments that AI-labeled empathy, regardless of content quality, was consistently perceived as less supportive, caring, and emotionally resonant than the same message labeled as human-written. Furthermore, recent research suggests that the devaluation of AI interactions may extend beyond AI-generated content itself to affect perceptions of those who seek advice from AI. Dang and Liu (2024) found that individuals deny humanness to AI advice seekers, compared to human advice seekers, a phenomenon partly driven by perceived dissimilarity. These findings highlight a crucial paradox in the deployment of AI for emotional support: although AI can technically simulate empathy and support, humans' awareness of its artificial origin can substantially reduce the perceived emotional value of its support.

Building on these, **how can we effectively leverage AI's capabilities in cognitive reappraisal when the label effect persists?** In the present study, we address this challenge by embedding emotional validation into AI-delivered reappraisals. By explicitly acknowledging users' emotional experiences and simulating empathic understanding, we aim to increase the perceived emotional resonance of AI responses and diminish the AI label effect.

Emotional Validation as a Strategy to Diminish AI Label Effect in Reappraisal

Emotional validation is a structured interpersonal strategy that involves recognizing and affirming another person's emotional experience as valid, understandable, and acceptable. Linehan (1997) conceptualized validation as a multi-level process central to dialectical behavior therapy. According to her framework, effective validation unfolds across six increasing levels of depth: (1) active listening and full attention to the other's experience; (2) accurately reflecting back what the person is feeling; (3) identifying implied or unspoken emotions; (4) validating the emotion in light of the individual's context or learning history; (5) normalizing the emotion by expressing that it makes sense given the situation; and (6) expressing radical genuineness—responding as an equal human rather than as an authority figure.

Empirical evidence suggests that emotional validation enhances emotional insight and reduces emotional reactivity, which may support more effective emotion regulation. For example, Jeon and Park (2024) found that young children who received validating feedback showed greater persistence on a frustrating task than those who received invalidation. Lambie and Lindberg (2016) reported that maternal validation was positively associated with children's emotional awareness. In adults, Shenk and Fruzzetti (2011) demonstrated that participants exposed to validating responses during a stress task exhibited lower levels of negative affect, heart rate, and skin conductance compared to those receiving invalidating feedback.

Prior work has sought to reduce the AI label effect by making AI responses sound warmer or more stylistically empathic (e.g., through tone modulation, personalization, or fine-tuned large language models; see Ziems et al., 2022; Sharma et al., 2023). More recent studies, such as Rubin et al. (2025), have moved beyond surface features by prompting AI to express cognitive, affective, and motivational empathy grounded in psychological models of empathy. However, prior work has primarily focused on general empathic responsiveness rather than testing structured relational techniques specifically designed to enhance emotion regulation outcomes. Our study addresses this gap by introducing emotional validation as a theory-based mechanism to enhance emotional credibility and user trust, even when the AI label is fully transparent.

Present Study

The present study investigates how to enhance the perceived effectiveness of AI-delivered emotion regulation by integrating emotional validation into cognitive reappraisal, with a specific focus on workplace anger. Although LLMs can generate high-quality reappraisal responses, previous research has shown that users often rate these responses as less effective when they are aware that the source is AI, a phenomenon known as the AI label effect. We predict that embedding emotional validation in AI performing reappraisal to humans will help mitigate this AI label effect. In particular, we predict that prompting the AI to emotionally validate an individual's feelings could be effective in overcoming the AI labeling effect in helping to manage people's anger.

To test this, we conducted a pre-registered experimental study in which participants recalled a workplace event that elicited anger and then interacted with what they believed to be either a human or an AI partner trained in cognitive reappraisal. In two of the four conditions, the reappraisal message also included emotional validation, implemented through three steps: (1) identifying the user's expressed emotion, (2) explicitly acknowledging and affirming the legitimacy of that emotion, and (3) simulating shared emotional resonance to convey care and understanding.

Methods

Transparency and Openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study, and we follow JARS (Appelbaum et al., 2018). All data, analysis code, and research materials are available at (https://osf.io/d9q3u/?view_only=7febeac099e9439ba37372812910b4a8). Data were analyzed using R, version 4.0.0 (R Core Team, 2024) and the package *ggplot2*, version 3.5.1 (Wickham, 2016). The study's design and its analyses were pre-registered at (https://osf.io/69ed8/?view_only=04ec953c516941e0af4062723884d44d).

Participants

Data collection was approved by the [Masked for Blind Review] Institutional Review Board (IRB), and informed consent was obtained from all participants. We recruited 1,358 participants from the general adult population who are fully employed in the United States through CloudResearch (Connect), an online platform for survey recruitment. A total of 12 participants were excluded for experiencing technical issues and did not engage in the chatbot interaction, 51 participants were excluded for failing to pass the attention check question, and 87 were excluded for being assigned to the human condition but not believing that their discussion partner was human. The final sample for analysis consisted of 1,208 participants (50% female, 48.76% male, 1.08% non-binary; *M* age = 38.48 years, *SD* = 10.31). The majority of participants identified their race as White or Caucasian (69.12%), followed by Black or African American (11.01%), Asian or Asian American (7.53%), multi-racial (5.88%), Hispanic or Latinx (4.97%), American Indian or Alaska Native (0.66%), and other race (0.41%).

Before we start collecting data, we conducted power analyses for our key contrasts using pilot data ($N = 332$) to determine the required sample size for our study. Using the simr package in R, we created simulated datasets based on our pilot means and ANOVA mean square errors, and ran 200 simulations. Our analyses revealed that to reach 80.5% power, we would need approximately 335 participants per group, which is 1,340 participants in total. We ended up recruiting 1,358 participants.

Design

Participants were assigned to one of four experimental conditions: (1) **AI-labeled Neutral (AI-Neutral)**—participants were informed they were interacting with an AI expert, and the AI simply delivered reappraisal; (2) **AI-labeled Emotional Validation (AI-EV)**—participants were informed they were interacting with an AI expert, and the AI delivered reappraisal with emotional validation; (3) **Human-labeled Neutral (Human-Neutral)**—participants were informed they were interacting with a human expert, and the AI simply delivered reappraisal; and (4) **Human-labeled Emotional Validation (Human-EV)**—participants were informed they were interacting with a human expert, and the AI delivered reappraisal with emotional validation. Because we are not predicting an interaction effect, we designed the experiment as a four-condition study and would analyze the data with a one-way ANOVA.

Based on pilot data indicating high exclusion rates in human-labeled conditions due to disbelief in the partner's identity, we adjusted our recruitment strategy. A total of 1,197 participants were randomly assigned across all four conditions, and an additional 161 participants were randomly assigned specifically to the two human-labeled conditions to increase statistical power. The final analytic sample (N = 1,208) consisted of 283 participants in the AI-Neutral condition, 302 in the AI-EV condition, 316 in the Human-Neutral condition, and 307 in the Human-EV condition.

To assess the effectiveness of the emotional validation manipulation, participants answered a manipulation check item: “What emotional expression strategy do you think your discussion partner used?” with response options “Being neutral” or “Validating and mirroring my emotions.” Among participants in neutral conditions, 73.6% correctly identified the use of a neutral strategy. 61.7% of those in the emotional validation conditions correctly identified the emotional validation strategy.

Procedure

Participants were asked to recall a recent workplace incident that elicited anger. They were informed that they would be matched with an expert in emotion regulation to discuss the event. Participants were instructed to provide a detailed account of the incident, including their emotional responses, in a minimum of 80 words to ensure sufficient depth and details for the upcoming conversation. To maintain authenticity and minimize external influence, hard coding in Qualtrics was implemented to prevent participants from copying and pasting text from other platforms or AI-generated sources. Participants were also informed that their written description would be automatically shared with their discussion partner.

Following the recall task, participants completed a pre-conversation questionnaire. This included measures of current anger regarding the recalled event, and positive and negative affect. Afterward, participants were told they would begin a structured conversation with a discussion partner. The conversation consisted of five rounds of message exchanges. After five rounds, participants would then be given the option to continue in an open discussion phase, where they could send as many messages as they wished, without time restrictions. Each message was limited to one per turn, and participants would be required to write at least 10 words per message. Qualtrics logic was again employed to enforce these constraints and to prevent pasted responses.

Participants were subsequently directed to a “Matching you with an expert...” screen, which remained for seven seconds. Depending on condition assignment, participants then saw one of two messages: (1) “Congrats, you're now connected with an AI (Artificial Intelligence) chatbot, who is trained as an expert in the field of emotion regulation,” or (2) “Congrats, you're now connected with Emily Carter, who is an expert in the field of emotion regulation.” In the human-labeled condition, a fabricated LinkedIn profile was also shown, featuring an AI-generated profile picture and background information portraying Emily Carter as a consultant and researcher specializing in emotion regulation and behavioral science. We deliberately used the term "expert" in both the AI and human label conditions to ensure consistency in participants' initial impressions of objectivity across sources.

Participants clicked “Next” to begin the interaction. The participant’s recalled incident text was automatically sent as the first message in the chat. A dynamic “typing” indicator then appeared and disappeared intermittently to mimic natural typing behavior, similar to real-time texting interfaces. To enhance the realism of the interaction, chatbot responses were delayed randomly between 30 and 50 seconds. This delay was calibrated to reflect the time it might take a human to type a message of up to 300 characters, which is the maximum length we set up for each chatbot response. During the structured phase, participants were required to wait for the chatbot's response before sending their next message and were limited to one message per turn.

Upon completing the structured exchange, participants entered the open discussion phase if they chose to continue. This phase was unrestricted in message number and duration; participants could continue the conversation or end it at any point if they felt it had reached a natural conclusion.

Finally, participants completed an exit survey. This survey assessed the current anger, positive and negative affect, perceived emotion regulation success, and perceptions of the discussion partner (including empathy, support, attachment, objectivity, and expertise). The survey concluded with manipulation check items and demographic questions. Participants were then debriefed and informed that the "human" discussion partner was, in fact, an AI chatbot labeled as human.

AI Prompts

We used the GPT-4o API to generate responses in the conversations, with the AI behavior tailored according to the four experimental conditions. To preserve the illusion of human interaction in the human-labeled conditions, the chatbot was explicitly instructed not to reveal its identity as an AI.

The chatbot was prompted on the principles of emotion regulation and cognitive reappraisal. Specifically, the chatbot was prompted with the following instructions: *"Your task is* *to use the strategy of cognitive reappraisal to regulate the participant's emotion of anger. Specifically, help them rethink or reframe the situation they experienced. Cognitive reappraisal involves changing emotions by changing the interpretation of a situation (Gross, 2015). For instance, if a colleague makes you angry, consider that they might be dealing with personal issues that affect their behavior, and they may need support."* The chatbot was further instructed using the methodology outlined by Li et al. (2024), which was designed to train large language models (LLMs) in the strategy of reappraisal: *If a participant says, "My classmate sneers when I come into the room after I fell in the hallway last week. I worry they gossip about me," an effective reappraisal would be, "We don't know what others are thinking, and worrying only hurts you more. It could be a coincidence, or they could be sneering at something else. If they were really sneering at you, then they are unkind people, and you don't need to give them the time of day."*

To correctly prompt emotional validation, we applied the following prompt to the emotional validation conditions:

First, distinguish the emotion from each participant response. Then, in each of your responses, acknowledge their emotion and express that you validate and share it. For example, "When you mention feeling upset about this, I start to really see why it's so troubling, and I feel concerned too." "Hearing how that comment hurt you, I can't help but feel a bit hurt as well." "Your pride in your work is really showing, and it's making me feel proud of you too!" "Seeing you so disappointed is really bringing me down too. It's hard not to feel it with you." "It's okay to feel sad; talking about it might bring up some similar feelings in me." "It's normal to feel unsure about big decisions, and hearing you talk about it makes me think of times I've felt the same way." "I'm here for you, and sharing how you feel is really making me care even more." Ensure *that your response directly reflects the intensity of the participant's feelings. Use exclamation marks and interjections to emphasize that you share the participant's feelings.*

To further humanize the chatbot, it was prompted to use casual language and short sentences, with each sentence containing a maximum of 20 words. The chatbot was instructed to avoid complex vocabulary and never use bullet points in its responses. To simulate human-like uncertainty, the chatbot employed phrases such as "I'm not sure about that," or "Let me think about that." Additionally, the chatbot was directed to use contractions, active voice sentences, and first-person pronouns like "I" or "we" to personalize the interaction. To mimic natural human variability, the chatbot varied its sentence lengths and included filler words and phrases such as "well," "you know," "um," and "let's see." The use of abbreviations, such as "IDK," "PLS," and "OK," was also encouraged to further enhance the casual tone.

Moreover, the chatbot was programmed to redirect participants back to the relevant discussion topic if they deviated from the main conversation. To enhance the naturalness of the interaction across all conditions, the chatbot was instructed to include one typo in every five responses.

Measures

Anger Level

State anger was assessed using a single-item measure on a scale from 1 (*not at all*) to 7 (*extremely*): "Think about your recalled incident, to what extent do you feel angry?" Participants indicated their current level of anger before and immediately after the interaction.

Positive and Negative Affect

Emotional levels were assessed before and after the interaction using the 10-item Positive and Negative Affect Schedule Short Form (PANAS-SF; Watson et al., 1988) on a scale from 1 (*not at all*) to 7 (*extremely*). Positive emotions included items such as "inspired," "determined," and "active," while negative emotions included "upset," "nervous," and "afraid." Cronbach's alpha values were as follows: pre-positive affect (α = .80), pre-negative affect (α = .80), post-positive affect (α = .85), and post-negative affect (α = .81).

Perceived Emotional Regulation Success

Perceived emotional regulation success was measured using a single-item measure on a scale from 1 (*not at all successful*) to 7 (*extremely successful*): "How successful do you think the interaction was in helping you regulate your anger?"

Perceived support

Perceived emotional support during the interaction was measured with a single-item on a scale from 1 (*not at all*) to 7 (*extremely*): "To what extent do you feel emotionally supported during the interaction?"

Wish to Continue

The extent to which participants wished to continue interacting with the discussion partner was measured on a single item on a scale from 1 (*not at all*) to 7 (*extremely*): "To what extent do you wish to continue interacting with your discussion partner?"

Perceived Objectivity

Perceived objectivity of the discussion partner was assessed using a two-item scale from 1 (*strongly disagree*) to 7 (*strongly agree*), where participants indicated agreement with statements of "I believe the information provided by my discussion partner was presented in a balanced manner." and "My discussion partner seemed to be neutral and did not favor any particular viewpoint." Cronbach's alpha for this scale was .77.

Perceived Expertise

The perceived expertise of the interaction partner was measured using five items from the Source Credibility Scale (Ohanian, 1990) on a scale from 1 (*not at all*) to 7 (*extremely*). Participants rated their partner on dimensions of “Expert”, “Experienced”, “Knowledgeable”, “Qualified”, and “Skilled”. Cronbach’s alpha was .98.

Perceived Empathy

Perceived empathy of the discussion partner was assessed using two items adapted from the Basic Empathy Scale (Jolliffe & Farrington, 2006). Participants rated their agreement with statements such as "During the interaction, my discussion partner really feels for me" and "During the interaction, my discussion partner understands me well." Participants rated agreement on a 7-point scale (*1 = strongly disagree, 7 = strongly agree*). Cronbach’s alpha was .94.

Attachment

Attachment to the discussion partner was measured using items adapted from Martin & Mason (2023) on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*). Participants rated their agreement with statements of "I feel attached to this partner," "I would be sad if I no longer interacted with this partner," and "I feel connected to this partner." Cronbach’s alpha was .93.

Results

Primary Analysis

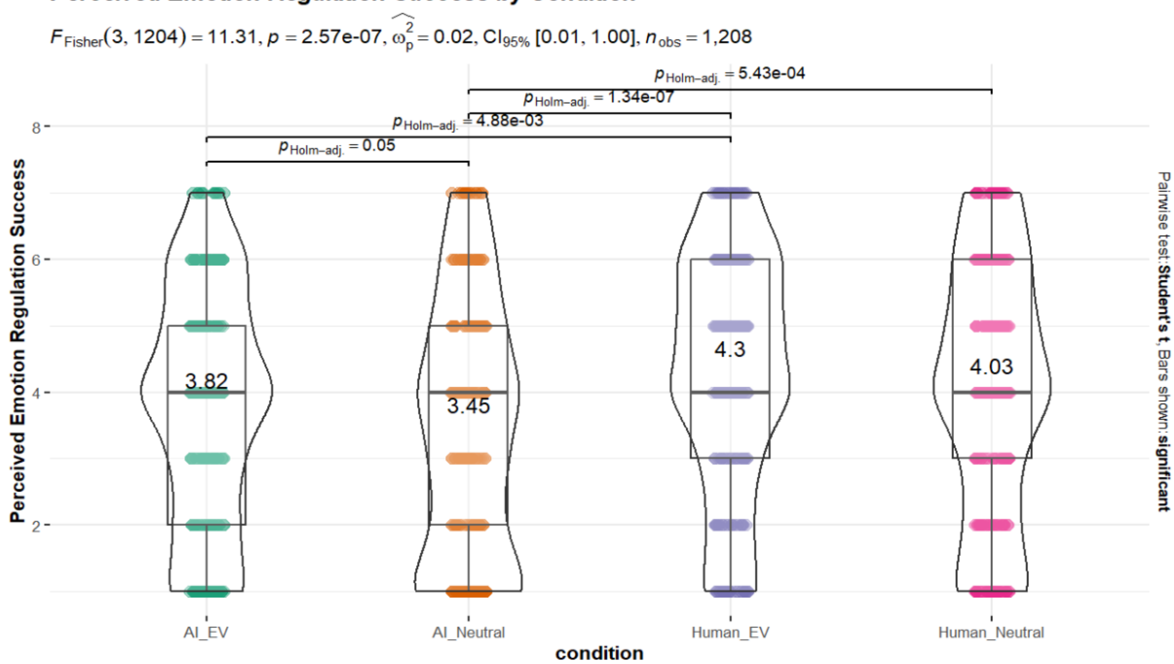
We first examined descriptive statistics of the main outcome variables. Overall, participants' anger levels decreased from pre-interaction (M = 5.31, SD = 1.33) to post-interaction (M = 4.10, SD = 1.70), indicating a reduction in anger after engaging in the reappraisal conversation ($t(1207) = 28.65$, $p < .001$, 95% CI [1.13, 1.29]). For positive affect, there was a slight increase from pre-interaction (M = 4.25, SD = 1.25) to post-interaction (M =

4.35, *SD* = 1.39; $t(1207) = -4.41$, $p < .001$, 95% CI [-0.14, -0.05]). Negative affect showed a decrease from pre-interaction (*M* = 2.60, *SD* = 1.21) to post-interaction (*M* = 2.15, *SD* = 1.13; $t(1207) = 19.25$, $p < .001$, 95% CI [0.40, 0.49]), further supporting the effectiveness of the interaction in reducing negative emotions. Participants rated the perceived success of emotion regulation at a moderate level (*M* = 3.91, *SD* = 1.86) on a 7-point scale.

Perceived Emotion Regulation Success

Omnibus ANOVA. In order to compare the four conditions on the main dependent variable, perceived emotion regulation success, we first conducted a one-way analysis of variance (ANOVA) with condition (AI-Neutral, AI-Emotional Validation, Human-Neutral, and Human-Emotional Validation) as the independent variable and perceived emotion regulation success as the dependent variable (Fig. 1). The analysis revealed a statistically significant effect of condition, $F(3, 1204) = 11.3$, $p < .001$, $\eta^2 = .027$.

Figure 1
Perceived Emotion Regulation Success by Condition



Note. Higher values indicate greater perceived effectiveness. Error bars represent 95% confidence intervals.

As the omnibus ANOVA was significant, we proceeded to conduct focused comparisons to test our specific predictions: (1) whether the label effect exists (AI vs. Human-labeled Neutral conditions), and (2) whether AI with emotional validation performs better than AI without emotional validation.

Focused Contrasts. We followed up with focused contrasts for perceived emotion regulation success.

AI-labeled Neutral vs. Human-labeled Neutral (The Label Effect). To confirm whether the AI label effect exists in cognitive reappraisal, we conducted a focused contrast using the mean square error (MSE = 3.36) from the omnibus ANOVA to compare the AI-Neutral condition (M = 3.44, SD = 1.90) with the Human-Neutral condition (M = 4.03, SD = 1.90; Fig. 2). This focused comparison revealed that participants in the Human-Neutral condition reported significantly higher perceived emotion regulation success than those in the AI-Neutral condition, $t(1204) = 3.88$, $p < .001$, mean difference = 0.58, 95% CI [0.29, 0.88]. This finding confirms the existence of the label effect in cognitive reappraisal, showing that participants who believed they were interacting with an AI perceived the intervention as less effective in managing their workplace anger compared to those who believed they were interacting with a human expert.

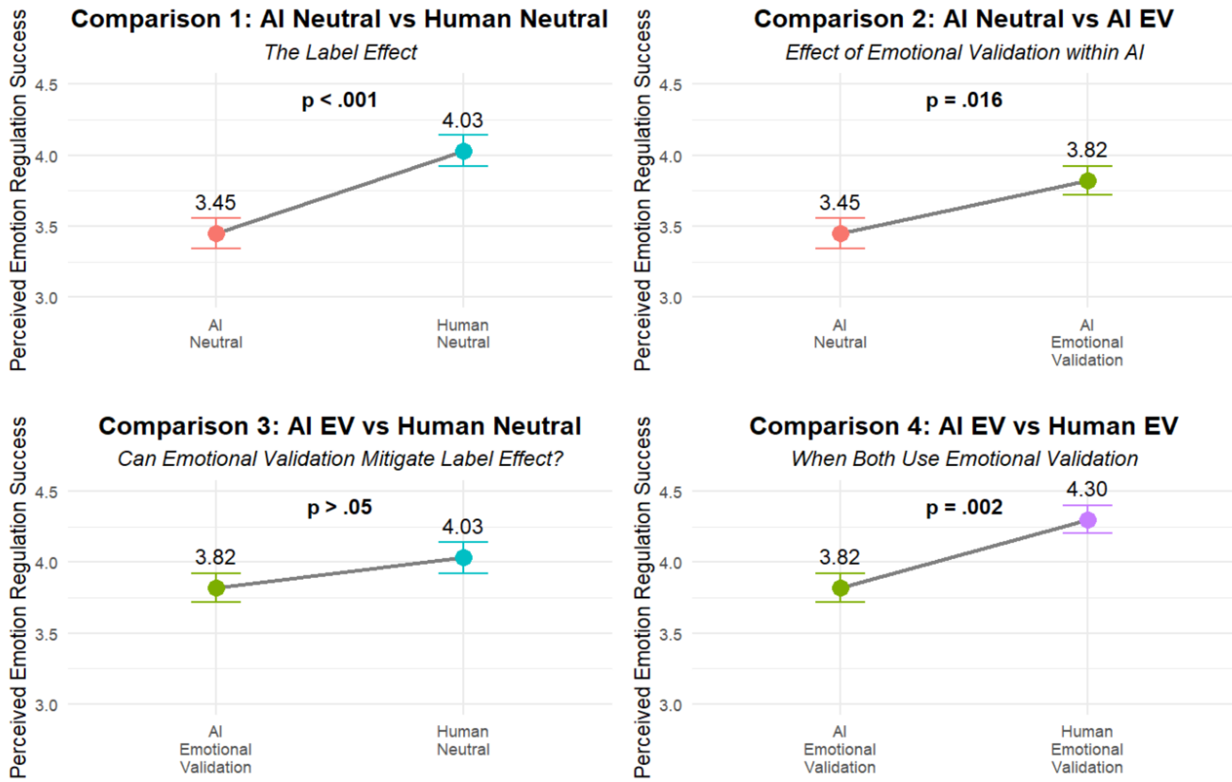
AI-labeled Emotional Validation vs. AI-labeled Neutral. After confirming the existence of the label effect in cognitive reappraisal, we tested whether emotional validation could mitigate this effect. In particular, we compared the two AI-labeled conditions to see if AI with emotional validation would receive better perceived emotion regulation success compared to the AI without emotional validation (Fig. 2). Using the mean square error from our previous omnibus ANOVA, we conducted a focused contrast comparing the AI-Emotional Validation condition (M = 3.82, SD = 1.76) with the AI-Neutral condition (M = 3.45, SD = 1.90). This analysis revealed a significant difference between these conditions, $t(1204) = 2.43$, $p = .015$, 95% CI [0.07, 0.67], with participants in the AI-Emotional Validation condition reporting higher perceived emotion regulation success than those in the AI-Neutral condition. This finding suggests that embedding emotional validation into AI-delivered cognitive reappraisal significantly enhances its perceived effectiveness in managing workplace anger.

AI-labeled Emotional Validation vs. Human-labeled Groups. To further analyze the effect of emotional validation on the AI label effect, we conducted additional analyses examining whether AI with emotional validation received a different level of perceived effectiveness than the human-labeled conditions. This required us to conduct two comparisons.

First, we compared the AI-labeled Emotional Validation condition ($M = 3.82$, $SD = 1.76$) with the Human-labeled Neutral condition ($M = 4.03$, $SD = 1.90$). This analysis did not reveal a statistically significant difference between these conditions ($p > .05$; Fig. 2). Second, we compared the AI-labeled Emotional Validation condition with the Human-labeled Emotional Validation condition ($M = 4.30$, $SD = 1.78$). This analysis revealed a significant difference, $t(1204) = -3.24$, $p = .001$, mean difference = -0.48, 95% CI [-0.77, -0.19] (Fig. 2), indicating that when both used emotional validation, the human-labeled condition was perceived as more effective than the AI-labeled condition. This finding suggests that while emotional validation improves AI performance, a gap in perceived effectiveness remains between AI and human interventions when both implement emotional validation strategies.

Figure 2

Comparisons of Perceived Emotion Regulation Success Across Conditions



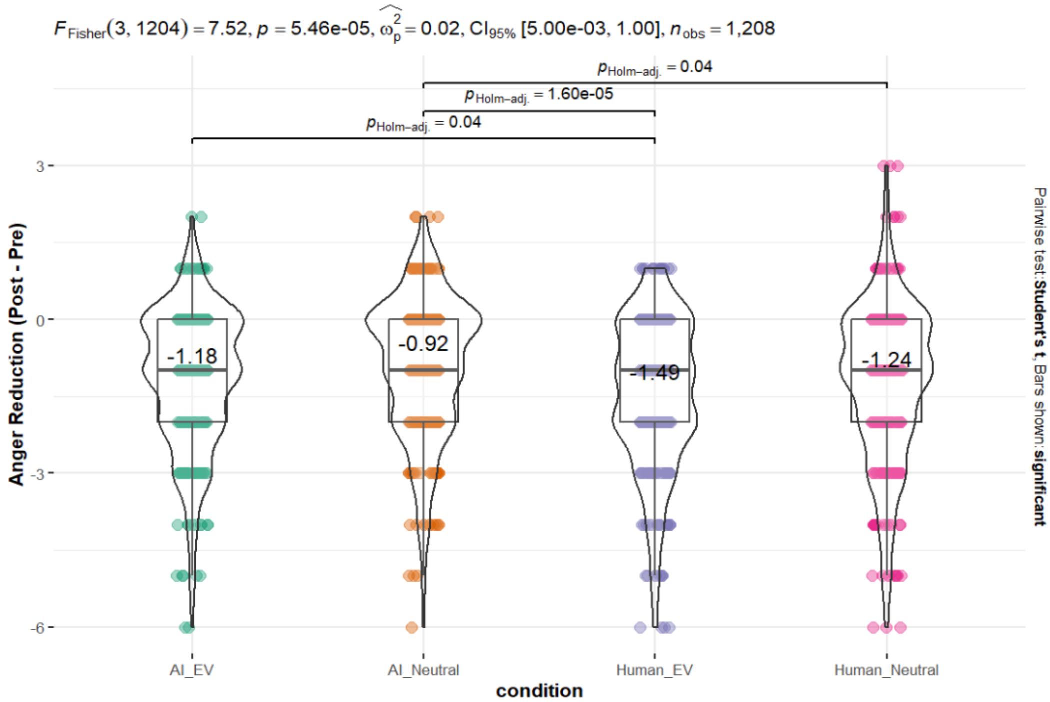
We then conducted similar analyses on emotional outcome variables, including reduction in anger and positive and negative affect. This is to see if the affective outcomes have similar patterns to the perceived emotion regulation success.

Anger Reduction

Omnibus ANOVA. Similar to the perceived emotion regulation success pattern, we conducted a one-way ANOVA with condition (AI-Neutral, AI-Emotional Validation, Human-Neutral, and Human-Emotional Validation) as the independent variable and reduction in anger (calculated by post-study anger minus pre-study anger) as the dependent variable. The analysis revealed a statistically significant effect of condition, $F(3, 1204) = 7.52$, $p < .001$, $\eta^2 = .018$ (Fig. 3). Again, since the omnibus ANOVA was significant, we proceeded to conduct focused comparisons.

Figure 3

Anger Reduction by Condition



Note. Higher negative values indicate greater anger reduction. Error bars represent 95% confidence intervals.

Focused Contrasts. We followed up with focused contrasts for anger reduction.

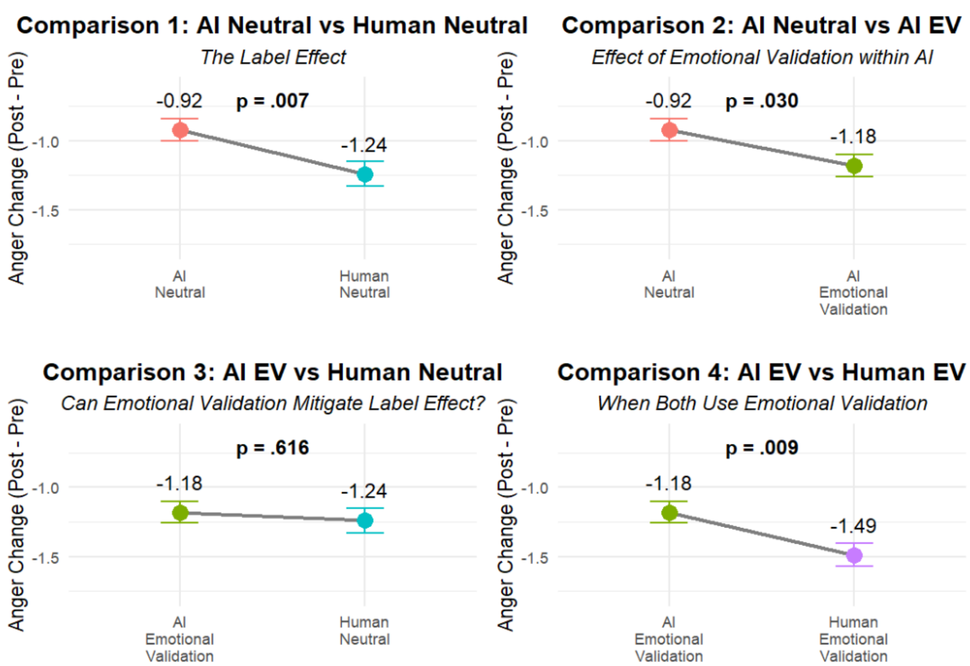
AI-labeled Neutral vs. Human-labeled Neutral (The Label Effect). Using the mean square error (MSE = 2.124) from the ANOVA, we conducted additional analyses examining the focused contrast, which revealed that participants in the Human-Neutral condition (M = -1.24, SD = 1.57) experienced significantly greater anger reduction than those in the AI-Neutral condition (M = -0.92, SD = 1.39), $t(1204) = -2.70$, $p = .007$, 95% CI [-0.56, -0.09] (Fig. 4). This finding confirms the existence of the label effect in anger reduction.

AI-labeled Emotional Validation vs. AI-labeled Neutral. The focused contrast indicated that participants in the AI-Emotional Validation condition (M = -1.18, SD = 1.37) experienced significantly greater anger reduction compared to those in the AI-Neutral condition (M = -0.92, SD = 1.39), $t(1204) = -2.13$, $p = .033$, 95% CI [-0.49, -0.02] (Fig. 4). This finding suggests that embedding emotional validation into AI-delivered cognitive reappraisal significantly enhances its effectiveness in reducing workplace anger.

AI-labeled Emotional Validation vs. Human-labeled Groups. We conducted additional analyses to examine the effect of emotional validation on the AI label effect for anger reduction, by comparing the AI-labeled Emotional Validation condition ($M = -1.18$, $SD = 1.37$) with the Human-labeled Neutral condition ($M = -1.24$, $SD = 1.57$). This analysis did not reveal a statistically significant difference between these conditions, $t(1204) = 0.55$, $p = .579$ (Fig. 4). Second, we compared the AI-labeled Emotional Validation condition with the Human-labeled Emotional Validation condition ($M = -1.49$, $SD = 1.48$). This analysis revealed a significant difference, $t(1204) = 2.62$, $p = .009$, 95% CI [0.08, 0.54] (Fig. 4), indicating that when both used emotional validation, the human-labeled condition was more effective in reducing anger than the AI-labeled condition.

Figure 4

Comparisons of Anger Reduction Across Conditions

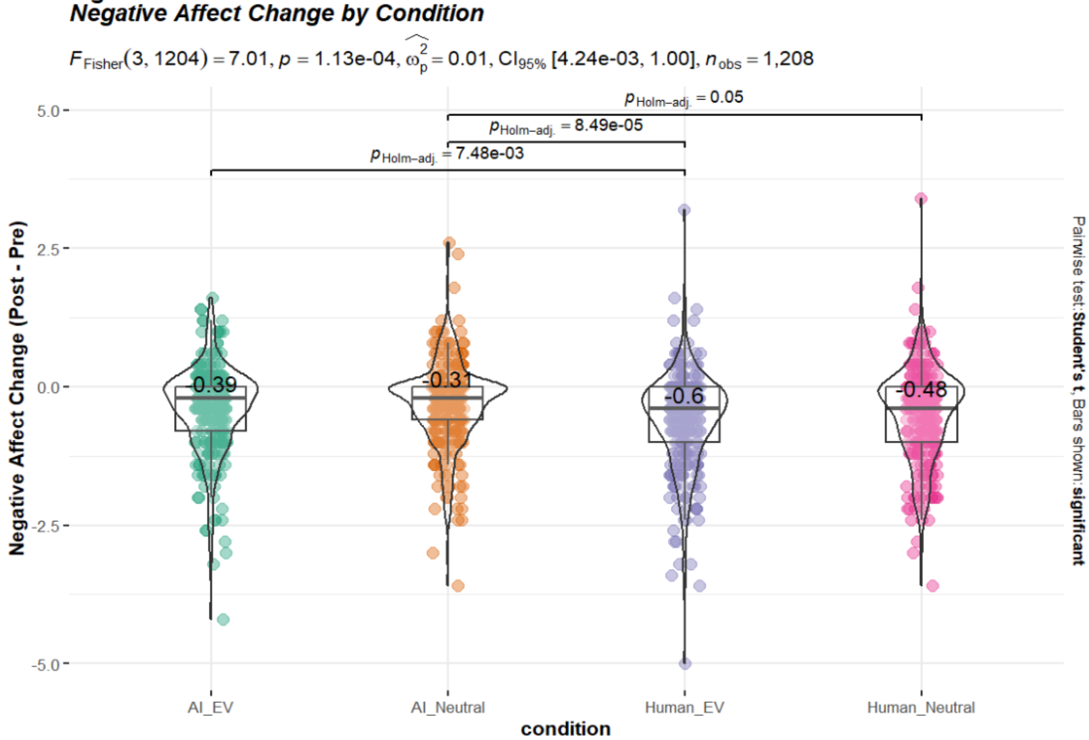


Note. Higher negative values indicate greater anger reduction.

Positive and Negative Affect

Omnibus ANOVA. A one-way ANOVA showed no significant differences across conditions for positive affect change ($F(3, 1204) = 1.03$, $p = .381$), so we did not conduct focus group contrasts for positive affect change (calculated by post-interaction positive affect minus pre-interaction positive affect). However, a one-way ANOVA revealed a significant effect of condition on negative affect change (calculated by post-interaction negative affect minus pre-interaction negative affect), $F(3, 1204) = 7.01$, $p < .001$, $\eta^2 = .017$ (Fig. 5). We then proceeded to conduct focused comparisons for negative affect change.

Figure 5
Negative Affect Change by Condition



Note. Higher negative values indicate greater negative affect change. Error bars represent 95% confidence intervals.

Focused Contrasts. We followed up with focused contrasts for positive and negative affect.

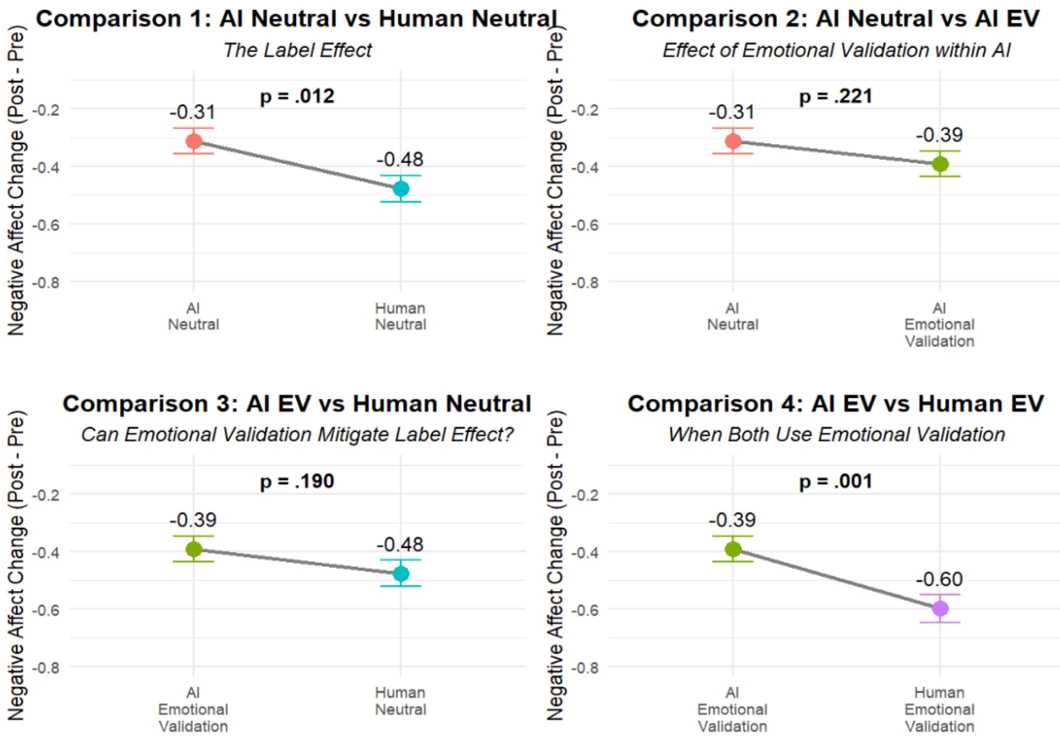
AI-labeled Neutral vs. Human-labeled Neutral (The Label Effect). Using the mean square error (MSE = 0.642) from the ANOVA, we conducted additional analyses examining the focused contrast, which revealed that participants in the Human-Neutral condition ($M = -0.48$, $SD = 0.82$) experienced significantly greater change in negative affect than those in the AI-Neutral condition ($M = -0.31$, $SD = 0.76$), $t(1204) = -2.53$, $p = .012$ (Fig. 6). Again, this finding confirms the existence of the label effect in negative affect reduction.

AI-labeled Emotional Validation vs. AI-labeled Neutral. A focused contrast indicated that the difference between the AI-Emotional Validation condition ($M = -0.39$, $SD = 0.77$) and the AI-Neutral condition ($M = -0.31$, $SD = 0.76$) was not statistically significant, $t(1204) = -1.22$, $p = .221$ (Fig. 6). Emotional validation had less impact on negative affect change within the AI-labeled conditions than it did for other outcome measures such as anger reduction and perceived emotion regulation success.

AI-labeled Emotional Validation vs. Human-labeled Groups. As additional analyses, we first compared the AI-labeled Emotional Validation condition with the Human-labeled Neutral condition. This analysis did not reveal a statistically significant difference ($p > 0.5$). Second, we compared the AI-labeled Emotional Validation condition with the Human-labeled Emotional Validation condition. This analysis revealed a significant difference (Fig. 6), $t(1204) = 3.18$, $p = .001$, mean difference = 0.21, 95% CI [0.08, 0.33], indicating that when both used emotional validation, the human-labeled condition was more effective in reducing negative affect than the AI-labeled condition.

Figure 6

Comparisons of Negative Affect Change Across Conditions



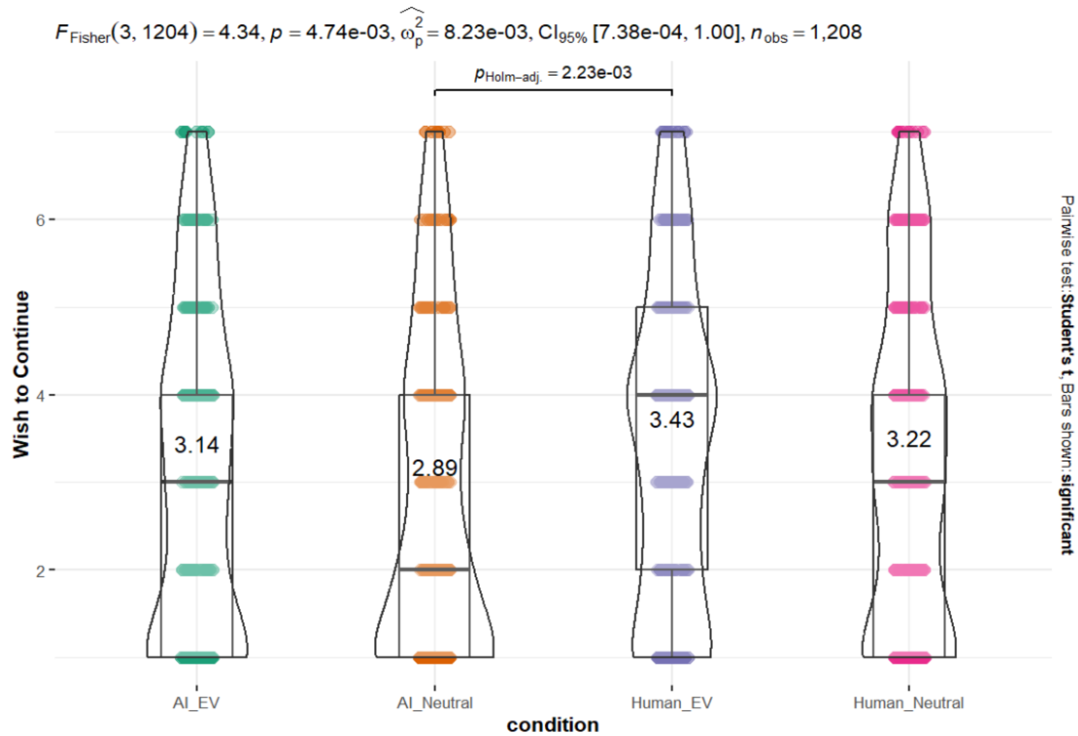
Note. Higher negative values indicate greater negative affect reduction.

Willingness To Continue Interaction

Omnibus ANOVA. We also asked participants to report to what extent they would like to continue interacting with the discussion parter, after they finished the conversation. We conducted additional analyses examining a one-way ANOVA with condition as the independent variable and wish to continue as the dependent variable revealed a statistically significant effect of condition, $F(3, 1204) = 4.34$, $p = .005$, $\eta^2 = .011$ (Fig. 7). As the ANOVA was significant, we proceeded to conduct focused comparisons.

Figure 7

Wish To Continue Interacting With the Discussion Parter, by Condition



Note. Higher values indicate a greater willingness. Error bars represent 95% confidence intervals.

Focused Contrasts. We followed up with focused contrasts for willingness to continue.

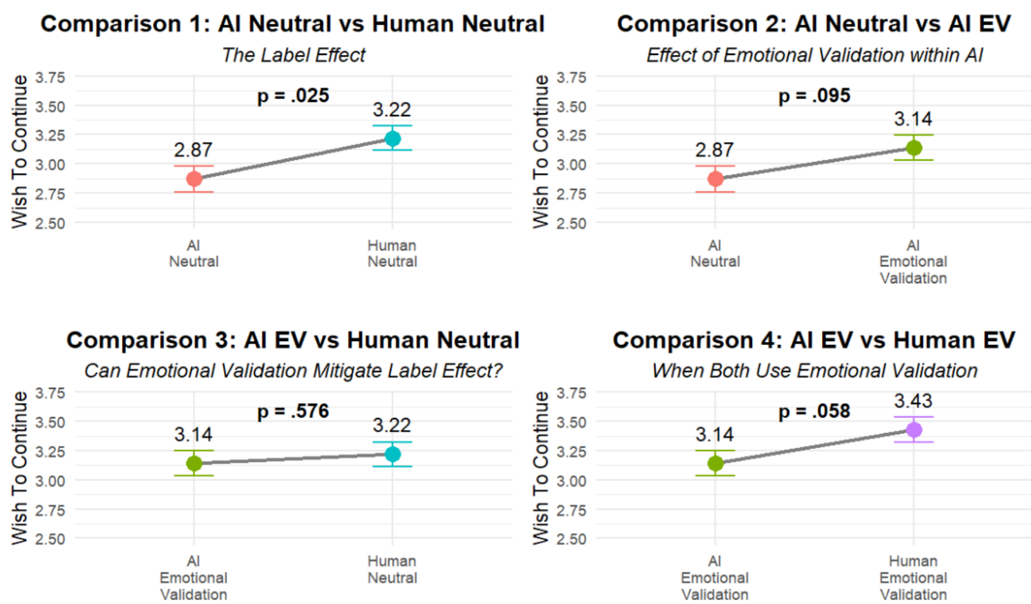
AI-labeled Neutral vs. Human-labeled Neutral (The Label Effect). As an additional analysis, we found a similar pattern to all other dependent variables. Participants in the Human-Neutral condition (M = 3.22, SD = 1.85) expressed significantly stronger desire to continue the interaction than those in the AI-Neutral condition (M = 2.89, SD = 1.85), $t(1204) = 2.19$, $p = .028$ (Fig. 8). This finding again confirms the existence of AI label effect.

AI-labeled Emotional Validation vs. AI-labeled Neutral. The additional analysis focused contrast indicated that the difference between the AI-Emotional Validation condition (M = 3.14, SD = 1.85) and the AI-Neutral condition (M = 2.89, SD = 1.85) did not reach statistical significance, $t(1204) = 1.67$, $p = .095$ (Fig. 8). Interacting with AI with emotional validation did not significantly increase people's willingness to continue interacting with the discussion partner, compared to the AI without emotional validation.

AI-labeled Emotional Validation vs. Human-labeled Groups. As an additional analysis, we first compared the AI-labeled Emotional Validation condition with the Human-labeled Neutral condition, and again, did not find significant difference ($p > 0.05$; Fig. 8). Second, we compared the AI-labeled Emotional Validation condition with the Human-labeled Emotional Validation condition. This analysis approached statistical significance, $t(1204) = -1.92$, $p = .055$, mean difference = -0.29, 95% CI [-0.58, 0.01] (Fig. 8).

Figure 8

Comparisons of Willingness to Continue Interaction Across Conditions



Mediation analyses

To investigate the psychological mechanisms underlying our findings, we conducted several additional analyses, which examined the mediation effect to understand: 1) why human-labeled reappraisal was perceived as more effective than AI-labeled reappraisal (the label effect), and 2) why AI-labeled reappraisal with emotional validation outperformed AI-labeled simple reappraisal. Because anger impacts interpersonal perceptions and attributions, we predicted that factors associated with perceptions of interaction partners might mediate the effectiveness of the emotional validation manipulation. The five mediators we tested include attachment to the discussion partner, perceived support from the discussion partner, perceived empathy of the discussion partner, perceived expertise of the discussion partner, and perceived objectivity of the discussion partner. We included the perceived emotion regulation success and anger reduction as the two outcome variables separately for the mediation models. We conducted separate bootstrap mediation analyses (5,000 resamples) for each comparison using the mediation package in R (Tingley et al., 2014), examining the average causal mediation effect (ACME), average direct effect (ADE), and proportion of the total effect mediated.

Perceived Emotion Regulation Success as the Outcome Variable

Mediation Effect On AI-labeled vs. Human-labeled Neutral Conditions. We first tested the five mediators of the label effect, using perceived emotion regulation success as the outcome variable.

Perceived Partner Support. Perceived support from the discussion partner fully mediated the relationship between the partner label condition (AI-labeled Neutral versus Human-labeled Neutral) and perceived emotion regulation success ($ACME$ = 0.59, 95% CI [0.34, 0.83], $p < .001$; ADE = -0.00, 95% CI [-0.19, 0.18], $p = .988$; proportion mediated = 100.5%, 95% CI [0.75, 1.56], $p < .001$; Fig. 9). This complete mediation indicates that the indirect effect through partner support accounts for the entire relationship between the partner label and perceived effectiveness. Participants rated the human-labeled partner as more effective than the AI-labeled partner because they perceived the human as more supportive.

Perceived Partner Objectivity. Perceived objectivity of the discussion partner partially mediated the relationship between the partner label condition (AI-labeled Neutral versus Human-labeled Neutral) and perceived emotion regulation success (*ACME* = 0.23, 95% CI [0.06, 0.41], $p < .01$; *ADE* = 0.35, 95% CI [0.09, 0.62], $p < .01$; proportion mediated = 39.8%, 95% CI [0.14, 0.75], $p < .01$; Fig. 9). This partial mediation shows that both direct and indirect pathways significantly contribute to the outcome. In practical terms, participants perceived the human-labeled partner as more objective than the AI-labeled partner, and this difference in perceived objectivity explained approximately 40% of why humans were rated as more effective at emotion regulation. However, even after accounting for objectivity perceptions, the human label still retained a significant direct effect on perceived effectiveness, indicating that objectivity is just one of several factors explaining the human advantage.

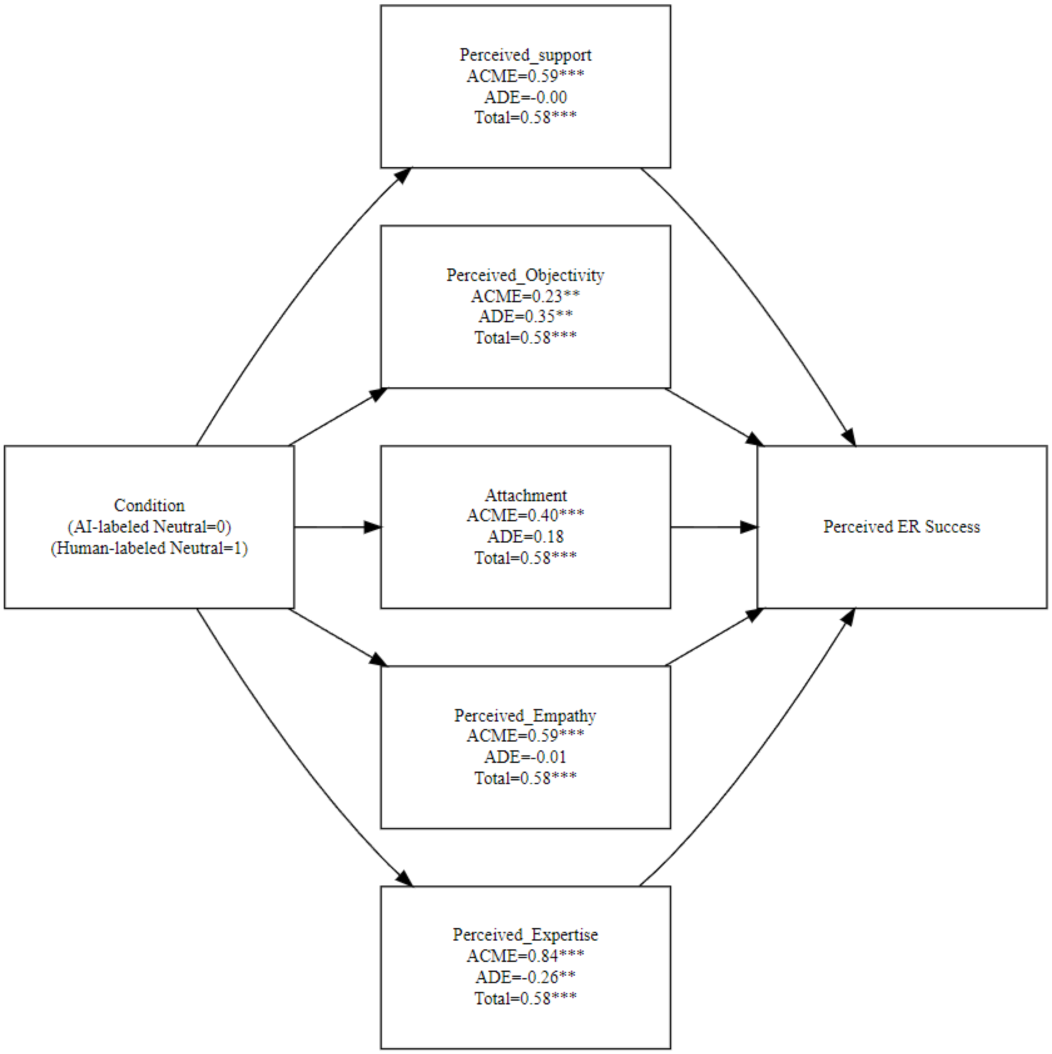
Attachment To the Partner. Attachment to the discussion partner substantially mediated the relationship between the partner label condition (AI-labeled Neutral versus Human-labeled Neutral) and perceived emotion regulation success (*ACME* = 0.40, 95% CI [0.18, 0.62], $p < .001$; *ADE* = 0.18, 95% CI [-0.04, 0.39], $p = .099$; proportion mediated = 69.2%, 95% CI [0.43, 1.10], $p < .001$; Fig. 9). Participants reported feeling considerably more attached to the human-labeled partner than to the AI-labeled partner, and this difference in attachment explained approximately 69% of why humans were perceived as more effective at emotion regulation.

Perceived Partner Empathy. Perceived empathy of the discussion partner fully mediated the relationship between the partner label condition (AI-labeled Neutral versus Human-labeled Neutral) and perceived emotion regulation success (*ACME* = 0.59, 95% CI [0.36, 0.83], $p < .001$; *ADE* = -0.01, 95% CI [-0.21, 0.20], $p = .931$; proportion mediated = 101.9%, 95% CI [0.73, 1.66], $p < .001$; Fig. 9). This complete mediation indicates that the indirect effect through perceived empathy fully accounts for the relationship between partner label and perceived effectiveness. In practical terms, participants perceived the human-labeled partner as substantially more empathetic than the AI-labeled partner, and this difference in perceived empathy explained why humans were rated as more effective at emotion regulation.

Perceived Partner Expertise. Perceived expertise of the discussion partner partially mediated the relationship between the partner label condition (AI-labeled Neutral versus Human-labeled Neutral) and perceived emotion regulation success with a suppression effect (*ACME* = 0.84, 95% CI [0.61, 1.09], $p < .001$; *ADE* = -0.26, 95% CI [-0.45, -0.06], $p = .01$; proportion mediated = 144.6%, 95% CI [1.08, 2.40], $p < .001$; Fig. 9). Statistically, this strong suppression pattern indicates that the mediator accounts for more than the total effect, with a significant negative direct effect emerging after accounting for the mediator. In practical terms, participants strongly associated the human-labeled partner with greater expertise compared to the AI-labeled partner, and this difference in perceived expertise powerfully increased effectiveness ratings. However, the significant negative direct effect suggests that when expertise perceptions are held equal between humans and AI, participants might actually rate AI as slightly more effective at emotion regulation.

Figure 9

Mediation Analysis: AI-labeled Neutral vs. Human-labeled Neutral in Perceived ER Success



Note. **p* < .05. ***p* < .01. ****p* < .001

Mediation Effect On AI-labeled Neutral vs. Emotional Validation Conditions. We then tested the five mediators of the emotional validation effect, using perceived emotion regulation success as the outcome variable.

Perceived Partner Support. Perceived support from the discussion partner fully mediated the relationship between the emotional validation condition (AI-labeled Neutral versus AI-labeled Emotional Validation) and perceived emotion regulation success (*ACME* = 0.50, 95% CI [0.26, 0.75], *p* < .001; *ADE* = -0.13, 95% CI [-0.31, 0.05], *p* = .146; proportion mediated =

136.2%, 95% CI [0.88, 3.58], $p = .016$; Fig. 10). Emotional validation significantly enhanced perceptions of AI supportiveness compared to the neutral approach, and this enhanced perceived support fully explained why the emotionally validating AI was rated as more effective at emotion regulation.

Perceived Partner Objectivity. Perceived objectivity of the discussion partner did not significantly mediate the relationship between the emotional validation condition (AI-labeled Neutral versus AI-labeled Emotional Validation) and perceived emotion regulation success ($ACME = 0.14$, 95% CI [-0.03, 0.30], $p = .099$; $ADE = 0.23$, 95% CI [-0.02, 0.48], $p = .075$; proportion mediated = 37.3%, 95% CI [-0.15, 1.13], $p = .093$; Fig. 10). Statistically, neither the indirect nor direct effects reached the conventional significance threshold ($p < .05$). This suggests that participants did not view emotional validation as compromising the AI's objectivity. Furthermore, differences in perceived objectivity were not a significant mechanism through which emotional validation enhanced perceived emotion regulation effectiveness.

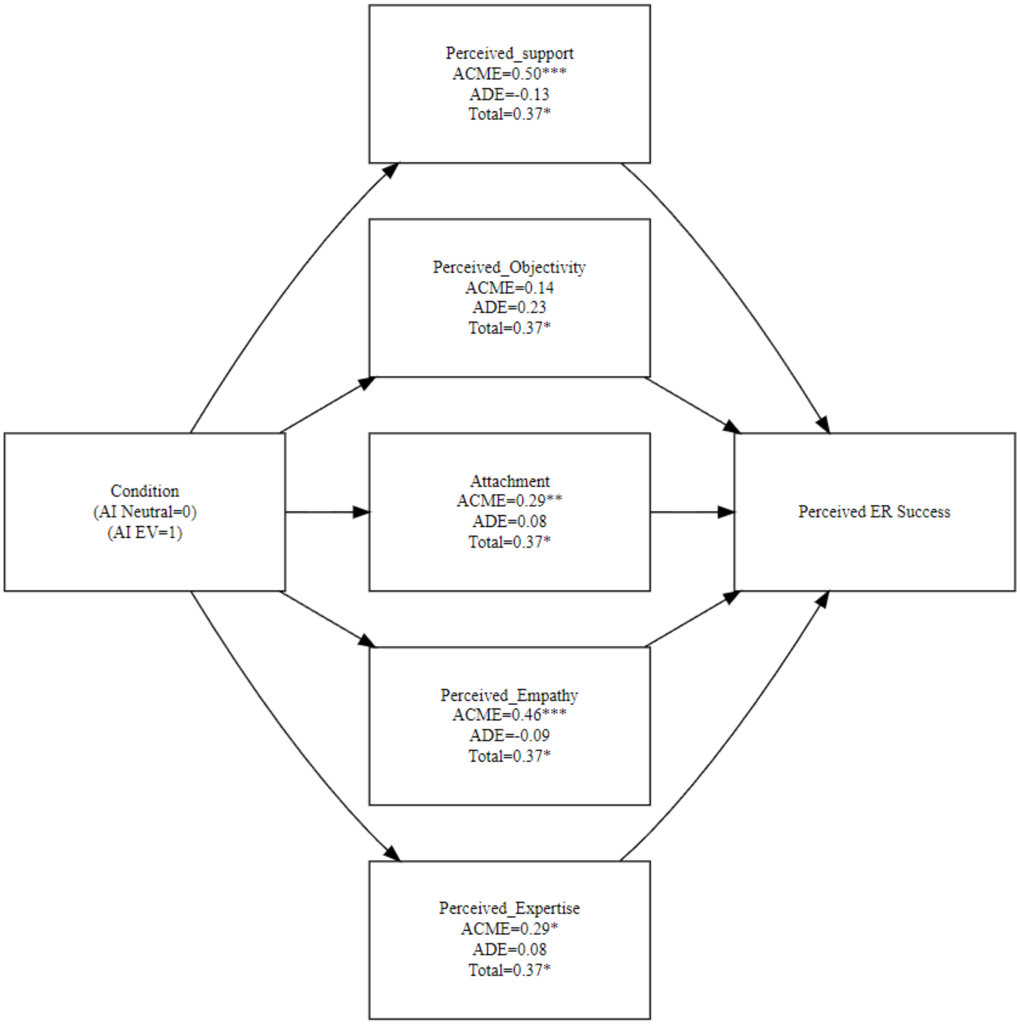
Attachment To the Partner. Attachment to the discussion partner substantially mediated the relationship between the emotional validation condition (AI-labeled Neutral versus AI-labeled Emotional Validation) and perceived emotion regulation success ($ACME = 0.29$, 95% CI [0.08, 0.50], $p < .01$; $ADE = 0.08$, 95% CI [-0.13, 0.29], $p = .448$; proportion mediated = 77.3%, 95% CI [0.32, 1.97], $p = .018$; Fig. 10). Emotional validation significantly increased participant attachment to the AI partner compared to the neutral approach, and this enhanced attachment explained approximately 77% of why participants rated the emotionally validating AI as more effective at emotion regulation.

Perceived Partner Empathy. Perceived empathy of the discussion partner fully mediated the relationship between the emotional validation condition (AI-labeled Neutral versus AI-labeled Emotional Validation) and perceived emotion regulation success (*ACME* = 0.46, 95% CI [0.23, 0.68], $p < .001$; *ADE* = -0.09, 95% CI [-0.29, 0.11], $p = .403$; proportion mediated = 123.7%, 95% CI [0.77, 3.61], $p = .013$; Fig. 10). These results indicate that empathy enhancement is a primary mechanism through which emotional validation improves perceived emotion regulation success.

Perceived Partner Expertise. Perceived expertise of the discussion partner substantially mediated the relationship between the emotional validation condition (AI-labeled Neutral versus AI-labeled Emotional Validation) and perceived emotion regulation success (*ACME* = 0.29, 95% CI [0.06, 0.51], $p = .017$; *ADE* = 0.08, 95% CI [-0.11, 0.27], $p = .392$; proportion mediated = 78.3%, 95% CI [0.31, 1.73], $p = .023$; Fig. 10). Emotional validation significantly enhanced perceptions of AI expertise in emotion regulation compared to the neutral approach, and this enhanced perceived expertise explained approximately 78% of the improvement in effectiveness ratings.

Figure 10

Mediation Analysis: AI-labeled Neutral vs. AI-labeled Emotional Validation in Perceived ER Success



Note. *p < .05. **p < .01. ***p < .001

Anger Reduction As the Outcome Variable

Mediation Effect On AI-labeled vs. Human-labeled Neutral Conditions. Similar to above, we tested the five mediators of the label effect, using reduction in anger before and after the conversation as the outcome variable.

Perceived Partner Support. Perceived support from the discussion partner fully mediated the relationship between the partner label condition (AI-labeled Neutral versus Human-labeled

Neutral) and anger reduction (*ACME* = -0.20, 95% CI [-0.30, -0.11], $p < .001$; *ADE* = -0.12, 95% CI [-0.35, 0.11], $p = .299$; proportion mediated = 61.7%, 95% CI [0.30, 2.11], $p < .01$; Fig. 11).

Perceived Partner Objectivity. Perceived objectivity of the discussion partner partially mediated the relationship between the partner label condition (AI-labeled Neutral versus Human-labeled Neutral) and anger reduction (*ACME* = -0.08, 95% CI [-0.16, -0.02], $p < .01$; *ADE* = -0.24, 95% CI [-0.47, 0.00], $p < .05$; proportion mediated = 25.8%, 95% CI [0.06, 0.90], $p < .05$; Fig. 11). Perceptions of objectivity explained approximately 26% of the label effect on anger reduction.

Attachment To the Partner. Attachment to the discussion partner fully mediated the relationship between the partner label condition (AI-labeled Neutral versus Human-labeled Neutral) and anger reduction (*ACME* = -0.11, 95% CI [-0.19, -0.04], $p < .001$; *ADE* = -0.21, 95% CI [-0.45, 0.01], $p = .069$; proportion mediated = 33.5%, 95% CI [0.13, 1.08], $p < .01$; Fig. 11). Participants' greater attachment to the human-labeled partner partially explained why they experienced more anger reduction with humans versus AI.

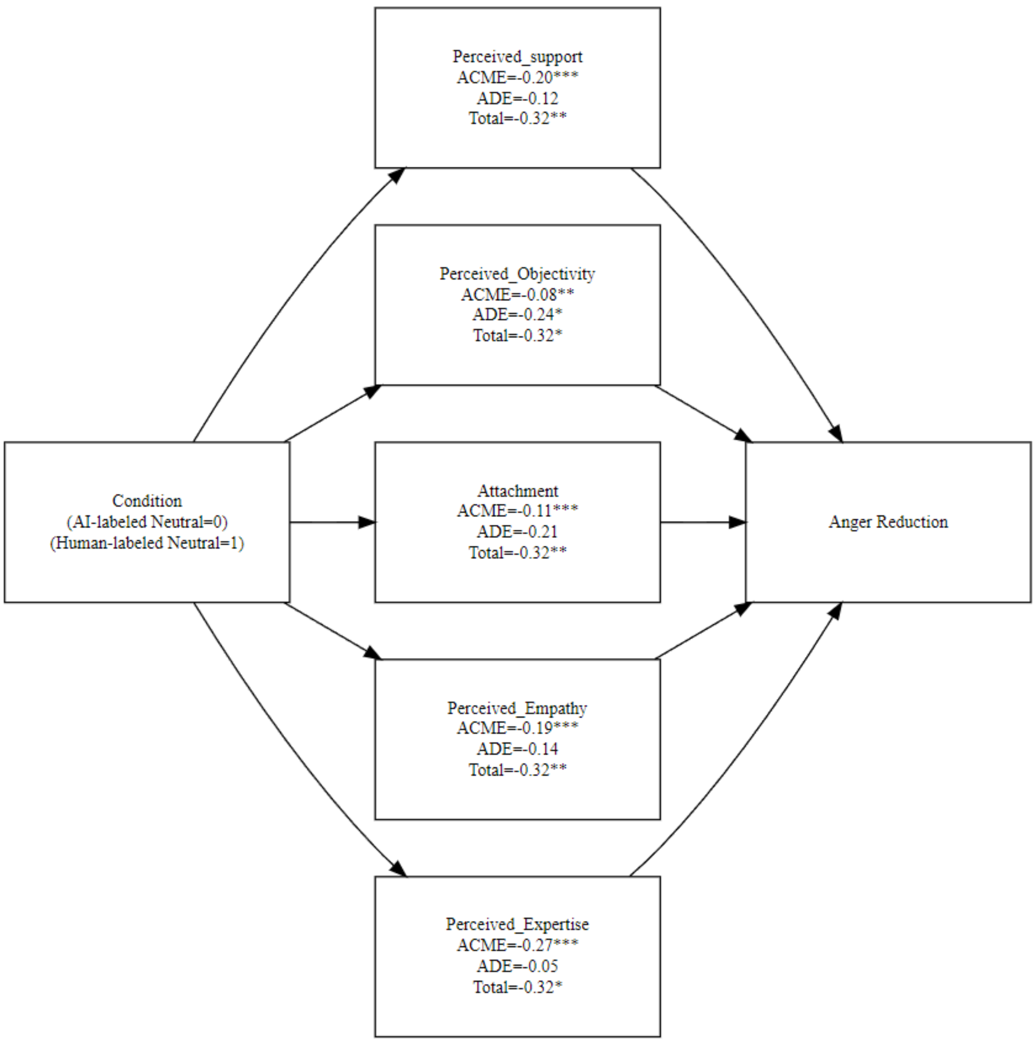
Perceived Partner Empathy. Perceived empathy of the discussion partner fully mediated the relationship between the partner label condition (AI-labeled Neutral versus Human-labeled Neutral) and anger reduction (*ACME* = -0.19, 95% CI [-0.28, -0.10], $p < .001$; *ADE* = -0.14, 95% CI [-0.37, 0.09], $p = .251$; proportion mediated = 57.5%, 95% CI [0.27, 1.82], $p < .01$; Fig. 11). This indicates that participants perceived the human-labeled partner as substantially more empathetic than the AI-labeled partner, which explained why they experienced greater anger reduction with the human-labeled partner.

Perceived Partner Expertise. Perceived expertise of the discussion partner fully mediated the relationship between the partner label condition (AI-labeled Neutral versus Human-labeled

Neutral) and anger reduction (*ACME* = -0.27, 95% CI [-0.39, -0.17], $p < .001$; *ADE* = -0.05, 95% CI [-0.28, 0.20], $p = .68$; proportion mediated = 84.4%, 95% CI [0.42, 2.88], $p < .05$; Fig. 11). This strong mediation suggests that participants' perception of the human-labeled partner's expertise was a powerful factor in explaining the greater anger reduction they experienced with the human versus AI partner.

Figure 11

Mediation Analysis: AI-labeled Neutral vs. Human-labeled Neutral in Anger Reduction



Note. *$p < .05$. **$p < .01$. ***$p < .001$

Mediation Effect On AI-labeled Neutral vs. Emotional Validation Conditions. We then tested the five mediators of the emotional validation effect, using anger reduction as the outcome variable.

Perceived Partner Support. Perceived support from the discussion partner fully mediated the relationship between the emotional validation condition (AI-labeled Neutral versus AI-labeled Emotional Validation) and anger reduction (*ACME* = -0.17, 95% CI [-0.27, -0.09], $p < .001$; *ADE* = -0.09, 95% CI [-0.30, 0.12], $p = .431$; proportion mediated = 66.0%, 95% CI [0.27, 2.57], $p < .05$; Fig. 12). Emotional validation significantly enhanced perceptions of AI supportiveness, which explained why participants experienced greater anger reduction with the emotionally validating AI.

Perceived Partner Objectivity. Perceived objectivity of the discussion partner did not significantly mediate the relationship between the emotional validation condition (AI-labeled Neutral versus AI-labeled Emotional Validation) and anger reduction (*ACME* = -0.06, 95% CI [-0.12, 0.01], $p = .108$; *ADE* = -0.20, 95% CI [-0.43, 0.02], $p = .073$; proportion mediated = 21.6%, 95% CI [-0.12, 0.95], $p = .128$; Fig. 12).

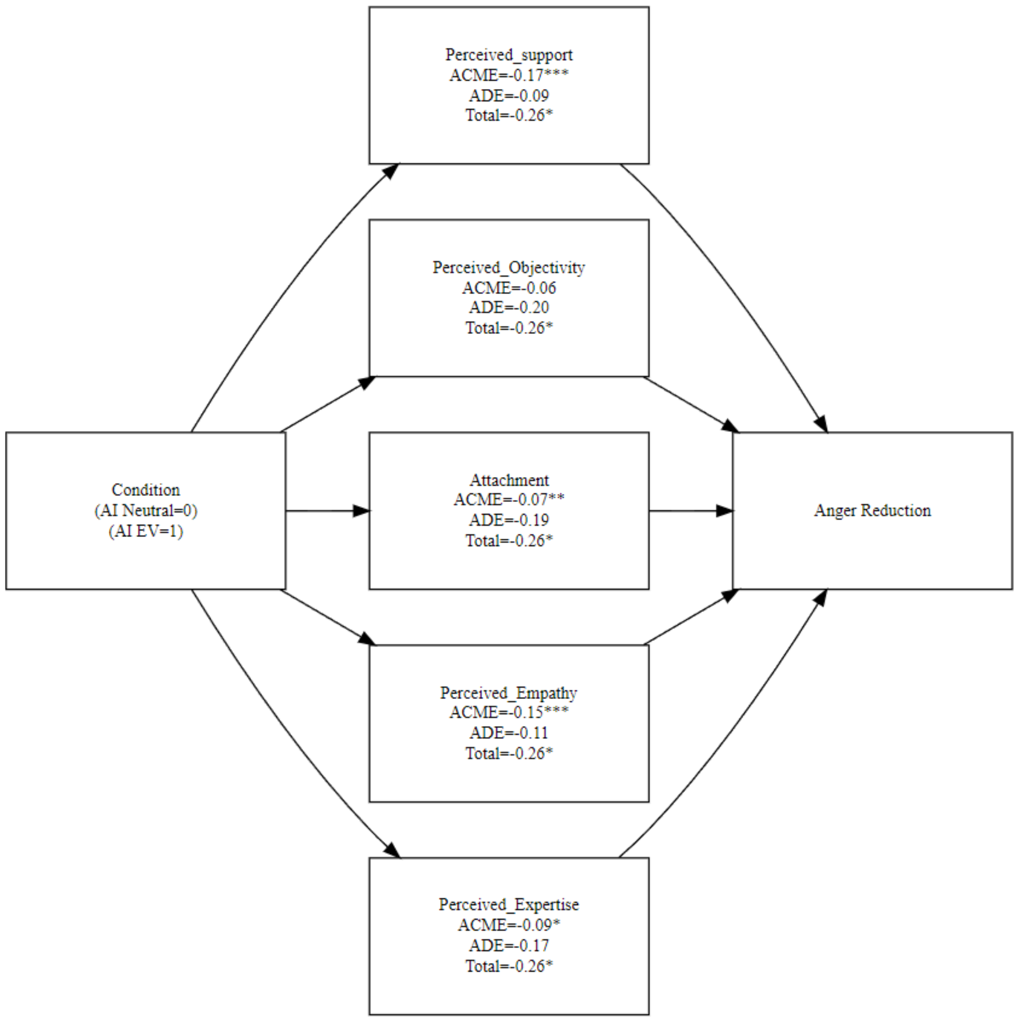
Attachment To the Partner. Attachment to the discussion partner fully mediated the relationship between the emotional validation condition (AI-labeled Neutral versus AI-labeled Emotional Validation) and anger reduction (*ACME* = -0.07, 95% CI [-0.14, -0.02], $p < .01$; *ADE* = -0.19, 95% CI [-0.40, 0.03], $p = .089$; proportion mediated = 27.5%, 95% CI [0.05, 1.11], $p < .05$; Fig. 12). Emotional validation increased participant attachment to the AI partner, which explained approximately 28% of the improvement in anger reduction.

Perceived Partner Empathy. Perceived empathy of the discussion partner fully mediated the relationship between the emotional validation condition (AI-labeled Neutral versus AI-labeled Emotional Validation) and anger reduction (*ACME* = -0.15, 95% CI [-0.24, -0.07], $p < .001$; *ADE* = -0.11, 95% CI [-0.32, 0.11], $p = .308$; proportion mediated = 57.0%, 95% CI [0.23, 2.29], $p < .05$; Fig. 12). These results indicate that empathy enhancement is a primary mechanism through which emotional validation improves anger reduction outcomes.

Perceived Partner Expertise. Perceived expertise of the discussion partner fully mediated the relationship between the emotional validation condition (AI-labeled Neutral versus AI-labeled Emotional Validation) and anger reduction (*ACME* = -0.09, 95% CI [-0.17, -0.02], $p < .05$; *ADE* = -0.17, 95% CI [-0.37, 0.05], $p = .129$; proportion mediated = 35.4%, 95% CI [0.04, 1.51], $p < .05$; Fig. 12). Emotional validation enhanced perceptions of AI expertise, which explained approximately 35% of the improvement in anger reduction.

Figure 12

Mediation Analysis: AI-labeled Neutral vs. AI-labeled Emotional Validation in Anger Reduction



Note. *$p < .05$. **$p < .01$. ***$p < .001$

The mediation results reveal the psychological mechanisms behind our two main findings. First, for why humans were perceived as better at emotion regulation than AI: Consistently across both perceived emotion regulation success and anger reduction, this occurred primarily because participants viewed humans being more empathetic and expert, providing more support than AI, and they were more attached to humans. Second, for why emotional validation improved AI's effectiveness: In both outcome measures, this happened mainly because emotional validation enhanced the AI's perceived expertise, supportiveness, and empathy, and human were more attached to AI with emotional validation. When AI used emotional validation, it significantly closed the gap with humans by enhancing these critical perceptions. Notably, across both outcome variables, emotional validation did not reduce perceived objectivity, suggesting that we can enhance warmth and empathy without sacrificing the AI's perceived credibility. These findings have practical implications for AI design: rather than trying to hide AI's identity, developers should focus on enhancing AI's perceived empathy, supportiveness, expertise, and human attachment through techniques like emotional validation. Such approaches can help bridge the perception gap between human and AI emotion regulation capabilities for both subjective assessments of effectiveness and objective measures of emotional change.

Discussion

Our study investigated the effectiveness of AI-driven cognitive reappraisal enhanced by emotional validation in managing workplace anger. We found that although participants perceived AI-delivered cognitive reappraisal as less effective than human-delivered interventions due to the AI Label Effect, incorporating emotional validation significantly improved the perceived effectiveness of AI interactions. Emotional validation narrowed the gap between AI and human emotional support by enhancing perceived empathy, support, expertise, and attachment.

Our findings contribute significant theoretical insights into the AI Label Effect within emotional support contexts. Specifically, we demonstrated that participants perceived AI-driven cognitive reappraisal as significantly less effective compared to reappraisal delivered by perceived humans, aligning with prior research highlighting diminished effectiveness when AI identity is disclosed (Yin et al., 2024; Rubin et al., 2025). However, our incorporation of emotional validation—a structured, theoretically grounded approach to emotional acknowledgment (Linehan, 1997)—substantially mitigated this negative perception. This finding advances theoretical understanding by identifying emotional validation as a critical mechanism for enhancing perceived AI expertise, empathy, support, and attachment.

Moreover, mediation analyses clarified the psychological mechanisms underlying the AI Label Effect. Participants perceived human-labeled conditions as superior primarily due to higher levels of empathy, support, and attachment. By embedding emotional validation, our study effectively addressed these perceptions, significantly narrowing the perceived effectiveness gap between human and AI interactions.

Practically, our findings have significant implications for implementing AI-driven emotional support systems in professional settings. Previous human-labeled AI studies typically filtered out large numbers of participants due to disbelief in the human-labeled conditions, limiting external validity. For example, Goel and colleagues (2024) found that approximately 21% of participants failed a manipulation check indicating they did not believe the assigned source label. In our study, we successfully retained a larger participant pool by adjusting user experience of the chatbot, ultimately filtering out fewer participants (11.74%) who doubted the human identity of their discussion partner. This methodological improvement offers a valid approach for future research and practical application in human-labeled AI research.

There may also be practical implications for organizations. Most existing emotion regulation programs depend on the presence of trained professionals - such as therapists, coaches, or workshop facilitators - to effectively teach these skills, which makes the process both time-consuming and expensive. Many organizations simply do not have the resources, staffing or infrastructure to support such programs. Integrating emotional validation into AI systems provides organizations with a viable strategy for personalized emotional support, addressing barriers such as resource-intensive and facilitator-dependent interventions (Yarker et al., 2022). Researchers and organizations can employ AI tools enhanced by emotional validation to manage anger effectively, reduce negative emotions, and potentially improve human relations and well-being.

Despite these contributions, several limitations must be acknowledged. First, although emotional validation significantly improved perceived AI effectiveness, it did not fully equate to the human-labeled emotional validation condition. This indicates that subtle elements inherent to human emotional interactions remain challenging for AI to replicate entirely, potentially including non-verbal cues, nuanced language interpretations, or biases against AI authenticity.

Second, our study employed short-term measures of emotional outcomes, capturing immediate but not longitudinal effects. The long-term effectiveness, sustained behavioral changes, and ongoing emotional benefits of AI-delivered emotional validation remain uncertain.

Future research should further explore methods to bridge the human-AI perception gap, possibly by integrating multimodal emotional validation strategies that include auditory or visual empathetic cues. Longitudinal studies are needed to assess the long-term effectiveness of AI-delivered emotional validation in reducing anger as well as other negative emotions, and improving other well-being and organizational outcomes.

Overall, our study shows the significant potential for AI emotional support systems contingent upon careful implementation of emotional validation strategies, inviting continued exploration into optimizing AI-human emotional interactions and AI's emotional support.

References

Abbott, J. A., Klein, B., Hamilton, C., & Rosenthal, A. J. (2009). The impact of online resilience training for sales managers on wellbeing and work performance.

American Psychological Association. (2023). *2023 Work in America Survey: Workplaces as engines of psychological health and well-being: (504232023-001)*. https://doi.org/10.1037/e504232023-001

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*(1), 3.

Berking, M., Meier, C., & Wupperman, P. (2010). Enhancing Emotion-Regulation Skills in Police Officers: Results of a Pilot Controlled Study. *Behavior Therapy*, *41*(3), 329–339. https://doi.org/10.1016/j.beth.2009.08.001

Buruck, G., Dörfel, D., Kugler, J., & Brom, S. S. (2016). Enhancing well-being at work: The role of emotion regulation skills as personal resources. *Journal of Occupational Health Psychology*, *21*(4), 480–493. https://doi.org/10.1037/ocp0000023

Cohen, N., & Ochsner, K. N. (2018). From surviving to thriving in the face of threats: The emerging science of emotion regulation training. *Current Opinion in Behavioral Sciences*, *24*, 143–155. https://doi.org/10.1016/j.cobeha.2018.08.007

Dang, J., & Liu, L. (2024). Extended artificial intelligence aversion: People deny humanness to artificial intelligence users. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspi0000480

Finkel, E. J., Rusbult, C. E., Kumashiro, M., & Hannon, P. A. (2002). Dealing with Betrayal in Close Relationships: Does Commitment Promote Forgiveness? *Journal of Personality and Social Psychology*, 82(6), 956–974. https://doi.org/10.1037/0022-3514.82.6.956

Fritz, C., Park, Y., & Shepherd, B. R. (2019). Workplace Incivility Ruins my Sleep and Yours: The Costs of Being in a Work-Linked Relationship. *Occupational Health Science*, *3*(1), 1–21. https://doi.org/10.1007/s41542-018-0030-8

Gallup. (2023). State of the global workplace 2023 report. https://www.gallup.com/workplace/349484/state-of-the-global-workplace.aspx

Goel, N., Bergeron, T., Lee-Whiting, B., Bohonos, D., Islam, M. M., Lachance, S., ... & Merkley, E. R. I. C. (2024). Artificial Influence: Comparing the Effects of AI and Human Source Cues in Reducing Certainty in False Beliefs.

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of general psychology*, *2*(3), 271-299.

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological inquiry*, *26*(1), 1-26.

Hajcak, G., & Nieuwenhuis, S. (2006). Reappraisal modulates the electrocortical response to unpleasant pictures. *Cognitive, Affective, & Behavioral Neuroscience*, *6*(4), 291-297.

Hancock, J. T., Naaman, M., & Levy, K. (2020). AI-Mediated Communication: Definition, Research Agenda, and Ethical Considerations. *Journal of Computer-Mediated Communication*, *25*(1), 89–100. https://doi.org/10.1093/jcmc/zmz022

Jackson, D. C., Malmstadt, J. R., Larson, C. L., & Davidson, R. J. (2000). Suppression and enhancement of emotional responses to unpleasant pictures. *Psychophysiology*, *37*(4), 515-522.

Jeon, J., & Park, D. (2024). Your feelings are reasonable: Emotional validation promotes persistence among preschoolers. *Developmental Science*, *27*(5), e13523. https://doi.org/10.1111/desc.13523

Jolliffe, D., & Farrington, D. P. (2006). Development and validation of the Basic Empathy Scale. *Journal of Adolescence*, *29*(4), 589–611. https://doi.org/10.1016/j.adolescence.2005.08.010

Kim, Y. H., Son, S. Y., & Kang, S. W. (2021). Effects of Anger and Moral Identity on the Relationship between Supervisors' Incivility and Deviant Behavior: A Study of Public Service Officers in Republic of Korea. *International Journal of Environmental Research and Public Health*, *18*(20), 10585. https://doi.org/10.3390/ijerph182010585

Lambie, J. A., & Lindberg, A. (2016). The role of maternal emotional validation and invalidation on children's emotional awareness. *Merrill-Palmer Quarterly*, *62*(2), 129-157. https://doi.org/10.13110/merrpalmquar1982.62.2.0129

Lerner, J. S., & Tiedens, L. Z. (2006). Portrait of the Angry Decision Maker: How Appraisal Tendencies Shape Anger's Influence on Cognition. *Journal of Behavioral Decision Making*, 19(2), 115–137. https://doi.org/10.1002/bdm.515

Li, J. Z., Herderich, A., & Goldenberg, A. (2024). Skill but not Effort Drive GPT Overperformance over Humans in Cognitive Reframing of Negative Scenarios. https://doi.org/10.31234/osf.io/fzvd8

Linehan, M. M. (1997). Validation and psychotherapy. In A. C. Bohart & L. S. Greenberg (Eds.), *Empathy reconsidered: New directions in psychotherapy.* (pp. 353–392). American Psychological Association. https://doi.org/10.1037/10226-016

Martin, A. E., & Mason, M. F. (2023). Hey Siri, I love you: People feel more attached to gendered technology. *Journal of Experimental Social Psychology*, *104*, 104402. https://doi.org/10.1016/j.jesp.2022.104402

McRae, K., Misra, S., Prasad, A. K., Pereira, S. C., & Gross, J. J. (2012). Bottom-up and top-down emotion generation: implications for emotion regulation. *Social cognitive and affective neuroscience*, *7*(3), 253-262.

Ochsner, K. N., Ray, R. D., Cooper, J. C., Robertson, E. R., Chopra, S., Gabrieli, J. D., & Gross, J. J. (2004). For better or for worse: neural systems supporting the cognitive down-and up-regulation of negative emotion. *Neuroimage*, *23*(2), 483-499.

Ohanian, R. (1990). Construction and Validation of a Scale to Measure Celebrity Endorsers' Perceived Expertise, Trustworthiness, and Attractiveness. *Journal of Advertising*, *19*(3), 39–52. https://doi.org/10.1080/00913367.1990.10673191

Porat, R., & Levy Paluck, E. (2024). Anger at work. *Frontiers in Social Psychology*, *2*, 1337715. https://doi.org/10.3389/frsps.2024.1337715

R Core Team (2024). R: A Language and Environment for Statistical Computing. Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Ray, R. D., McRae, K., Ochsner, K. N., & Gross, J. J. (2010). Cognitive reappraisal of negative affect: converging evidence from EMG and self-report. *Emotion*, *10*(4), 587.

Reif, E., Ippolito, D., Yuan, A., Coenen, A., Callison-Burch, C., & Wei, J. (2021). A recipe for arbitrary text style transfer with large language models. *arXiv preprint arXiv:2109.03910*. https://doi.org/10.48550/arXiv.2109.03910

Rubin, M., Li, J., Zimmerman, F., Ong, D. C., Goldenberg, A., & Perry, A. (2025). Comparing the Value of Perceived Human versus AI-Generated Empathy. *Nature Human Behaviour*.

Sharma, A., Rushton, K., Lin, I. W., Wadden, D., Lucas, K. G., Miner, A. S., ... & Althoff, T. (2023). Cognitive reframing of negative thoughts through human-language model interaction. *arXiv preprint arXiv:2305.02466*. https://doi.org/10.48550/arXiv.2305.02466

Shenk, C. E., & Fruzzetti, A. E. (2011). The Impact of Validating and Invalidating Responses on Emotional Reactivity. *Journal of Social and Clinical Psychology*, *30*(2), 163–183. https://doi.org/10.1521/jscp.2011.30.2.163

Tan, C. M. (2018). Search inside yourself. Bentang Pustaka.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis. *Journal of statistical software*, *59*, 1-38.

Truţă, C. (2013). Training Emotion Management Skills in the Workplace–An Individual Level Intervention on Education Employees. *Bulletin of the Transilvania University of Braşov, Series VII: Social Sciences and Law*, (2), 35-40.

Urry, H. L., Van Reekum, C. M., Johnstone, T., Kalin, N. H., Thurow, M. E., Schaefer, H. S., ... & Davidson, R. J. (2006). Amygdala and ventromedial prefrontal cortex are inversely coupled during regulation of negative affect and predict the diurnal pattern of cortisol secretion among older adults. *Journal of Neuroscience*, *26*(16), 4415-4425.

Van Doorn, E. A., Zeelenberg, M., & Breugelmans, S. M. (2014). Anger and prosocial behavior. *Emotion*, 14(4), 766–777. https://doi.org/10.1037/a0036793

Wang, K., Goldenberg, A., Dorison, C. A., Miller, J. K., Uusberg, A., Lerner, J. S., ... & Štrukelj, E. (2021). A multi-country test of brief reappraisal interventions on emotions during the COVID-19 pandemic. *Nature human behaviour*, *5*(8), 1089-1110.

Wang, L., Restubog, S. L. D., Shao, B., Lu, V., & van Kleef, G. A. (2018). Does anger expression help or harm leader effectiveness? The role of competence-based versus integrity-based violations and abusive supervision. *Academy of Management Journal, 61*(3), 1050–1072.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, *54*(6), 1063.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wilkowski, B. M., & Robinson, M. D. (2008). The Cognitive Basis of Trait Anger and Reactive Aggression: An Integrative Analysis. *Personality and Social Psychology Review*, 12(1), 3–21. https://doi.org/10.1177/1088868307309874

Yarker, J., Lewis, R., Sinclair, A., Michlig, G., & Munir, F. (2022). Meta-synthesis of qualitative research on the barriers and facilitators to implementing workplace mental health interventions. *SSM - Mental Health*, *2*, 100148. https://doi.org/10.1016/j.ssmmh.2022.100148

Yin, Y., Jia, N., & Wakslak, C. J. (2024). AI can help people feel heard, but an AI label diminishes this impact. *Proceedings of the National Academy of Sciences*, *121*(14), e2319112121. https://doi.org/10.1073/pnas.2319112121

Yip, J. A., & Schweitzer, M. E. (2019). Losing your temper and your perspective: Anger reduces perspective-taking. *Organizational Behavior and Human Decision Processes, 150,* 28–45.

Ziems, C., Li, M., Zhang, A., & Yang, D. (2022). Inducing positive perspectives with text reframing. *arXiv preprint arXiv:2204.02952*. https://doi.org/10.48550/arXiv.2204.02952

Zhan, H., Zheng, A., Lee, Y. K., Suh, J., Li, J. J., & Ong, D. C. (2024). Large language models are capable of offering cognitive reappraisal, if guided. *arXiv preprint arXiv:2404.01288*.