Seen but not Heard:

The Effects of Race and Emotional Expression on Jurors' Influence in Deliberation

by

Hannah Phalen

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved June 2023 by the
Graduate Supervisory Committee:

Jessica Salerno, Chair
Nicholas Schweitzer
Nicholas Duran

ARIZONA STATE UNIVERSITY

August 2023

ABSTRACT

Emotions are an important part of persuasion. Experimental research suggests that White and male jurors can use emotion to increase their influence, while other jurors cannot. This research builds on prior research by examining the relationship between naturally occurring emotion during mock jury deliberations and the influence that jurors hold. Participants (N = 708) in 153 mock juries watched a murder trial video and deliberated on a verdict. Participants self-reported their experienced emotions and rated their perceptions of the other jurors' emotion and influence. After data was collected, I extracted acoustic indicators of expressed emotion from each deliberation and used a speech emotion recognition model to classify each mock juror's emotional expression. I hypothesized that there would be an overall effect of emotional expression on influence such that as mock jurors' emotion increased, their influence would also increase. However, I hypothesized that a juror's race and gender would moderate the relationship between emotion and influence such that White male jurors will be seen as more influential when they are more emotional, and that female jurors and jurors of color will be seen as less influential when they are more emotional. I also hypothesized that female jurors of color will be doubly penalized for being emotional, due to their "double-minority" status. Bayesian model averaging suggested that the data was most probable under models that included perceived emotion, race, and the interaction between the two, compared to models that did not. Consistent with the hypothesis, as participants were perceived as more emotional, their influence increased. In contrast to the hypotheses, being perceived as more emotional increased influence for both White and non-White mock jurors but the effect was stronger for non-White jurors. In other words, while all jurors benefited from being perceived as more emotional, non-White jurors benefited more than White jurors. Male jurors were more influential than female jurors, and gender did not interact with emotion.. Although being perceived as more emotional predicted increased influence for all participants, this research demonstrates that there are racial and gender disparities in the level of influence that someone might hold on a jury.

DEDICATION

I dedicate this dissertation to my family. Your unwavering support throughout my graduate career has helped me achieve more than I ever would have thought possible.

First to my parents, Tom and Angie. Thank you for raising me to relentlessly pursue my dreams and believe in myself. Throughout my entire life, your steadfast belief that I can do anything that I could imagine has been a touchstone that I was always able to return to in moments of doubt.

Then to my siblings, Haeli and Landon. I am incredibly lucky to have younger siblings who are there to encourage me, laugh with me, and enrich my life in ways that I cannot describe.

And finally, to my husband, Garrett. You pick me up when I am down; you believe in me when I forget how to believe in myself; and you have built a life with me around my dreams and ambitions. Thank you for being a sounding board, for reminding me to eat, and for everything that you do to make our life together possible.

ACKNOWLEDGEMENTS

I would like to begin by thanking my graduate advisor, Dr. Jessica Salerno for her support and mentorship both in my graduate career and in the preparation of this dissertation. Words cannot appropriately express the profound impact that she has had on both my graduate education and my future in academia. Everything that I have achieved in graduate school was made possible through her mentorship and support. Anything that I achieve in my career will be shaped and made better through her mentorship. I genuinely do not believe that I could have had a better mentor and advisor.

I would also like to thank my dissertation committee members, Dr. Nicholas Schweitzer and Dr. Nicholas Duran for their time and contributions on both this project and on many other aspects of my graduate career.

I would also like to thank: Drs. Janice Nadler and Nicole Roberts and Susan Bandes for their assistance in developing the trial materials and study protocol that was used in this study; Drs. Robert Beattey, Douglas Kelley, Kristen Mickelson; Matthew Simonton, Stacia Stolzenberg, and Rick Trinkner, and Will Knight for acting as the judge, lawyers, and witnesses in the stimulus video; Anna Fine, Samantha Bean, Taylor Bettis, Madison Adamoli, Logan Bussey, Breanna Olson, Robin Moore, Kylie Kulak, Ali McCoy, and Brianna Bailey for their help with data collection; Emily Line for her feedback and advice on data analysis; and the National Science Foundation and the American Bar Foundation for their financial support of this project.

Finally, I would like to thank all the friends that I made during my seven years in graduate school for their support and friendship across the years.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

Seen but not Heard:

The Effects of Race and Emotional Expression on Jurors' Influence in Deliberation

The ideal American jury is one where jurors from a variety of diverse backgrounds and perspectives participate fully and equally in deliberation (Cornwell & Hans, 2011). But even as courts attempt to increase diversity on juries (Corley, 2021), diversity on juries will do little good if jurors of different races and genders are not able to participate fully and equally in deliberation. Specifically, even when a jury appears diverse, the ideal American jury will never be realized if the voices of White and male jurors are heard, and the voices of non-White and female jurors are ignored. In this research, I investigate one potential area where there might be inequality in the deliberation process: jurors' emotional expression.

Jury trials are often emotionally charged and during deliberation, jurors might want to express those emotions during discussions about the case. However, it is possible that some jurors might have more to gain from expressing those emotions than other jurors. This research aimed to examine the extent to which jurors of different demographics can leverage a common persuasive tool: emotional expression. I examined two independent but intersecting factors that might predict the level of influence held by a juror: 1) emotional expression by jurors; and 2) the gender and race of the jurors who were expressing emotion.

Specifically, I examined whether emotional jurors were more influential and whether a juror's race and gender moderated the effect of emotions on influence such that White men were more influential when they were more emotional, but women and people of color were less influential when they were more emotional. I also examined several exploratory moderators (case evidence, jury instructions about emotion, and deliberation sample) for the relationship between a jurors' emotions and their race.

Emotions and Social Influence

Emotions play an important role in social influence. The Emotions as Social Information Theory (EASI) posits that emotions provide information to observers, which might influence their cognitions, attitudes, and behavior (Keltner & Haidt, 1999; Van Kleef, 2009). People can infer information about the situation from others' emotional reactions. For example, research has shown that people perceive a situation to be less cooperative when another person expresses anger, compared to happiness (Van Doorn et al., 2012). This perception can bleed into behavior, with some research suggesting that people are less cooperative in a negotiation with an angry counterpart, compared to a happy counterpart (Van Dijk et al., 2008). This might suggest that, in a jury setting, people will perceive a juror who expresses anger more negatively than a juror who does not express anger and, in turn, that juror might hold less influence on the jury.

However, other research suggests that emotions might have the opposite effect in that they might increase influence. For example, research has shown that fear appeals can increase influence (Tannenbaum et al., 2015). Positive emotions, like enthusiasm and amusement can also facilitate persuasion (Griskevicius et al., 2010). Similarly, despite the research suggesting that people might be less cooperative when their negotiation partner is angry, other research suggests that participants are more willing to negotiate with an angry counterpart when the anger is directed at the offer, rather than the participant (Steinel et al., 2008). And a meta-analysis on anger's impact on persuasion found that message-relevant anger (e.g., anger about the crime the jury is deliberating about) can lead to attitude change that is consistent with the anger, especially when the message-relevant anger is paired with a call to action that might alleviate the anger (Walter et al., 2019).

Additionally, other research suggests that the impact of emotion on influence might depend on the appropriateness of the emotion. Specifically, when anger is inappropriate, expressing anger might trigger a strong negative affective reaction about the person who expressed the anger from the perceiver (Van Kleef & Côté, 2007), which might translate into the perceiver rating the person who expressed inappropriate anger as less influential. Similarly, other research suggests that this effect is not limited to anger: Inappropriate emotional expression in general reduces persuasion.

In contrast, expressing emotions that match the perceiver's expectations might trigger a strong positive affective response (DeSteno et al., 2004; Rose et al., 2006), which might translate into the increased influence. Given that popular press often portrays jury deliberation as a highly emotional, contentious discussion (e.g., 12 Angry Men; The Juror; Trial by Jury), people might see emotional expression as appropriate during jury deliberation and might, in fact, expect that other jurors will be emotional during deliberation. Taken together, this suggests that in a jury setting, people will perceive a juror who expresses anger more positively than a juror who does not express anger and, in turn, that juror might hold more influence on the jury.

In sum, research suggests that expressing emotions, and anger specifically, might be a useful tool in persuading others. However, there is a risk that anger expression might backfire because the other person might perceive the person who expresses anger more negatively. That is, when anger is seen as message-focused and appropriate, the person who expressed anger might be seen as more influential. But, when anger is seen as person-focused and inappropriate, the person who expressed anger might be seen as less influential. While it is possible that emotions might generally be seen as appropriate, and therefore more influential in a jury deliberation context, it is also possible that emotions might be seen as selectively appropriate depending on the juror that is expressing emotion. In this study, I have examined two important boundary conditions, the mock juror's race and gender, which might alter the impact of emotional expression on influence.

Gender and Race as Moderators of the Effect of Emotions on Influence

Juror Gender

One potential moderator of the effect of emotions on influence is the gender of the juror who is expressing emotion. Outside the legal system, research suggests that expressing anger is more beneficial for men than women. When men express anger in a professional setting, they are seen as more competent and effective but when women express the same anger, they are seen as less competent and effective (Brescoll & Uhlmann, 2008; Lewis, 2000; Tiedens, 2001). Additionally, people perceive anger to be detrimental to interpersonal workplace interactions when a woman is expressing anger but not when a man is expressing anger (Gibson et al., 2009).

Within the legal system, research follows the same general trend. Qualitative research suggests that White men not only use emotion in their own arguments to exert influence in a jury, but they also police the emotions of other jurors (Lynch & Haney, 2015). Other research suggests that male attorneys and jurors are received more positively when they express anger (compared to no emotion), but the opposite is true for female attorneys and jurors (Salerno & Peter-Hagene, 2015; Salerno et al., 2019).

One potential reason that women might be penalized for expressing anger while men are rewarded is because anger is an emotion that is stereotypically associated with men (Brescoll & Uhlmann, 2008; Tiedens, 2001). Research has shown that women are penalized for violating gender stereotypes in their social relationships (Robnett et al., 2016), their workplace relationships (Berdahl, 2007; Rudman & Glick, 2001), and in the legal system (Salerno & Peter-Hagene, 2015; Salerno et al., 2019; Salerno & Phalen, 2019). In sum, men might be rewarded for expressing anger in deliberation but, women might be punished for expressing the same level of anger because expressing anger violates gender stereotypes.

Juror Race

Additionally, a juror's race might moderate the effect of emotions on influence. Research on the effect of the interaction between race and emotions on influence is limited. In contrast to the research that shows that women might be penalized when they express emotions that violate a gender stereotype, research suggests that people will cooperate more when emotion violates a racial stereotype. Specifically, participants made more concessions to an angry negotiation partner when their negotiation partner was East Asian, compared to White but concessions did not differ based on race when the negotiation partner was not angry (Adam & Shirako, 2013). In that research, anger expression was the most effective when the participant held the stereotype that East Asian people were emotionally inexpressive, suggesting that emotion was most effective when that emotion violates a stereotype. Other research also suggests that Black people are punished for expressing anger (Motro et al., 2022; Salerno et al., 2019), even though anger is an emotion that is stereotypically associated with Black people (Walley-Jean, 2009). Taken together, this research suggests, in contrast to violations of gender stereotypes, people might be punished for conforming to racial stereotypes about emotion and rewarded for violating those stereotypes.

If people are indeed punished for both violating gender stereotypes and punished for conforming to racial stereotypes, it follows that Black women might be uniquely penalized for expressing anger, because their expression of anger both violates the gender stereotypes and conforms to racial stereotypes. However, research on the interactive effect of gender and race on influence is mixed. In one study, anger expression negatively impacted perceptions of a Black female employee more than it impacted perceptions of White employees or Black male employees, perhaps due to a Black woman's "double-minority" status (Motro et al., 2022). But in another study, there was no evidence of an interaction between mock jurors' gender and race on influence (Salerno et al., 2019). These two studies, taken together, suggest that more research is needed to more deeply understand how intersectional identities can impact the effect of emotion on influence.

Importantly, most of the research on how gender and race moderate the impact of emotions on influence has experimentally manipulated emotional expression. While these experimental manipulations of emotion are important for making causal claims about the impact of demographics and emotion on influence, it is possible that gender and race differences exist in these experiments because the emotional expression is artificial. It is possible that when people spontaneously express emotion, they express emotion differently than those emotions are expressed in this research. Or, men and women might naturally express anger differently in the real world. People might respond differently to those spontaneous expressions of emotion than to the artificially manipulated expressions of emotion seen in prior research. Specifically, it is possible that, when emotion is genuine and not manipulated, the emotion comes off as more authentic and, in turn, more persuasive, regardless of the expresser's race and gender. When research does examine naturally occurring emotion in deliberation, that research involves either qualitative investigations of the use of emotion in deliberation (e.g., Lynch & Haney, 2015) or asking real jurors to self-report their experiences after deliberation has finished (e.g., Hickerson & Gastil, 2008).

In this research, I have expanded on previous research on how gender and race interact with emotions to impact influence by quantitatively examining whether an increase in naturally occurring emotional expression within deliberating mock juries increases influence. I have also built on prior research by examining how well multiple measures of emotion predict influence (rather than relying entirely on post-deliberation self-reports.

Exploratory Moderators of the Interactions between Juror Demographics and Emotion

I also tested whether several exploratory moderators (e.g., the presence of emotionally evocative stimuli, jury instructions that caution jurors about the impact of emotions, whether the deliberation was in-person or online) of the interactions between juror gender and emotion and juror race and emotion. These moderators were elements of the study design (described more below in the procedures section) that might have moderated the hypothesized interactions.

First, I examined whether the interaction between demographics and emotional expression on influence will attenuate when mock jurors' gruesome photographs (compared to when they do not). Based on research that suggests that people often attribute the emotions expressed by women and people of color to an internal cause and the emotions expressed by White men to an external cause (e.g., Barrett & Bliss, 2009; Brescoll & Uhlmann, 2008; Motro et al., 2022), it is possible that when mock jurors are provided with an external justification (the gruesome photographs) for any emotional expression, they might attribute all emotions to that external cause. Further, because research suggests that emotional expression is more effective when the emotions are perceived as appropriate (Van Kleef & Côté, 2007), it is possible that mock jurors will see emotional expression as more appropriate when there is more emotional evidence (i.e., when jurors see are gruesome photographs).

Second, I examined whether the interactions between demographic information and emotional expression on influence were more extreme when mock jurors are given an emotion-awareness instruction, which informs them that they might get emotional when they view the trial evidence and that they should be aware of how those emotions impact their decisions, compared to a control instruction. Because research suggests that emotional expression is less effective when that expression violates norms (Van Kleef & Côté, 2007) and because research suggests that women and people of color are already penalized for expressing emotions (e.g., Salerno et al., 2019), the emotion-awareness instruction might be seen as a norm that women and people of color violate when they express emotion in deliberation.

Finally, I examined whether the interactions between demographic information and emotional expression on influence would be more extreme online than in-person. Research suggests that women and people of color struggle to be heard and are interrupted more in online video calls, compared to in-person conversations (Catalyst, 2022). Women and people of color also experience Zoom fatigue sooner and in a more extreme way than men and White people (Fauville, et al., 2023; Ratan, et al., 2022). Struggling to be heard on Zoom and Zoom fatigue might both contribute to a decrease in the amount of influence that women and people of color hold when deliberating on Zoom, compared to in-person.

Measures of Emotion

Research suggests that emotion can be measured at three levels: a target's experienced emotion, expressed emotion, and emotions as perceived by others, (Bastiaansen et al., 2019; Ekman, 1992). Research suggests that the correlation between these different measures of emotions is at best small and inconsistent (Mauss & Robinson, 2009), but it is not clear from current research which, if any, measures of emotion are better predictors of influence. Self-report measures of experienced emotion might be the most accurate indicators of how a person feels but because other people are not aware of these self-report measures, they often have to rely on behavioral and physiological cues to determine someone's emotional state. Thus, it is important to understand whether a person's self-report measures of experienced emotion (or their own evaluation of their emotions) predict the level of influence that they hold as well as other's perceptions of their emotion. Further, research suggests that people are more accurate in assessing the emotions of members of their ingroup, compared to members of their outgroup (Elfenbein & Ambady, 2002; Matsumoto, 2002). So, it is possible that peoples' ratings of perceived emotion, and in turn, their judgments of the other person's level of influence, might not accurately represent the level of emotion that someone is expressing. By comparing how well perceived emotion and acoustic indicators of expressed emotion predict influence, I can begin to draw conclusions about whether perceptions of another person's behavior are more predictive of the perceiver's opinion of that person than the actual expressed behavior. I measured emotion using self-report measures to assess *experienced emotion*, co-jurors' subjective assessments of each others' emotionality to assess *perceived emotion*, and acoustic indicators to assess *expressed emotion* in order to identify how and when emotion predicts influence.

Experienced Emotion

While self-report measures of emotion are the one of the most commonly used metrics in studies that examine how emotion influences behavior, there are a number of limitations to the sole use of self-report measures of emotion. In most research, self-report measures are collected after exposure to stimuli, which means that participants are retroactively reporting their recollections of the emotions that they were experiencing at an earlier point in time (Robinson & Clore, 2002). Recalling a past emotion, however, might not be the most accurate way to measure experiences. Memory (e.g., Thomas & Diener, 1990), a person's personality traits (e.g., Barrett, 1997; Bartz et al., 1996), and individual differences in language (Barrett, 2006) can all influence the accuracy with which someone might recall their own emotions. In addition to these limitations on the *accuracy* of self-report emotions, it is possible that a person's experienced emotion might not match their expressed emotion. Given that an observer has to rely on their perceptions of another’s emotion, rather than that person’s experienced emotion, it is possible that self-reported experienced emotion might be less predictive of influence over others than perceived emotion.

Perceived Emotion

However, the accuracy of perceived emotion might also be limited. Specifically, research suggests that people might be more accurate at identifying emotions of members of their ingroup than their outgroup (Elfenbein & Ambady, 2002; Matsumoto, 2002). And these potential inaccuracies could have an important impact on influence. For example, trying to increase influence by using emotional arguments might be less effective if others are not accurately identifying the person’s emotion. Alternatively, if someone is inaccurately perceived as expressing too much emotion, they might lose influence because that emotional expression is seen as inappropriate. That is to say, given the important impact of emotion on influence and the potential inaccuracies in how a target of influence is perceives someone’s emotion, it is important to examine, not only how experienced and expressed emotion predict influence, but also how perceived emotion predicts influence.

Expressed Emotion

Given the potential limitations in measuring both experienced and perceived emotion, some researchers have attempted to measure emotion using more objective measures of emotional expression. Specifically, researchers have attempted to predict a person’s emotion by analyzing acoustic indicators of emotion in speech using a multidimensional approach (Goudbeek et al., 2009; Magdin et al., 2019) that examines three dimensions of speech: arousal, valence, and power.

When examined individually, many aspects of speech are useful in identifying one dimension of emotion. For example, research has shown that variations in pitch are indicative of emotional arousal (Bachorowski & Owren, 1995; Scherer et al., 1991) but less useful in identifying valence or power. For example, high arousal emotions such as fear, joy, and anger are all associated with a higher pitch than low arousal emotions such as sadness or love. While variations in intensity are also positively associated with arousal (especially in men), intensity is also positively associated with power (Pereira, 2000). That is, as speech becomes more intense, the speaker is more likely to be expressing a high arousal, high power emotion, such as anger or joy. Research also suggests that the frequency of different formant (the frequency components of speech that are used to make phenomes) are associated with different levels of arousal, power, and different valences (Goudbeek et al., 2009). A multidimensional approach to classify emotion in speech that incorporates the parts of speech that are associated with arousal, valence, and power can be used to provide a fairly accurate picture of someone's emotional expression. Of course, as with the other measures of emotion, there are potential limitations in using acoustic indicators of emotion to measure expressed emotion. For example, these measures often do not take into account that people might express multiple emotions simultaneously (Devillers et al., 2005). Additionally, most of the models that have been developed to predict emotion train the model on manufactured or scripted emotion (e.g., databases where actors express specific emotions). The rare models that are developed using naturally occurring emotion necessarily rely on either self-reports of experienced emotion or ratings of perceived emotion to judge the accuracy of the model, which means that those models are subject to the same limitations as the self-report and perceived emotion measures (Schuller et al., 2011).

That is to say, all three types of emotion measures are subject to some limitations in their accuracy. However, these measures can still provide important information about how emotions influence judgments. In this research, I examine the predictive value of all three types of emotion measure (experienced, perceived, and expressed) in order to develop a fuller picture of how emotion might predict a person's influence.

Research Overview

While the literature provides some indication of how emotional expression might impact influence differently depending on the mock juror's race and gender, there is less focus on whether these mock jury experiments replicate when emotion is naturally occurring. This research begins to answer that question. Specifically, I examined the impact of naturally occurring emotion on influence using three different measures of emotion. Then, I examined whether a mock juror's race and gender moderate the impact of emotion on influence. Finally, I tested several exploratory moderators of the impact of emotion and race on influence and the impact of emotion and gender on influence. The data analysis plan was pre-registered and all data analysis scripts are available on the Open Science Framework (OSF): https://osf.io/vtzf3/?view_only=6e5d98e1e4d94867b5718028a1217eab.

Hypothesis 1: Emotion and Influence

I hypothesized that there will be an overall effect of emotional expression on influence. Based on research that suggests that emotion is related to increased persuasion, when jurors express more emotion, I hypothesized that they will have more influence on their co-jurors.

Hypothesis 2: Emotion and Influence Moderated by Gender and Race

However, I hypothesized that the overall effect of emotional expression on influence will be moderated by a juror's race and gender. Specifically, because research suggests that women and people of color are more likely to be penalized when they express anger, relative to men and White people, I hypothesized that male jurors and White jurors will be seen as more influential when they express negative emotions (especially anger), compared to when they do not, but I hypothesize that female jurors and jurors of color will be seen as less influential when they express negative emotion (especially anger), compared to when they do not. I hypothesized that female jurors of color will be doubly penalized for expressing emotion, due to their "double-minority" status.

Method

I will conduct a secondary data analysis on deliberation data that was collected to test the impact of seeing gruesome photographs, jury instructions, and deliberation on juror decisions. The Institutional Review Board at Arizona State University approved this research (Protocol #00009106). See Appendix A for the IRB approval.

Participants

We collected data from 1138 mock jurors in 201 mock juries that comprised between 4 and 10 mock jurors. During the first half of the study, participants were recruited from Craigslist and Amazon's Mechanical Turk to attend an in-person study at Arizona State University. Participants were recruited from the Phoenix metropolitan area.

The first half of the data collection took place in person and then, due to health and safety protocols during the COVID-19 pandemic, the second half of data collection took place online over Zoom. During the second half of the study (the half that took place online), participants were initially recruited from Craigslist in five cities (Sacramento, Albuquerque, Cincinnati, Hartford, and Jacksonville) and from Amazon's Mechanical Turk. However, the Craigslist advertisements were quickly sent overseas, and we had to discontinue Craigslist advertising because participants were using VPN connections to circumvent our requirement that they be US citizens residing in the US. Therefore, most of the data collection in the second half of the study took place on Amazon's Mechanical Turk. Before attending the study, participants completed a pre-screening that determined if they were eligible for the study. Participants were not eligible if one of the following was true: 1) they were unwilling to attend a two to three hour session; 2) they did not consent to seeing gruesome photographs of a murder victim; 3) they were unwilling to be video-recorded; 4) they reported that they were not eligible for jury service; or 5) they had a vison or hearing impairment. In the second half of data collection, potential participants were also not eligible to participate if one of the following was true: 1) they did not computer access in a private room; 2) they did not have a webcam or were unwilling to use their webcam; 3) they did not have a microphone; or 4) they had an internet upload speed of less than 1.5 mbps. (See Appendix B for the in-person version of the eligibility pre-screener and Appendix C for the online version of the eligibility pre-screener).

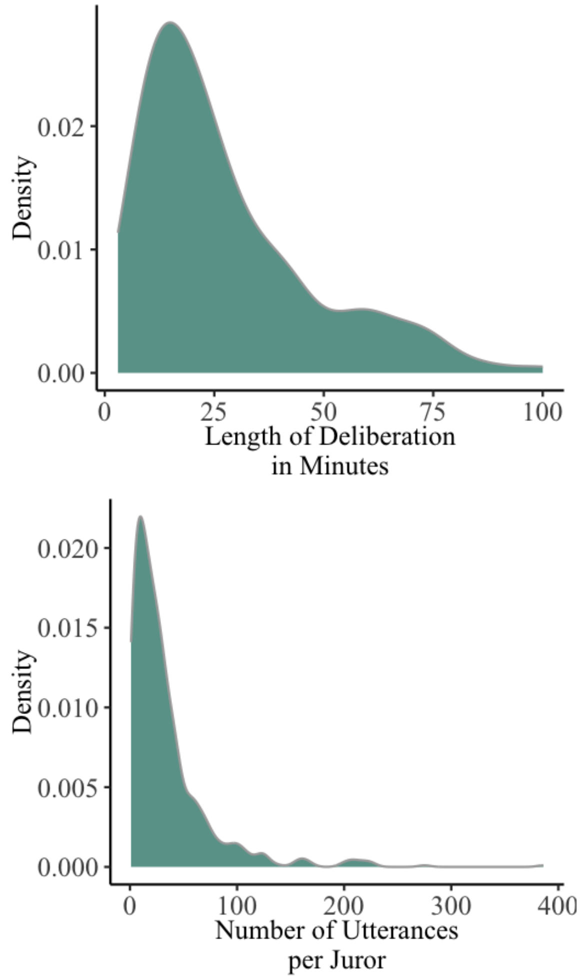
I excluded mock jurors from these analyses if they 1) failed a manipulation check about the evidence that they saw during trial (one of the exploratory moderators; n = 22, 1.93%); 2) reported that they had previously participated in a study using the same stimuli (n = 2, 0.18%); 3) joined a Zoom deliberation using an IP address that was not located within the US (n = 84, 7.38%); 4) incorrectly completed the dependent measures (e.g., if they provided influence ratings for multiple jurors who were not on their jury or failed to rate themselves on the influence measures; n = 332, 29.17%); or 5) I was not able to extract acoustic indicators of emotion from their voice (either because they did not participate in deliberation or because they did not speak audibly enough to be recorded; n = 186, 16.34%). In order to conduct Bayesian model comparison, all models must be built from the same data. Therefore, all exclusion criteria were used for all models. For example, because I used self-reported emotion as an independent variable in several models, participants who failed to rate themselves on the influence measure were not included in any analyses. Similarly, if I was unable to extract acoustic indicators of emotion from the participants' voices, those participants had missing data for the acoustic measures of emotion and were, therefore, excluded from all analyses.

We excluded entire juries if 1) more than 50% of the jurors would have been excluded for joining a Zoom deliberation using an IP address that was not located within the US (n_{jury} = 11, 5.47%, n_{juror} = 68, 5.98%); 2) a technical issue made the entire deliberation unusable (n_{jury} = 21, 10.45%, n_{juror} = 124, 10.90%). After exclusions, 6 juries (3.95%) had 2 eligible jurors and 24 juries (15.79%) had three eligible jurors. The other 122 (80.26%) juries had 4 or more eligible jurors.

The final sample was 708 mock jurors who deliberated in 153 juries. Participants were 73.31% White, 9.46% Black, 5.79% Hispanic/Latino, 5.93% Asian, 2.54% Mixed Race, 1.84% Indigenous, 0.14% Native Hawaiian or Pacific Islander, and 0.99% selected

“Other”. Participants had a mean age of 39.59 (*SD* = 14.32) and were 60.59% women. The average jury size in the final sample was 5.65 (*SD* = 1.47) jurors. Figure 1 shows distributions of deliberation length, the length of time each juror spoke, the number of utterances per juror, and the percentage of utterances in each deliberation spoken by each juror. Deliberations ranged from 3 minutes long to 100 minutes long, with a mean length of 27.61 minutes (*SD* = 19.93). Almost all of the deliberations (98.04%, *n* = 150) lasted over 5 minutes and the vast majority (82.35%, *n* = 126) lasted over 10 minutes. Within a deliberation, jurors spoke for, on average, 3.26 minutes (*SD* = 3.75). And while each juror spoke an average of 34.00 times (*SD* = 39.80), around one-third of those utterances were under 5 seconds long.

Figure 1.
Length of Deliberation and Deliberation Participation



Sensitivity Analysis

Because this was a secondary data analysis of another research question, I conducted a sensitivity analysis using the powerlmm package in R (Magnusson, 2018). The powerlmm package conducts simulated sensitivity analyses for nested three-level models (Influence Score nested within Juror who are nested within Jury). For the purposes of the sensitivity analysis, I assumed that each jury was made up of 5 jurors and that there was an intra-class correlation of .5 between jurors' ratings of the other jurors. The sensitivity analysis indicated that the final sample of 708 participants yields power of .80 to detect effects as small as d > .20, power of .90 to detect effects of d > .22, and power of .95 to detect effect of d > .26.

Replication Power Analysis

However, the sensitivity analysis that I conducted was limited in that powerlmm is not set up to detect interactions or unequal sample sizes, both of which could reduce the power to detect the above stated effects. Therefore, I also followed Kruschke's (2014) method of conducting a replication power analysis. A replication power analysis examines the probability that the same results would be achieved if the study was replicated exactly. This is calculated by generating new simulated data that is representative of the original data and fitting a new model that uses the posterior that was derived from the original data as new priors to that new, simulated data. This process is tantamount to conducting an exact replication on the exact same population and fitting a new model to the new data using the original data as the prior. This process is repeated 500 times and then, I calculated the probability that I would achieve the same pattern of results that I describe below. I used the most predictive model to conduct this replication power analysis and found that the replication power was 88.62%. In other words, if the experiment was replicated 500 times, the same pattern of results would emerge 88.62% of the time.

Procedure

Participants viewed a trial video in which a man was accused of murdering his wife, while the defense argued she committed suicide. During the trial video, participants were randomly assigned to either (a) view pathologist testimony without victim photographs or (b) view the same testimony with gruesome photographs of an alleged murder victim superimposed on the trial video. They were also randomly assigned to hear either (a) standard jury instructions or (b) emotion awareness instructions. In all conditions, an expert pathologist described the injuries shown in the gruesome photographs so that all participants had similar information about the victim's injuries. After watching the trial videos, participants deliberated in 4-10-person juries until they reached agreement (or 30 minutes remained in the session) either online or in person. After deliberation, they completed measures rating the other jurors on their level of influence, persuasiveness, likeability, competence, warmth, and emotionality. Participants also self-reported how emotional they were while viewing the trial evidence.

I used Praat (Boersma & Weenink, 2022) and Parselmouth (Feinberg, 2022) to extract auditory features from each deliberation. First, I split every deliberation into utterances. Each utterance is a statement from a juror that is more than one second long. Then, I ran every utterance through Praat to extract the pitch, intensity, and formant position of the utterance. Finally, I used machine learning (described below) to conduct speech emotion recognition (Rockikz, 2019).

Materials

See Appendix D for a transcript of the trial video, the photograph stimuli, and both sets of jury instructions. The trial was based on an actual case where a man was accused of murdering his wife (*R v. Valevski*, 2000). The gruesome photographs and basic information about the original trial came from Bright and Goodman-Delahunty (2006) but the trial transcript was developed independently. In all conditions, the trial contained excerpts from jury instructions, opening statements and closing arguments from the prosecution and defense, direct and cross-examination of three prosecution witnesses, direct and cross-examination of one defense witness, closing arguments from the prosecution and defense, and non-gruesome photographs of a door and locks. A lawyer reviewed the materials for their validity.

Testimony from the defendant's sister established that the defendant and the victim had an argument the night of the murder and the victim locked herself in their bedroom after telling him that he would be sorry when she was gone. The defendant and his neighbor found the victim in the bedroom the next morning behind a locked door. A locksmith testified that the defendant could have maneuvered the lock to make it appear locked from the inside and a forensic pathologist testified that the bodily evidence was consistent with homicide. The couple's neighbor testified about the victim's depression and the defendant's desire to reconcile with his wife.

Gruesome Photographs

Participants saw the same neutral photographs and heard the same testimony by a forensic pathologist describing the photographs. This information included a description of the size and shape of the wounds. Participants were randomly assigned to view no additional gruesome photographs or four gruesome photographs in color superimposed over the trial video. The photographs were taken from the original case evidence. One photograph depicted the victim at the crime scene and the other three photographs were autopsy photographs of the face, upper body, and neck. All three photographs show a knife wound on the neck.

Jury Instructions

Participants were randomly assigned to see either standard jury instructions or emotion-awareness jury instructions at the end of the case evidence. The standard instructions were modified from actual Illinois pattern jury instructions for first-degree murder. The authors—two of whom are law professors—wrote the emotional instructions. The emotional instructions were loosely based off New Jersey pattern instructions that attempt to teach jurors how to deal with emotional evidence. In contrast to the New Jersey instructions, however, the emotional instructions were informed by research on juror reactions to emotional evidence. Specifically, in the emotional instructions, the judge read seven additional sentences that were not included in the control instructions:

> In this case, photographs of the deceased might be admitted in evidence. If so, these photographs have been admitted to provide details about the victim's injuries and to help you visualize issues relevant to the case. You may, understandably, find the photographs upsetting. Be aware that in addition to helping you resolve the issues in the case, the photographs may also influence your decision in inappropriate ways. Being upset about the disturbing events depicted in the photograph might make you want to convict someone for the crime. This desire to convict someone might lower your threshold for how much proof you need to believe that the State has met the "beyond a reasonable doubt" standard and convict. The desire to convict might also influence you to pay more attention to evidence that supports a guilty verdict than you pay to evidence that supports a not guilty verdict.

Measures

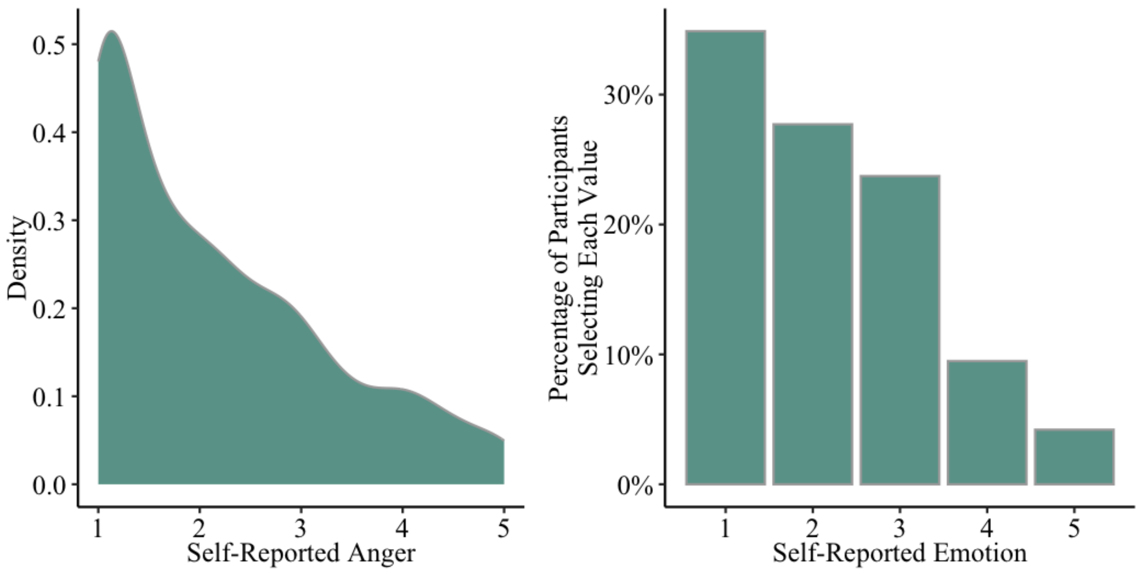
Measures of Emotion

I collected several measures of emotion that break down into three categories: 1) self-report measures of emotion; 2) perceptions of others' emotions; and 3) acoustic indicators of emotion.

Self-Report Measures of Emotion. Participants reported their own emotions in two measures. First, participants reported the extent to which they felt contempt, outrage, anger, and infuriated when viewing the evidence on 5-point scales from *Not at all* to *Very Much*. I averaged these four measures together to create a *self-reported anger* score (α =

0.91). Second, participants rated how emotional they were during deliberation on a 5-point scale from *Not at all* to *Extremely*. I used this measure as the *self-reported emotion* predictor. As shown in Figure 2, participants tended to rate themselves as having relatively low levels of emotion on both measures (M_{anger} = 2.15, SD_{anger} = 1.12; $M_{emotionality}$ = 2.23, $SD_{emotionality}$ = 1.16).

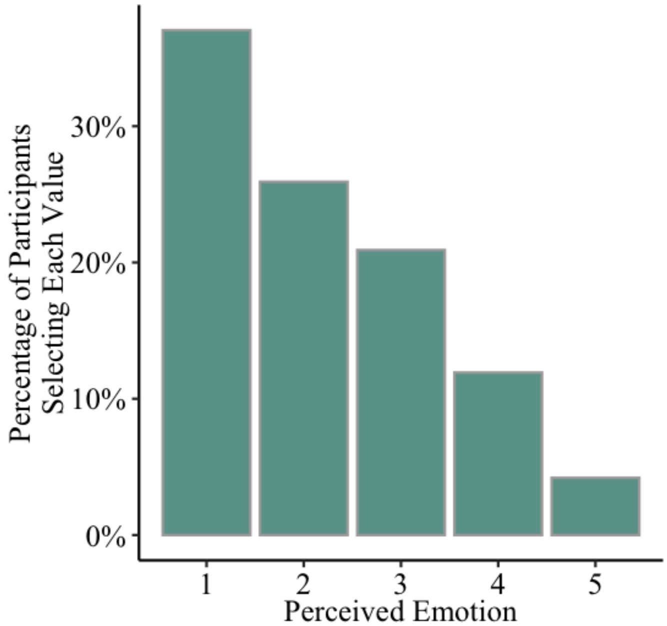
Figure 2.
Distribution of Self-Reported Emotion



Others' Perceptions of The Target's Emotions. Participants were also asked to rate the other participants on the extent to which they were emotional during deliberation on 5-point scales from *Not at all* to *Extremely*. This was done in a round-robin design so that each juror rated themselves (the self-report measure of emotion mentioned above) and the other jurors on their emotionality. The *perceived emotion* measure represented how emotional every other juror on the jury felt the target juror was and each juror was given a number of perceived emotion ratings equal to the number of other jurors on their jury. For example, if a jury was made up of five jurors, each juror would have four unique perceived emotion ratings, each one representing how emotional the other four jurors thought that juror was. Again, as shown in Figure 3, other participants tended to rate the target participant as having low levels of emotion (*M* = 2.20, *SD* = 1.18).

Figure 3.

Distribution of Ratings of Perceived Emotion of Other Jurors

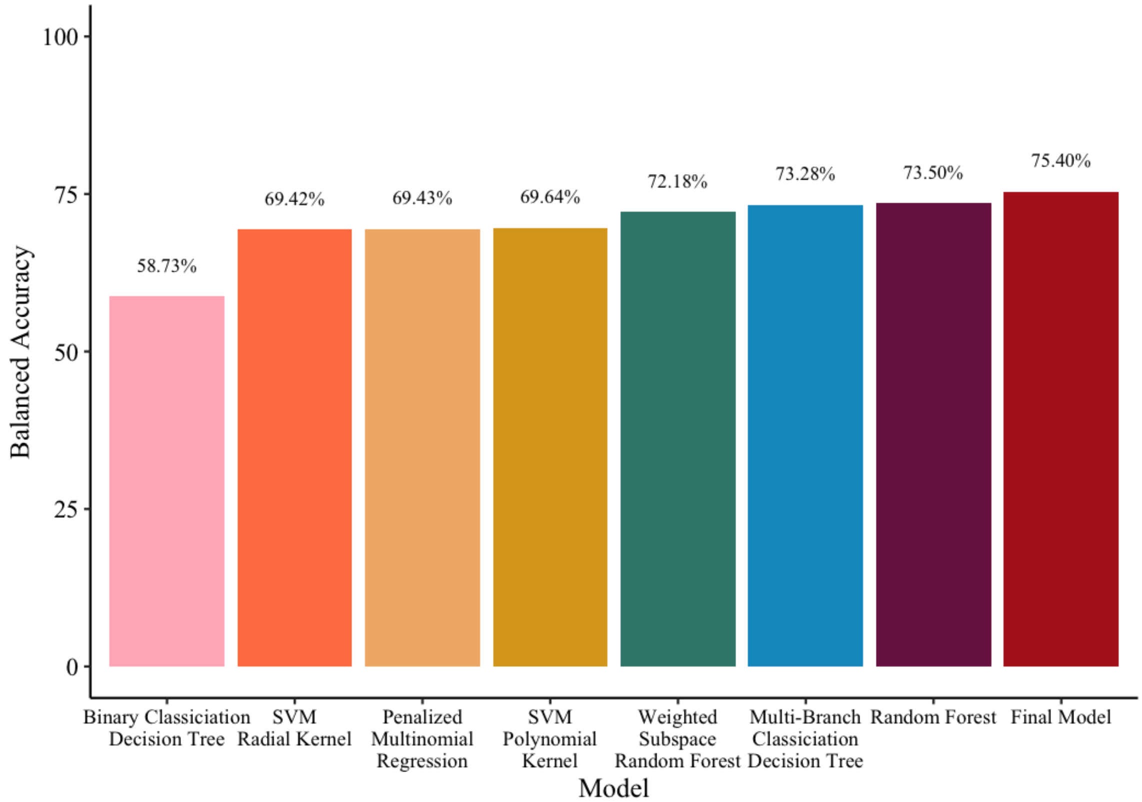


Acoustic Indicators of Emotion. I used Praat to collect data on each juror's pitch, formants, and intensity during deliberation. Then, I used a speech emotion recognition model that was trained on the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) to classify each juror's average emotion across the deliberation as one of the following: 1) calm (*n* = 13); 2) happy (*n* = 56); 3) sad (*n* = 38); 4) angry (*n* = 418); 5) fearful (*n* = 5); 6) surprised (*n* = 42); or 7) disgusted (*n* = 135). The speech emotion recognition model was an ensemble model that used stacking to combine seven weaker machine learning models into one more powerful meta-model using a weighted vote (Alhamid, 2022; Freund & Schapire, 1996). Each of the seven weaker models used a different machine learning method. The model was trained on 80% of the RAVDESS audio files and tested on the remaining 20% of the RAVDESS audio files. In other words, after the model was trained on 80% of the RAVDESS dataset, I used the model to classify emotions from the remaining 20% of the RAVDESS dataset. Then, I compared the model's classification with the ground truth from the RAVDESS dataset in order to determine the accuracy of the model. Figure 4 shows the balanced accuracy for each of the seven weaker models and for the final meta-model. The resulting model predicted emotion with 75.40% accuracy and anger specifically with 74.00% accuracy, which is considered ideal model performance (Barkved, 2022). Figure 5 shows the confusion matrix from the final model. A confusion matrix shows both the correct classifications and, when there was an incorrect classification, how the model was incorrect.

I used four acoustic indicators of emotion as predictors in the models discussed below. First, I used the classifications from the speech recognition model as a measure of *expressed emotion*. Then, because I was particularly interested in the impact of anger on influence, I also recoded the expressed emotion measure into a dichotomous variable (*expressed anger*) where a 1 indicated that the speech recognition model had classified the participant's average emotion as angry (n = 418) or as any other emotion (n = 289). Finally, because pitch and maximum intensity are often associated with anger, I conducted analyses examining participants' *mean pitch* and *maximum intensity* across all of their utterances in deliberation. Figure 6 shows the variations in maximum intensity and mean pitch both over the course of the deliberation and across jurors.

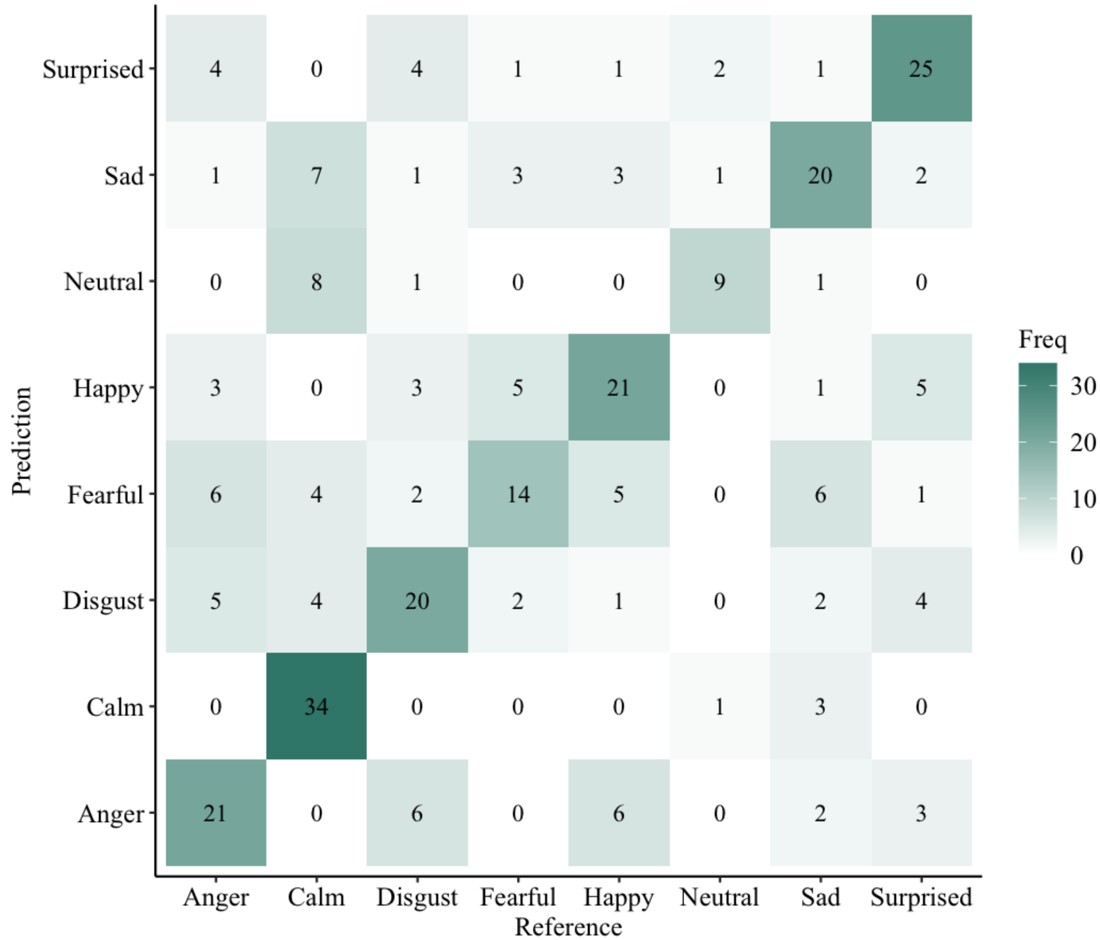
Figure 4.

Balanced Accuracy for Each of Machine Learning Model



Note. Balanced accuracy is a measure of how well the model predicted the test dataset. The models were trained on 80% of the RAVDESS dataset and tested on 20% of the RAVDESS dataset. Balanced accuracy is calculated by averaging the true positive rate (or how often the model correctly identified a participant in the test set as expressing an emotion when they were, in fact, expressing that emotion) and the true negative rate (or how often the model correctly *failed* to identify a participant in the test set as expressing an emotion when they were, in fact, *not* expressing that emotion).

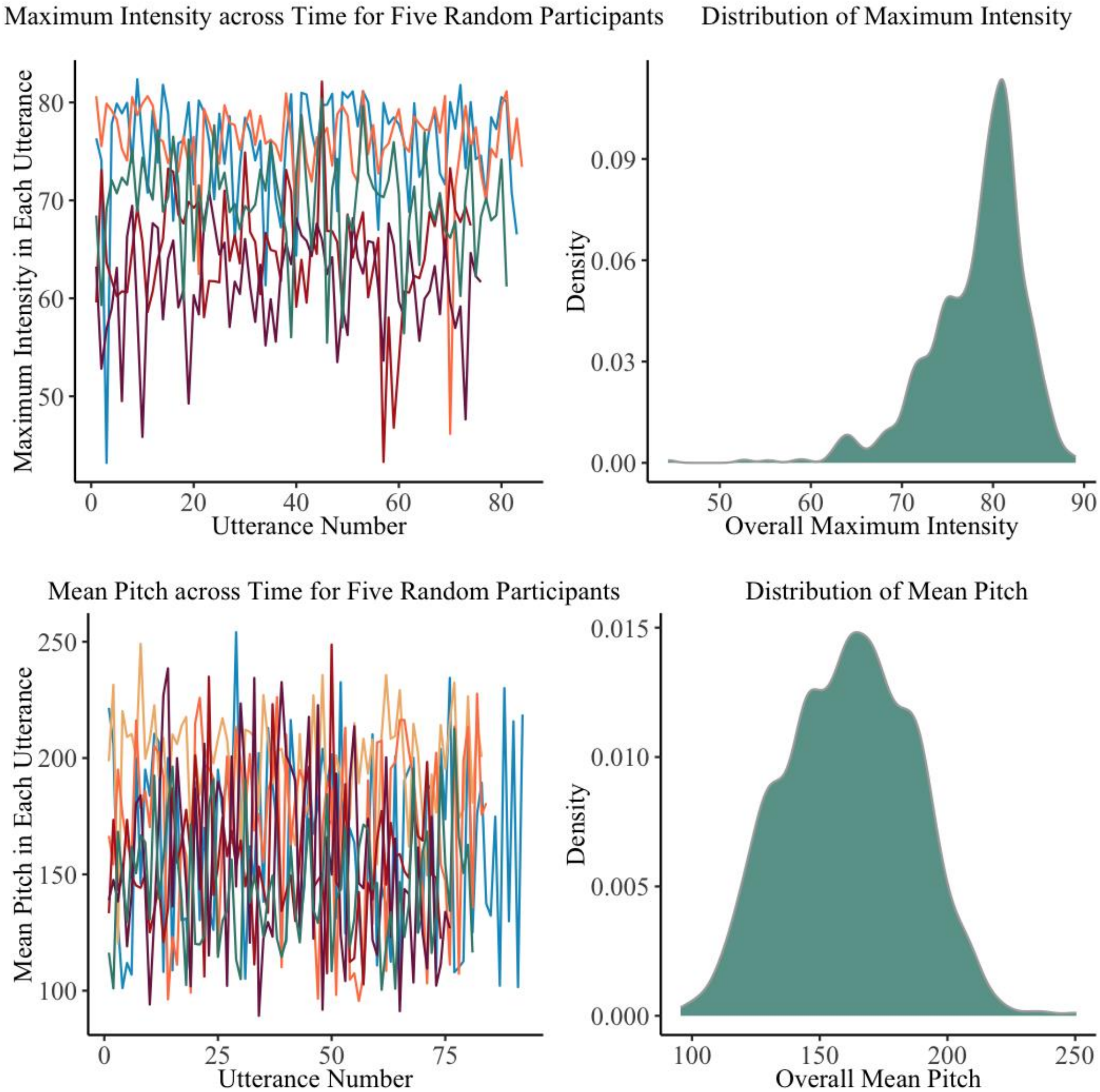
Figure 5.
Confusion Matrix in the Final Model



Note. The confusion matrix represents how well the model predicted emotion in the RAVDESS test dataset. The x-axis is the correct emotion that the RAVDESS participant was actually expressing, and the y-axis is the model's prediction. Darker colors indicate more frequent guesses.

Figure 6.

Variations in Maximum Intensity and Mean Pitch



Note. Maximum intensity is measured in dB and mean pitch is measured in Hz. The average maximum intensity for a calm voice in the RAVDESS dataset is 55.70 dB (*SD* = 5.59) and the average maximum intensity for an angry voice in the RAVDESS dataset is 75.71 dB (*SD* = 8.12). The average pitch for a calm voice in the RAVDESS dataset is 159.14 Hz (*SD* = 47.23). The average pitch for an angry voice in the RAVDESS dataset is 199.80 Hz (*SD* = 35.12).

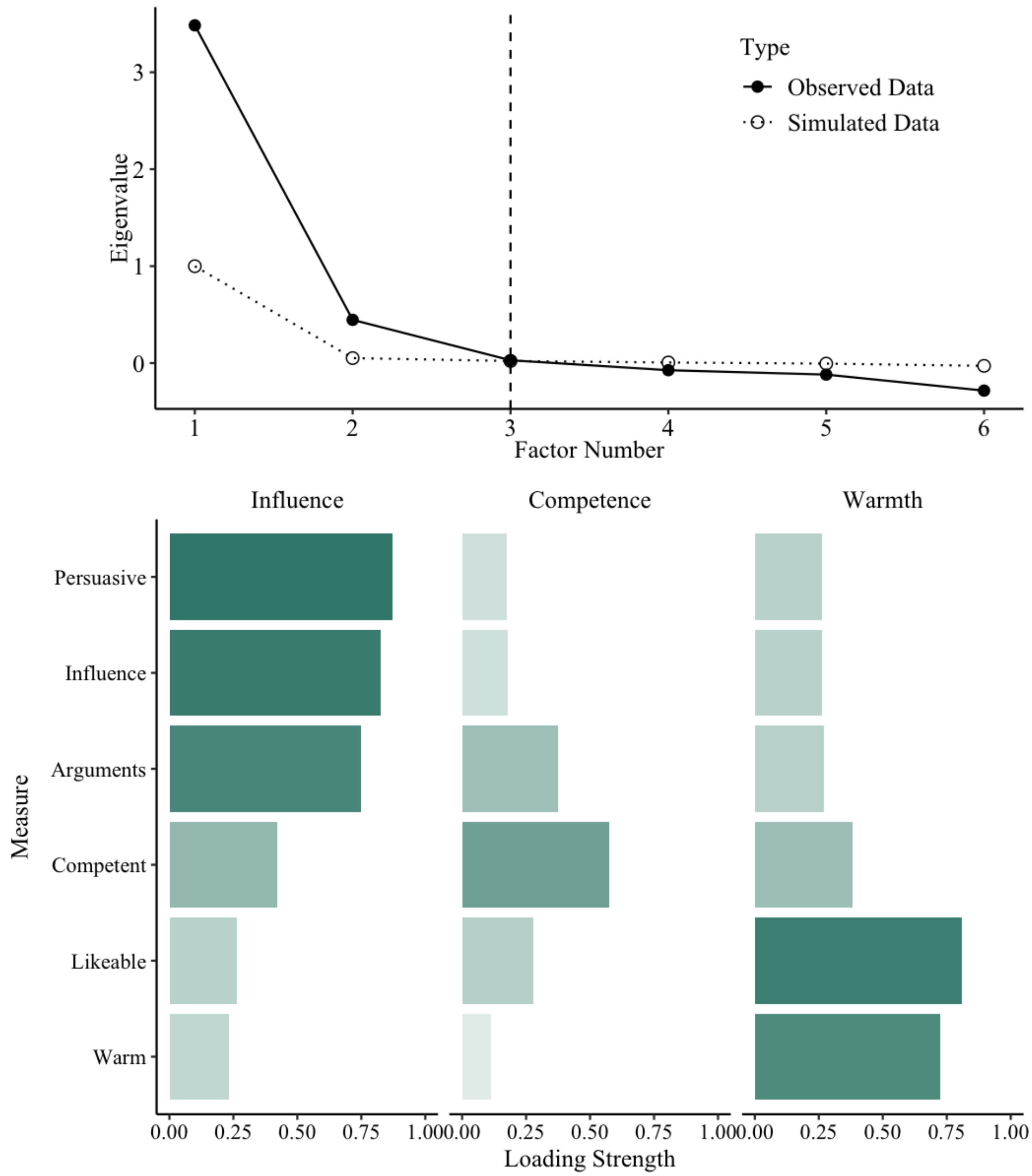
Juror Race and Gender.

Jurors self-reported their race and gender. Because all jurors in the final sample reported that they were either men or women, I dummy-coded gender (man = 0 [n = 278] and woman = 1 [n = 429]). Because of the relatively small percentage of non-White jurors, I dummy-coded race as either White (0, n = 518) or non-White (1, n = 189).

Response Variable (Influence)

Participants were asked to rate each juror on the extent to which they were 1) influential, 2) persuasive, 3) presented high quality arguments, 4) likable, 5) competent, and 6) warm on 5-point scales from Not at all to Very Much. This was done in a round-robin design so that each juror rated themselves and the other jurors on each of the measures. I conducted an exploratory factor analysis, which indicated that influential, persuasive, and presented high quality arguments all loaded onto one factor (Figure 7). I averaged participants' scores on those three items to create an influence score (α = 0.92). Because influence was the primary response variable of interest, I did not examine the other two factors.

Figure 7.
Results of the Exploratory Factor Analysis



Analytic Approach

I conducted a series of Bayesian multi-level models using the R package *brms* (Bürkner, 2017) to address my hypotheses. There are several benefits to using Bayesian modeling for these analyses, over null hypothesis significance testing (NHST). These benefits include: 1) being able to make more concrete probabilistic statements about the probability of each hypothesis, given the data; 2) more precise estimation of parameters in complex models; and 3) the ability to estimate the relative likelihood of two alternative models, given the data.

Bayesian modeling is a type of statistical analysis that uses preexisting beliefs (the prior) and the collected data to generate a posterior, or a mathematical estimate of the likelihood of the hypothesis, given the data (Kruschke, 2014). Unlike NHST, the goal of Bayesian analysis is not to determine if there is enough evidence against the null hypothesis to reject that null hypothesis. Rather, the goal of Bayesian analysis is to quantify the probability that a hypothesis is true, given prior beliefs about the hypothesis and the data. Another key difference between NHST and Bayesian analysis is the use of credible intervals, rather than confidence intervals. While NHST assumes that the true parameter value is a fixed value, Bayesian analysis assumes that the true parameter value is a random variable. Where confidence intervals are a probabilistic statement about the likelihood that the interval contains the true parameter, credible intervals are a probabilistic statement about the location of the true parameter value. In other words, where a 95% confidence interval can be interpreted as a range of values that would contain the true parameter value in 95% of repeated samples, a 95% credible interval can be interpreted as a 95% probability that the true parameter value is within the interval (Kruschke, 2014).

In a multi-level modeling context, Bayesian analyses have several advantages over NHST such as the ability to estimate parameters in complex models that involve moderator variables, the ability to make statistically reliable inferences with small sample sizes, and the ability to deal with random missing data (Bates et al., 2015; Lüdtke et al., 2013). Bayesian analyses are also more reliable than NHST in models that involve cross-level interactions (such as the interaction between perceived emotion [a level 1 variable] and juror race [a level 2 variable] and jury instructions [a level 3 variable]; Stegmueller, 2013).

Bayesian analyses also take a different approach to model comparison. Where model comparison in a frequentist framework compares the model likelihood, or the probability of the data, given the model, across two or more models, model comparison in a Bayesian framework compares the model posteriors, or the probability of the model, given the data, across two or more models (VanderPlas, 2015). In this way, under a Bayesian framework, we can determine the relative evidence of one model over another.

I implemented the Bayesian workflow described by Martin et al. (2021):

1. Collect the data.
2. Build the models.
3. Fit the models to the data.
4. Evaluate the posterior predictive distribution.
5. Compare the models.
6. Select the models that are most likely, given the data.
7. Summarize the posterior of the best model.

Results

Overview of the Bayesian Workflow

A more thorough description of the model building process can be found in Appendix E, but I have summarized the process here. To build the models that address hypotheses 1 and 2, I began by determining the distribution that best describes the observed data. Influence was slightly left-skewed (skewness = -0.27) and platykurtic (kurtosis = -0.68). However, these values for skew and kurtosis fall within the recommended range to conclude that data is normally distributed (George & Mallery, 2010).

Then, I determined the random effects structure that best modeled the data collection process. In line with Barr et al. (2013) recommendation that researchers should use the maximal random effects structure that is justified by the design, I began by testing a model with the maximal random effects structure such that rating and rated juror were crossed random effects that were nested within dyad and dyad was, in turn, nested within jury. However, that model failed to converge, which can increase the risk of Type I errors. Therefore, I followed the recommendations of Bates et al. (2015) and I removed random effects until I found the most parsimonious model. Based on visual inspection of the pairs plot from the most complex model (See Appendix E) and because the research questions are primarily focused on perceived influence, rather than susceptibility to influence, I first removed the random effect relating to Rating Juror and then, when the model still failed to converge, I removed the random effect relating to Dyad. The new model converged and the final random effects structure was:

$$Influence \sim 1 + (1 \mid Jury) + (1 \mid Jury{:}RatedJuror)$$

Finally, I developed a set of weakly informative, regularizing priors for all models (Gill & Witko, 2013; Stan Development Team, 2023). Weakly informative, regularizing priors provide some information about the expected scale of the dependent variable while containing enough uncertainty to prevent the prior from having undue weight in the model. They also assume there will be no effect of the predictors on the response variable but allow for a high amount of uncertainty in the estimate. These priors are useful because they can prevent some over-fitting to the data and make the model more robust to outliers. The final priors that I used in all models were:

$$\beta_0 \sim N(3, .75)$$

$$\beta_n \sim N(0, 1)$$

$$\sigma \sim HalfT(3,0,.1)$$

$$\Sigma \sim HalfT(3,0,2.5)$$

β_0 represents the intercept prior for every model; β_n represents the prior for all predictors in the model; σ represents the prior for the standard deviation of all random effects; and Σ represents the prior for the residual standard deviation.

After building the Intercept-Only model, I fit 28 Bayesian mixed effects models to the data in order to address Hypotheses 1 and 2. Seven models contained one of the seven emotion measures. Seven models contained an interaction between one of the emotion measures and the gender of the rated juror (i.e., the juror whose level of influence was being judged). Seven models contained an interaction between one of the emotion measures and the race of the rated juror. Seven models contained a three-way interaction between one of the emotion measures, the gender of the rated juror, and the race of the rated juror. Including the Intercept-Only model, I built 29 models. I examined the posterior predictive distribution for each model and after finding that each model was operating as expected, I compared the performance of the 28 models to each other and to the performance of an Intercept-Only model that did not include any predictors.

I used two model comparison methods to determine the best model: Bayes factors and leave-one-out cross-validation (LOO-CV). I used $bayestestR$ (Makowski et al., 2019) to calculate Bayes Factors and $brms$ (Bürkner, 2017) to conduct leave-one-out cross-validation. These two methods serve two unique but important purposes. Bayes factors can be used to determine the marginal likelihood of each model to determine which model best describes the data. In its simplest form, a Bayes factor is the ratio of the posterior probabilities of two models (Wetzels et al., 2011). Because Bayes factors are a ratio, they can be interpreted in a way that is similar to an odds ratio. For example, a Bayes factor of 2 implies that the data are twice as likely under the alternative hypothesis, compared to the reference group. Jeffreys (1961) provides a widely accepted set of verbal labels to categorize Bayes factors (see Table 1).

Below, I report the results of the best performing models (the models that outperformed the Intercept-Only model based on their Bayes factors and LOO-CV) and I briefly summarize the results of the other models and the results of exploratory hypotheses in the Alternative Models section. Correlations between all continuous independent variables are shown in

Figure **8**. Descriptive statistics for all categorical variables are shown in Table 1.

Table 1.

Jeffreys (1961) Suggested Interpretation of Bayes Factors

Bayes factor	Interpretation
>100	Decisive evidence for H_A
30–100	Very strong evidence for H_A
10–30	Strong evidence for H_A
3–10	Substantial evidence for H_A
1–3	Anecdotal evidence for H_A
1	No evidence
1/3–1	Anecdotal evidence against H_A
1/10–1/3	Substantial evidence against H_A
1/30–1/10	Strong evidence against H_A
1/100–1/30	Very strong evidence against H_A
<1/100	Decisive evidence against H_A

Figure 8.

Correlations between all Continuous Variables

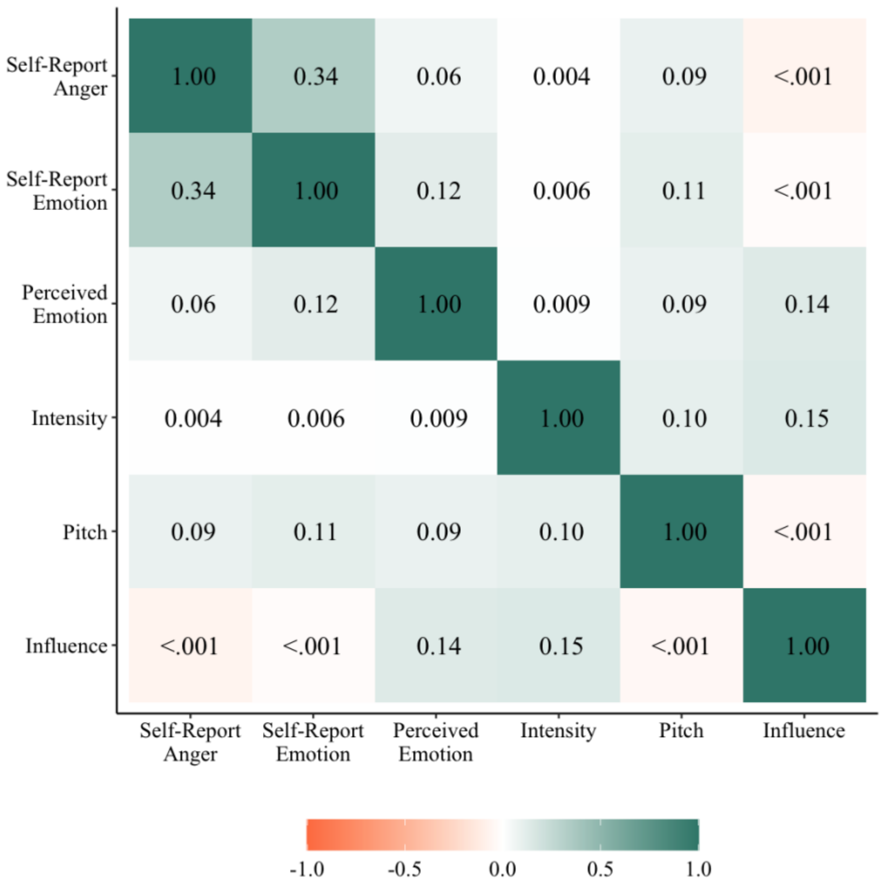


Table 1.

Descriptive Statistics of all Continuous Predictors and Influence by all Categorical Predictors

Categorical Predictor	Category	*n* (%)	Pitch	Intensity	Perceived Emotion	Self-Report Anger	Self-Report Emotion	Influence
Gender	Female	429 (60.68%)	172.20 (21.71)	77.87 (5.15)	2.34 (1.23)	2.31 (1.16)	2.34 (1.18)	3.11 (1.09)
	Male	278 (39.32%)	146.85 (21.71)	78.43 (4.92)	2.00 (1.08)	1.91 (1.02)	2.06 (1.10)	3.24 (1.01)
Race	Non-White	189 (26.73%)	160.43 (25.56)	77.36 (5.20)	2.18 (1.18)	2.11 (1.03)	2.21 (1.25)	3.03 (1.10)
	White	518 (73.27%)	162.89 (24.77)	78.36 (5.00)	2.21 (1.18)	2.16 (1.15)	2.24 (1.12)	3.21 (1.04)
Acoustic Indicators of Emotion	Calm	13 (1.84%)	120.32 (19.30)	65.22 (8.57)	2.20 (1.12)	2.40 (1.15)	2.38 (0.96)	2.78 (1.15)
	Angry	418 (59.12%)	162.99 (23.93)	80.56 (2.84)	2.26 (1.22)	2.16 (1.14)	2.24 (1.19)	3.29 (1.04)
	Disgust	135 (19.09%)	159.58 (23.29)	73.61 (4.33)	2.25 (1.17)	2.23 (1.09)	2.13 (1.08)	2.94 (1.05)
	Fearful	5 (0.71%)	163.59 (57.62)	76.13 (3.96)	2.29 (1.20)	1.85 (1.02)	1.60 (0.89)	3.74 (0.93)
	Happy	56 (7.92%)	166.83 (22.76)	79.45 (2.38)	1.98 (1.06)	2.11 (1.26)	2.46 (1.29)	3.08 (1.06)
	Sad	38 (5.37%)	161.76 (30.65)	76.38 (5.63)	2.09 (1.15)	2.01 (0.95)	2.29 (1.04)	3.23 (1.05)
	Surprised	42 (5.94%)	170.29 (22.21)	71.86 (3.68)	1.91 (0.98)	1.95 (0.99)	2.07 (1.02)	2.95 (1.09)
	Not Angry	289 (40.88%)	161.13 (26.44)	74.52 (5.45)	2.13 (1.13)	2.14 (1.09)	2.21 (1.11)	3.01 (1.07)
Total		707 (100.00%)	161.53 (25.13)	77.79 (5.16)	2.20 (1.18)	2.13 (1.12)	2.20 (1.14)	3.17 (1.06)

Note. For each continuous measure, the means are reported with the standard deviation in parentheses. Shading was done to improve the readability of the table. Within the acoustic indicators of emotion, pitch was lower on average than the average pitch of the speakers in the RAVDESS dataset (e.g., the mean pitch of female speakers in the RAVDESS dataset was 213.11 and the mean pitch of male speakers in the RAVDESS dataset was 154.72). Average maximum intensity was higher in the sample than in the RAVDESS dataset. Descriptive statistics for the RAVDESS dataset are included in Appendix F. The “not angry” category indicates that the speech recognition model had classified the participant’s average emotion as any emotion other than angry.

Hypothesis 1

I used 8 of the 29 models described above (seven emotion models and the Intercept-Only model) to examine the effect of emotion on influence. The model syntax for the eight models is as follows:

- Model 0: An Intercept Only Model

  $Influence \sim 1 + (1 \mid Jury) + (1 \mid Jury{:}\,RatedJuror)$

- Model 1: Self-Report Anger

  $Influence \sim Self-Report\ Anger + (1 \mid Jury) + (1 \mid Jury{:}\,RatedJuror)$

- Model 2: Self-Report Emotion

  $Influence \sim Self-Report\ Emotion + (1 \mid Jury) + (1 \mid Jury{:}\,RatedJuror)$

- Model 3: Perceived Emotion

  $Influence \sim Perceived\ Emotion + (1 \mid Jury) + (1 \mid Jury{:}\,RatedJuror)$

- Model 4: Expressed Emotion

  $Influence \sim Expressed\ Emotion + (1 \mid Jury) + (1 \mid Jury{:}\,RatedJuror)$

- Model 5: Expressed Anger

  $Influence \sim Expressed\ Anger + (1 \mid Jury) + (1 \mid Jury{:}\,RatedJuror)$

- Model 6: Mean Pitch

  $Influence \sim Mean\ Pitch + (1 \mid Jury) + (1 \mid Jury{:}\,RatedJuror)$

- Model 7: Maximum Intensity

  $Influence \sim Maximum\ Intensity + (1 \mid Jury) + (1 \mid Jury{:}\,RatedJuror)$

The results replicated under four different sets of priors (vague priors, strong priors, weakly informative priors, and priors weakly biased towards the alternative hypothesis—See Appendix G).

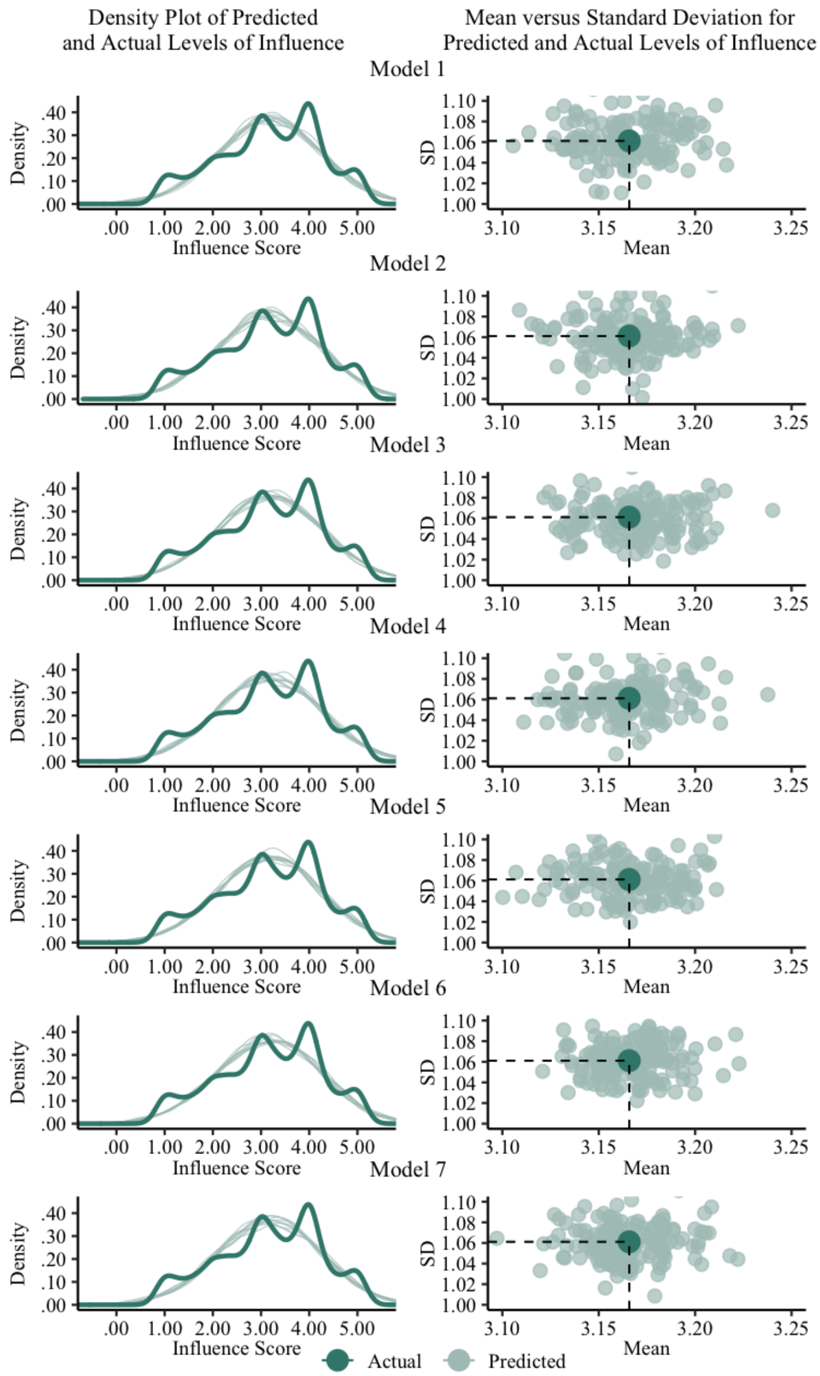
Evaluate the Posterior Predictive Distribution

After running the eight models, I examined whether the models were functioning properly by examining the residual plots and ensuring that all chains mixed (See Appendix H). All of the residuals were slightly light-tailed but overall, reasonably consistent with normality. Trace plots of parameters indicate that all chains were mixed, which indicates that the models converged and were functioning properly.

After determining that the models were functioning properly, I evaluated the posterior predictive distribution for each model. The purpose of posterior predictive checks is to examine whether the model has adequate within-sample prediction (Gelman et al., 2020). That is, I examined whether the models were a good fit for the data. As shown in Figure 9, all models had good within-sample prediction. Because all of the models had good within-sample prediction, I was able to conduct model comparison to determine which model was the most likely, given the data and which model had the best out-of-sample performance.

Figure 9.

Posterior Predictive Checks for All Hypothesis 1 Models



Note. The density plots show the distribution of the actual influence scores (the dark thick line) and the distribution of the ten predicted influence scores. The scatterplots show the mean and standard deviation of actual influence scores (the large, dark dot) and 150 predicted influence scores. When the actual mean and standard deviation of influence fall around the middle of the scatterplot, prediction was relatively consistent with actual influence scores.

Model Comparison

Table 2 shows all pairwise model comparisons between the eight models described above. Perceived emotion was the best predictor of influence, with Model 3 best describing the data, compared to all of the other models (all BF_{10}s > 3.04×10^5). Both self-report measures of emotion were relatively poor predictors of influence, with anecdotal evidence against Model 1 (BF_{10}= 0.50), compared to the Intercept-Only Model, and decisive evidence against Model 2 (BF_{10}= 0.003), compared to the Intercept-Only Model. Two of the four acoustic measures of emotion (Expressed Anger–Model 5 and Maximum Intensity–Model 7) were relatively good predictors of influence, with decisive evidence for Models 5 (BF_{10}= 976.03) and 7 (BF_{10} = 45897.92), compared the Intercept-Only Model. In sum, as shown in Table 2, Models 3 (perceived emotion), 5 (expressed anger), and 7 (maximum intensity) were the most likely, given the data but there was decisive evidence in favor of Model 3 (perceived emotion), compared to all other models.

I confirmed these results using cross-validation. Where Bayes factors measure which model best predicts the data, LOO-CV estimates the out-of-sample prediction. LOO-CV is a method of K-fold cross validation where K is equal to the number of data-points in a set. The dataset is split into a training set, which contains N-1 observations, and a testing set, which contains the final observation. The model is trained on the training set and then used to predict the final observation. I used the loo package (Vehtari et al., 2019), which calculates the log predictive density (lpd), or the likelihood of that the observation in the test set falls within the posterior distribution produced by the training set. Then, all of the lpd values are averaged together to create the estimated log predicted density (elpd), which functions as a measure of how well the model predicts new data. While higher elpd values indicate better predictive accuracy, they are not easily interpretable on their own and they are more informative when used to compare the accuracy of multiple models (Vehtari et al., 2019). Therefore, I examined the change in elpd values across models. When the elpd difference is greater than two standard errors of the difference, the model with the higher elpd is considered have better predictive accuracy (Vehtari et al., 2019).

As shown in Table 3, Model 3 (the model that included perceived emotion) had the best out-of-sample prediction, followed by Model 7 (the model that included maximum intensity). All other models were relatively similar in predictive accuracy to the Intercept-Only model (all elpds were within 2 standard errors of the difference).

Table 2.

Bayes Factors for All Hypothesis 1 Pairwise Comparisons

Denominator	Numerator							
	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Model 0	1.00	0.50	0.003	1.40×10^{10}	0.17	976.03	0.02	45897.92
Model 1	2.02	1.00	0.006	2.82×10^{10}	0.34	1971.73	0.04	92720.68
Model 2	341.22	168.91	1.00	4.76×10^{12}	57.49	3.33×10^{5}	6.69	1.57×10^{7}
Model 3	7.17×10^{-11}	3.55×10^{-11}	2.10×10^{-13}	1.00	1.21×10^{-11}	6.99×10^{-8}	1.41×10^{-12}	3.29×10^{-6}
Model 4	5.93	2.94	0.02	8.28×10^{10}	1.00	5792.60	0.12	2.72×10^{5}
Model 5	0.001	5.07×10^{-5}	3.00×10^{-6}	1.43×10^{7}	1.73×10^{-4}	1.00	2.01×10^{-5}	47.02
Model 6	50.98	25.24	0.15	7.11×10^{11}	8.59	49758.22	1.00	2.34×10^{6}
Model 7	2.18×10^{-5}	1.08×10^{-5}	6.39×10^{-8}	3.04×10^{5}	3.67×10^{-6}	0.02	4.27×10^{-7}	1.00

Note. Values greater than one indicate that there is more support for the numerator model than the denominator model. Values less than one indicate that there is more support for the denominator model than the numerator model. Shading was done to improve the readability of the table. The following are the emotion measures in each model: Model 0: Intercept-Only; Model 1: Self-Report Anger; Model 2: Self-Report Emotion; Model 3: Perceived Emotion; Model 4: Expressed Emotion; Model 5: Expressed Anger; Model 6: Mean Pitch; and Model 7: Maximum Intensity.

Table 3.

Δelpd and Δsd for All Hypothesis 1 Pairwise Comparisons

Comparison	Reference						
	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Model 0	0						
Model 1	3.31 (4.23)	0					
Model 2	0.31 (1.86)	-3.00 (3.88)	0				
Model 3	29.09 (17.10)	25.78 (17.64)	28.77 (17.21)	0			
Model 4	5.06 (7.29)	1.75 (8.25)	4.75 (7.55)	-24.03 (18.44)	0		
Model 5	4.88 (6.28)	1.57 (7.46)	4.56 (6.53)	-24.21 (18.03)	-0.18 (3.46)	0	
Model 6	-0.33 (1.85)	-3.64 (4.38)	-0.64 (2.30)	-29.42 (17.22)	-5.39 (7.69)	-5.21 (6.61)	0
Model 7	7.74 (7.01)	4.43 (8.02)	7.42 (7.24)	-21.35 (18.51)	2.68 (6.28)	2.86 (6.39)	8.07 (7.36)

Note. The Δelpd is reported with 2×*Δsd* in paratheses. Positive Δelpd values indicate that the comparison group has better predictive accuracy than the reference group. Negative Δelpd values indicate that the reference group has better predictive accuracy than the comparison group. If 2×*Δsd* is greater than Δelpd, the models are considered similar in predictive accuracy.

Summarize the Posterior of the Best Model

In both model comparison methods, Model 3 (perceived emotion) was the best model. However, given that perceived emotion and influence were measured simultaneously, it is possible that these measures were related due to an artifact of the study design. Therefore, because Model 7 was also well-supported by the data, I examined whether both Models 3 and 7 supported the hypothesis. Unlike NHST, Bayesian methods do not rely on p-values to draw conclusions about whether to accept or reject the null hypothesis. Rather, Bayesian credible intervals can be used to understand the certainty surrounding an effect. If a credible interval falls entirely above zero, that means that we are 95% certain that there is a positive effect. When a credible interval does not cross zero, we can conclude that the model provides meaningful support for the hypothesis.

Hypothesis 1 was meaningfully supported in both Model 3 and Model 7. As shown in Figure 10., as participants were perceived as more emotional, they were rated as more influential, b =0.12, $\beta = 0.14$, $SE = 0.02$, 95% CI [0.10, 0.17]. Similarly, as shown in Figure 11., as participants' max intensity increased, they were rated as more influential, b =0.03, $\beta = 0.18$, $SE = 0.03$, 95% CI [0.12, 0.24]. The regression results for all other models are summarized in the Alternative Models section and described in detail in Appendix I.

Figure 10.
Influence by Perceived Emotion

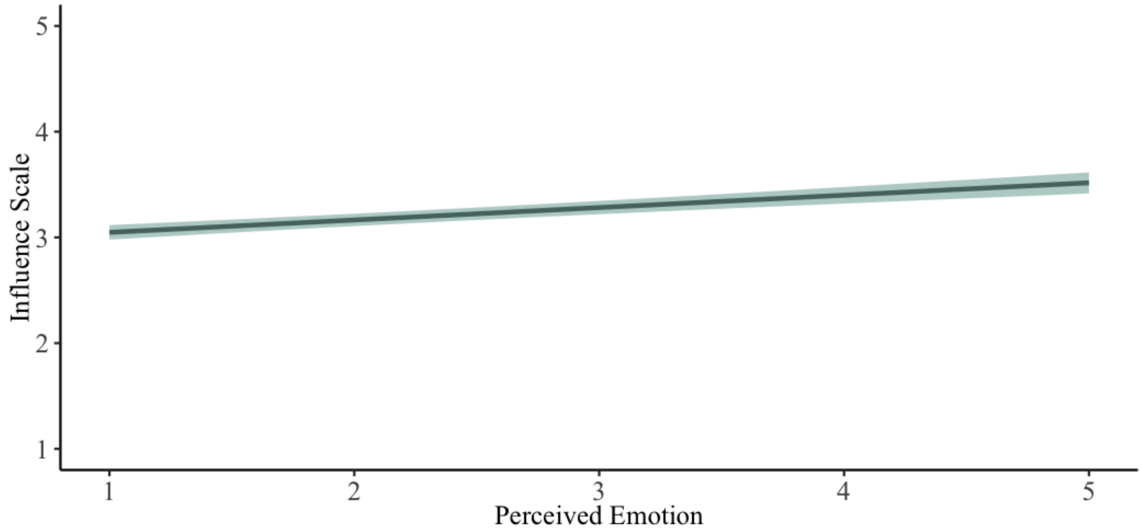
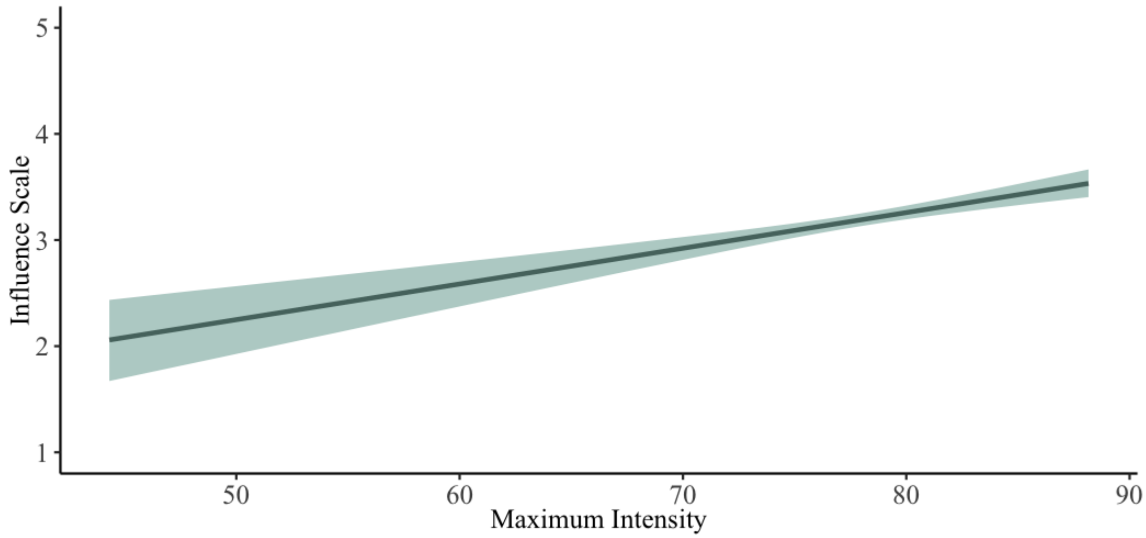


Figure 11.
Influence by Maximum Intensity



Hypothesis 2

I ran 21 additional models (three models per emotion measure) to examine how a juror's race and gender moderated the relationship between emotion and influence. The following are the model syntax for the 21 additional models:

- Model 1A: Self-Report Anger by Gender

  $Influence \sim Self - Report\ Anger * Gender + (1 | Jury) + (1 | Jury: RatedJuror)$

- Model 1B: Self-Report Anger by Race

$$Influence \sim Self\text{–}Report\ Anger * Race + (1 \mid Jury) + (1 \mid Jury{:}\ RatedJuror)$$

- Model 1C: Self-Report Anger by Gender by Race

$$Influence \sim Self\text{–}Report\ Anger * Gender * Race + (1 \mid Jury) + (1 \mid Jury{:}\ RatedJuror)$$

- Model 2A: Self-Report Emotion by Gender

$$Influence \sim Self\text{–}Report\ Emotion * Gender + (1 \mid Jury) + (1 \mid Jury{:}\ RatedJuror)$$

- Model 2B: Self-Report Emotion by Race

$$Influence \sim Self\text{–}Report\ Emotion * Race + (1 \mid Jury) + (1 \mid Jury{:}\ RatedJuror)$$

- Model 2C: Self-Report Emotion by Gender by Race

$$Influence \sim Self\text{–}\ Report\ Emotion * Gender * Race + (1 \mid Jury) + (1 \mid Jury{:}\ RatedJuror)$$

- Model 3A: Perceived Emotion by Gender

$$Influence \sim Perceived\ Emotion * Gender + (1 \mid Jury) + (1 \mid Jury{:}\ RatedJuror)$$

- Model 3B: Perceived Emotion by Race

$$Influence \sim Perceived\ Emotion * Race + (1 \mid Jury) + (1 \mid Jury{:}\ RatedJuror)$$

- Model 3C: Perceived Emotion by Gender by Race

$$Influence \sim Perceived\ Emotion * Gender * Race + (1 \mid Jury) + (1 \mid Jury{:}\ RatedJuror)$$

- Model 4A: Expressed Emotion by Gender

$$Influence \sim Expressed\ Emotion * Gender + (1 \mid Jury) + (1 \mid Jury{:}\ RatedJuror)$$

- Model 4B: Expressed Emotion by Race

$$Influence \sim Expressed\ Emotion * Race + (1 \mid Jury) + (1 \mid Jury{:}\ RatedJuror)$$

- Model 4C: Expressed Emotion by Gender by Race

$$Influence \sim Expressed\ Emotion * Gender * Race + (1 \mid Jury) + (1 \mid Jury{:}\ RatedJuror)$$

- Model 5A: Expressed Anger by Gender

$$Influence \sim Expressed\ Anger * Gender + (1 \mid Jury) + (1 \mid Jury{:}\ RatedJuror)$$

- Model 5B: Expressed Anger by Race

$$Influence \sim Expressed\ Anger * Race + (1 \mid Jury) + (1 \mid Jury{:}\ RatedJuror)$$

- Model 5C: Expressed Anger by Gender by Race

$$Influence \sim Expressed\ Anger * Gender * Race + (1 \mid Jury) + (1 \mid Jury{:}RatedJuror)$$

- Model 6A: Mean Pitch by Gender

$$Influence \sim Mean\ Pitch * Gender + (1 \mid Jury) + (1 \mid Jury{:}RatedJuror)$$

- Model 6B: Mean Pitch by Race

$$Influence \sim Mean\ Pitch * Race + (1 \mid Jury) + (1 \mid Jury{:}RatedJuror)$$

- Model 6C: Mean Pitch by Gender by Race

$$Influence \sim Mean\ Pitch * Gender * Race + (1 \mid Jury) + (1 \mid Jury{:}RatedJuror)$$

- Model 7A: Maximum Influence by Gender

$$Influence \sim Maximum\ Intensity * Gender + (1 \mid Jury) + (1 \mid Jury{:}RatedJuror)$$

- Model 7B: Maximum Influence by Race

$$Influence \sim Maximum\ Intensity * Race + (1 \mid Jury) + (1 \mid Jury{:}RatedJuror)$$

- Model 7C: Maximum Influence by Gender by Race

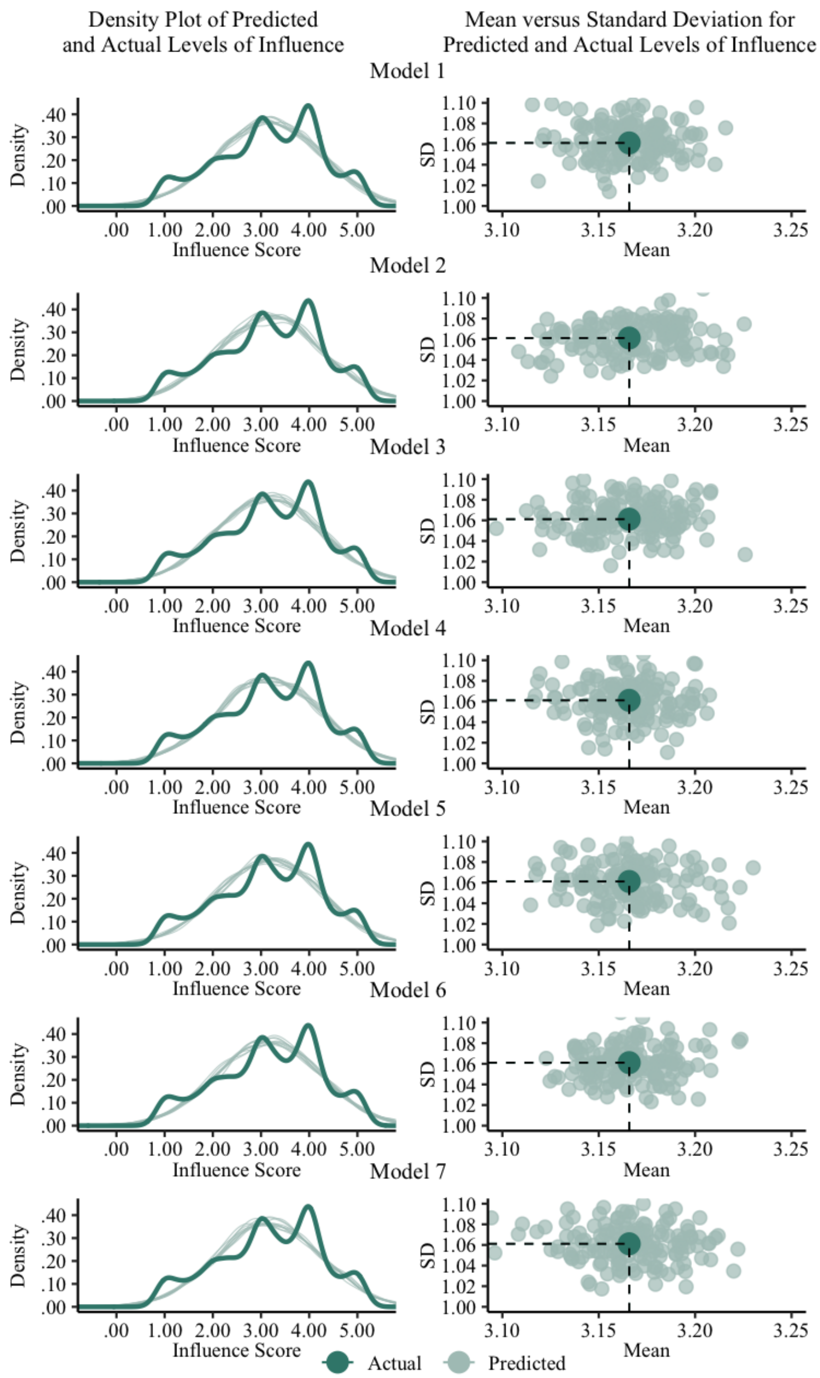
$$Influence \sim Maximum\ Intensity * Gender * Race + (1 \mid Jury) + (1 \mid Jury{:}RatedJuror)$$

Evaluate the Posterior Predictive Distribution

Again, I examined whether the models were functioning properly by examining the residual plots and ensuring that all chains mixed (See Appendix H). All of the residuals were slightly light-tailed but overall, reasonably consistent with normality across the models involving the emotion by gender interactions, the emotion by race interactions, and the 3-way interactions. Trace plots of parameters indicate that all chains were mixed. Then, I evaluated the posterior predictive distribution for each model. As shown in Figures Figure **12**, Figure **13**, and Figure **14**, all models had acceptable within-sample prediction.

Figure 12.

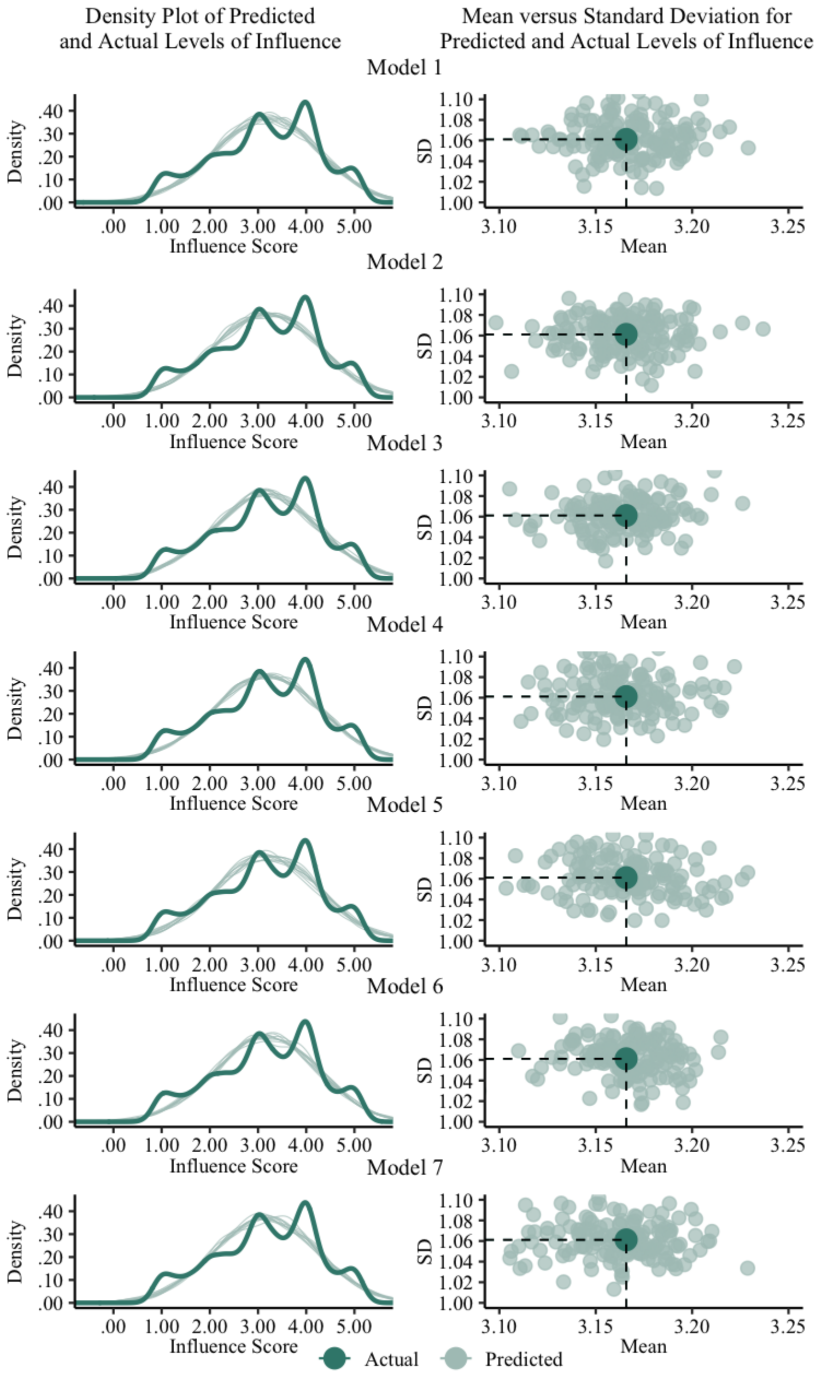
Posterior Predictive Checks for Models that Involve 2-Way Interactions with Gender



Note. The density plots show the distribution of the actual influence scores (the dark thick line) and the distribution of the ten predicted influence scores. The scatterplots show the mean and standard deviation of actual influence scores (the large, dark dot) and 150 predicted influence scores. When the actual mean and standard deviation of influence fall around the middle of the scatterplot, prediction was relatively consistent with actual influence scores.

Figure 13.

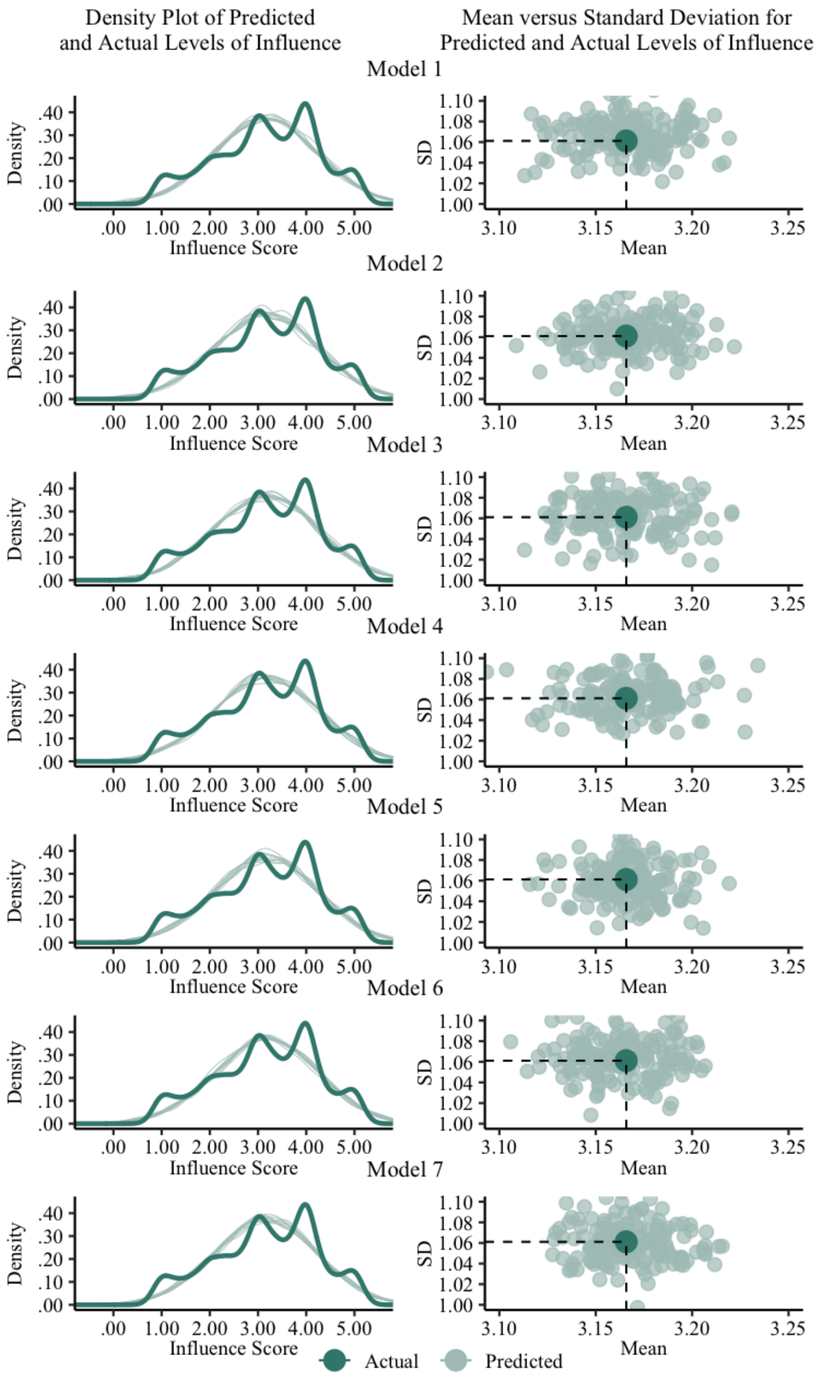
Posterior Predictive Checks for Models that Involve 2-Way Interactions with Race



Note. The density plots show the distribution of the actual influence scores (the dark thick line) and the distribution of the ten predicted influence scores. The scatterplots show the mean and standard deviation of actual influence scores (the large, dark dot) and 150 predicted influence scores. When the actual mean and standard deviation of influence fall around the middle of the scatterplot, prediction was relatively consistent with actual influence scores.

Figure 14.

Posterior Predictive Checks for Models that Involve 3-Way Interactions



Note. The density plots show the distribution of the actual influence scores (the dark thick line) and the distribution of the ten predicted influence scores. The scatterplots show the mean and standard deviation of actual influence scores (the large, dark dot) and 150 predicted influence scores. When the actual mean and standard deviation of influence fall around the middle of the scatterplot, prediction was relatively consistent with actual influence scores.

Model Comparison

Again, I used Bayes factors and leave-one-out cross-validation to determine the best model. I examined the pairwise model comparisons between the 29 models (the 7 emotion-only models, the 21 models that tested moderation, and the Intercept-Only model). Model 3B (perceived emotion by race) best described the data, compared to all of the other models (all $BFs_{10} > 18.82$). Again, both self-report measures of emotion were relatively poor predictors of influence. All models that contained a self-report measure of emotion performed worse than the Intercept-Only Model (Self-Reported Anger: All $BFs_{10} < 0.50$; Self-Reported Emotion: All $BFs_{10} < 0.005$). While models that included expressed anger and maximum intensity generally described the data better than the Intercept-Only Model, the model that included the three-way interaction between expressed anger, race, and gender actually performed worse than the Intercept-Only Model. Table 4 includes all pairwise model comparisons involving perceived emotion, expressed anger, and maximum intensity. All other pairwise comparisons between the 29 models can be found in Appendix I.

Once I identified the models that performed better than the Intercept-Only model, I examined whether the effects were being driven by any specific predictors. Specifically, I examined whether the data was more probable under models that included a specific predictor, compared to models without that predictor (Clyde et al., 2011; Hinne et al., 2020). Inclusion Bayes factors quantify the change from prior inclusion odds (the prior probability that a predictor is included in a model) to the posterior inclusion odds (the probability that a predictor is included in the model, given the data). As shown in Table 5, the data were more probable under models that included perceived emotion ($BF_{inclusion}$

$= 7.89 \times 10^5$), race ($BF_{inclusion}$ = 18.70), and the interaction between perceived emotion and race ($BF_{inclusion}$ = 88.17), compared to models that did not.

Then I examined the change in elpd values across all models to measure out-of-sample prediction. Consistent with hypothesis 1, only models that included perceived emotion and models that included maximum intensity had better out-of-sample prediction than the Intercept-Only model. Table 6 includes the Δelpd values for all pairwise comparisons that included either perceived emotion or maximum intensity.

Again, models that included perceived emotion had better out-of-sample prediction than models that included maximum intensity. However, all of the models that included perceived emotion were relatively similar to each other in out-of-sample prediction (all Δelpd values were within 2 standard errors of the difference).

Table 4.

Bayes Factors for Hypothesis 2 Pairwise Comparisons for all Models that Include Perceived Emotion, Expressed Anger, or Maximum Intensity

Denominator	Numerator												
	Model 0	Model 3	Model 3A	Model 3B	Model 3C	Model 5	Model 5A	Model 5B	Model 5C	Model 7	Model 7A	Model 7B	Model 7C
Model 0	1	1.40×10^{10}	2.43×10^{9}	2.63×10^{11}	1.66×10^{7}	976.03	29.62	117.05	0.01	6.15×10^{5}	63950.28	1.15×10^{5}	77.42
Model 1	2.02	2.82×10^{10}	4.91×10^{9}	5.31×10^{11}	3.35×10^{7}	1971.73	59.83	236.46	0.03	1.24×10^{6}	1.29×10^{5}	2.33×10^{5}	156.39
Model 1A	680.16	9.49×10^{12}	1.65×10^{12}	1.79×10^{14}	1.13×10^{10}	6.64×10^{5}	20144.91	79614.52	9.79	4.19×10^{8}	4.35×10^{7}	7.85×10^{7}	52654.82
Model 1B	25.48	3.56×10^{11}	6.19×10^{10}	6.69×10^{12}	4.22×10^{8}	24867.95	754.63	2982.35	0.37	1.57×10^{7}	1.63×10^{6}	2.94×10^{6}	1972.44
Model 1C	3.62×10^{6}	5.06×10^{16}	8.80×10^{15}	9.52×10^{17}	6.01×10^{13}	3.54×10^{9}	1.07×10^{8}	4.24×10^{8}	52183.29	2.23×10^{12}	2.32×10^{11}	4.18×10^{11}	2.81×10^{8}
Model 2	341.22	4.76×10^{12}	8.29×10^{11}	8.96×10^{13}	5.66×10^{9}	3.33×10^{5}	10106.35	39941.22	4.91	2.10×10^{8}	2.18×10^{7}	3.94×10^{7}	26416.01
Model 2A	17670.9	2.47×10^{14}	4.29×10^{13}	4.64×10^{15}	2.93×10^{11}	1.72×10^{7}	5.23×10^{5}	2.07×10^{6}	254.46	1.09×10^{10}	1.13×10^{9}	2.04×10^{9}	1.37×10^{6}
Model 2B	204.35	2.85×10^{12}	4.96×10^{11}	5.37×10^{13}	3.39×10^{9}	1.99×10^{5}	6052.42	23919.72	2.94	1.26×10^{8}	1.31×10^{7}	2.36×10^{7}	15819.83
Model 2C	1.16×10^{7}	1.62×10^{17}	2.82×10^{16}	3.05×10^{18}	1.93×10^{14}	1.13×10^{10}	3.44×10^{8}	1.36×10^{9}	1.67×10^{5}	7.15×10^{12}	7.43×10^{11}	1.34×10^{12}	8.99×10^{8}
Model 3	7.17×10^{-11}	1.00	0.17	18.82	0.001	6.99×10^{-8}	2.12×10^{-9}	8.39×10^{-9}	1.03×10^{-12}	4.41×10^{-5}	4.58×10^{-6}	8.27×10^{-6}	5.55×10^{-9}
Model 3A	4.12×10^{-10}	5.74	1.00	108.12	0.007	4.02×10^{-7}	1.22×10^{-8}	4.82×10^{-8}	5.93×10^{-12}	0.0003	2.63×10^{-5}	4.75×10^{-5}	3.19×10^{-8}
Model 3B	3.81×10^{-12}	0.05	0.009	1.00	6.31×10^{-5}	3.72×10^{-9}	1.13×10^{-10}	4.46×10^{-10}	5.48×10^{-14}	2.34×10^{-6}	2.43×10^{-7}	4.40×10^{-7}	2.95×10^{-10}
Model 3C	6.03×10^{-8}	841.79	146.54	15843.29	1.00	5.89×10^{-5}	1.79×10^{-6}	7.06×10^{-6}	8.69×10^{-10}	0.04	0.004	0.007	4.67×10^{-6}
Model 4	5.93	8.28×10^{10}	1.44×10^{10}	1.56×10^{12}	9.84×10^{7}	5792.60	175.78	694.69	0.09	3.65×10^{6}	3.80×10^{5}	6.85×10^{5}	459.45
Model 4A	1.22×10^{6}	1.70×10^{16}	2.96×10^{15}	3.20×10^{17}	2.02×10^{13}	1.19×10^{9}	3.61×10^{7}	1.43×10^{8}	17565.02	7.51×10^{11}	7.80×10^{10}	1.41×10^{11}	9.44×10^{7}
Model 4B	2.43×10^{5}	3.39×10^{15}	5.91×10^{14}	6.39×10^{16}	4.03×10^{12}	2.37×10^{8}	7.20×10^{6}	2.85×10^{7}	3501.55	1.50×10^{11}	1.56×10^{10}	2.81×10^{10}	1.88×10^{7}
Model 4C	1.71×10^{16}	2.39×10^{26}	4.16×10^{25}	4.50×10^{27}	2.84×10^{23}	1.67×10^{19}	5.07×10^{17}	2.00×10^{18}	2.46×10^{14}	1.05×10^{22}	1.09×10^{21}	1.98×10^{21}	1.33×10^{18}
Model 5	0.001	1.43×10^{7}	2.49×10^{6}	2.69×10^{8}	16984.11	1.00	0.03	0.12	1.48×10^{-5}	630.51	65.52	118.30	0.08
Model 5A	0.03	4.71×10^{8}	8.20×10^{7}	8.87×10^{9}	5.60×10^{5}	32.95	1.00	3.95	0.0005	20777.77	2159.17	3898.45	2.61
Model 5B	0.009	1.19×10^{8}	2.08×10^{7}	2.24×10^{9}	1.42×10^{5}	8.34	0.25	1.00	0.0001	5257.41	546.34	986.43	0.66
Model 5C	69.44	9.69×10^{11}	1.69×10^{11}	1.82×10^{13}	1.15×10^{9}	67779.47	2056.79	8128.62	1.00	4.27×10^{7}	4.44×10^{6}	8.02×10^{6}	5376.04
Model 6	50.98	7.11×10^{11}	1.24×10^{11}	1.34×10^{13}	8.45×10^{8}	49758.22	1509.93	5967.38	0.73	3.14×10^{7}	3.26×10^{6}	5.89×10^{6}	3946.66
Model 6A	201.47	2.81×10^{12}	4.89×10^{11}	5.29×10^{13}	3.34×10^{9}	1.97×10^{5}	5967.22	23582.97	2.90	1.24×10^{8}	1.29×10^{7}	2.33×10^{7}	15597.12
Model 6B	66.12	9.23×10^{11}	1.61×10^{11}	1.74×10^{13}	1.10×10^{9}	64532.21	1958.25	7739.18	0.95	4.07×10^{7}	4.23×10^{6}	7.63×10^{6}	5118.48
Model 6C	11900.5	1.66×10^{14}	2.89×10^{13}	3.13×10^{15}	1.97×10^{11}	1.16×10^{7}	3.52×10^{5}	1.39×10^{6}	171.37	7.32×10^{9}	7.61×10^{8}	1.37×10^{9}	9.21×10^{5}
Model 7	1.62×10^{-6}	22675.36	3947.30	4.27×10^{5}	26.94	0.002	4.81×10^{-5}	0.0002	2.34×10^{-8}	1.00	0.10	0.19	0.0001
Model 7A	1.56×10^{-5}	2.18×10^{5}	37985.08	4.11×10^{6}	259.22	0.02	0.0005	0.002	2.25×10^{-7}	9.62	1.00	1.81	0.001
Model 7B	8.66×10^{-6}	1.21×10^{5}	21038.13	2.27×10^{6}	143.57	0.008	0.0003	0.001	1.25×10^{-7}	5.33	0.55	1.00	0.0007
Model 7C	0.01	1.80×10^{8}	3.14×10^{7}	3.39×10^{9}	2.14×10^{5}	12.61	0.38	1.51	0.0002	7949.25	826.06	1491.49	1.00

Note. Values that are greater than one indicate that there is more support for the numerator model than the denominator model. Values less than one indicate that there is more support for the denominator model than the numerator model. Shading was done to improve the readability of the table.

Table 5.

Inclusion Bayes Factors for all Models that Include Perceived Emotion, Expressed Anger, or Maximum Intensity

Predictor	Prior Inclusion Probability	Posterior Inclusion Probability	$BF_{Inclusion}$
Perceived Emotion	0.31	1.00	7.89×10^{5}
Gender	0.46	0.009	0.01
Perceived Emotion × Gender	0.15	0.009	0.05
Race	0.46	0.94	18.70
Perceived Emotion × Race	0.15	0.94	88.17
Gender × Race	0.23	5.94×10^{-5}	0.0002
Perceived Emotion × Gender × Race	0.08	5.94×10^{-5}	0.0007
Expressed Anger	0.31	4.02×10^{-9}	9.05×10^{-9}
Expressed Anger × Gender	0.15	1.06×10^{-10}	5.84×10^{-10}
Expressed Anger × Race	0.15	4.20×10^{-10}	2.31×10^{-9}
Expressed Anger × Gender × Race	0.08	5.16×10^{-14}	6.19×10^{-13}
Maximum Intensity	0.31	2.85×10^{-6}	6.41×10^{-6}
Maximum Intensity × Gender	0.15	2.29×10^{-7}	1.26×10^{-6}
Maximum Intensity × Race	0.15	4.14×10^{-7}	2.28×10^{-6}
Maximum Intensity × Gender × Race	0.08	2.77×10^{-10}	3.33×10^{-9}

Note. Prior inclusion probability is the probability that a model includes a predictor. For example, because 4 out of the 13 models include the term “Perceived Emotion”, the prior inclusion probability of Perceived Emotion is .31. Posterior inclusion probability is the probability of the parameter, given the data. The Inclusion Bayes Factor is the change from the prior to the posterior inclusion odds. Shading was done to improve the readability of the table.

Table 6.

Δelpd and Δsd for Hypothesis 2 Pairwise Comparisons for all Models that Include Perceived Emotion, Expressed Anger, or Maximum Intensity

Comparison	Reference								
	Model 0	Model 3	Model 3A	Model 3B	Model 3C	Model 7	Model 7A	Model 7B	Model 7C
Model 0	0	-29.09 (17.10)	-29.81 (17.77)	-34.87 (19.29)	-34.43 (20.13)	-7.74 (7.01)	-7.97 (7.56)	-8.14 (7.83)	-7.69 (8.36)
Model 1	3.31 (4.23)	-25.78 (17.64)	-26.50 (18.17)	-31.56 (19.87)	-31.12 (20.57)	-4.43 (8.02)	-4.66 (8.32)	-4.83 (8.87)	-4.38 (9.15)
Model 1A	2.13 (4.83)	-26.96 (18.00)	-27.68 (18.06)	-32.74 (20.22)	-32.30 (20.48)	-5.61 (8.27)	-5.84 (7.85)	-6.01 (9.13)	-5.56 (8.72)
Model 1B	3.85 (5.74)	-25.24 (17.89)	-25.96 (18.44)	-31.02 (19.29)	-30.58 (20.03)	-3.89 (8.74)	-4.12 (9.08)	-4.29 (7.99)	-3.84 (8.34)
Model 1C	3.85 (6.59)	-25.24 (18.44)	-25.96 (18.51)	-31.02 (19.89)	-30.59 (20.04)	-3.89 (9.14)	-4.12 (8.78)	-4.29 (8.42)	-3.84 (7.77)
Model 2	0.31 (1.86)	-28.77 (17.21)	-29.50 (17.84)	-34.56 (19.41)	-34.12 (20.19)	-7.42 (7.24)	-7.65 (7.69)	-7.83 (8.06)	-7.38 (8.48)
Model 2A	0.77 (3.74)	-28.32 (17.71)	-29.04 (17.78)	-34.10 (19.95)	-33.66 (20.20)	-6.97 (7.61)	-7.20 (7.15)	-7.37 (8.51)	-6.92 (8.07)
Model 2B	1.93 (5.01)	-27.15 (17.73)	-27.88 (18.38)	-32.94 (19.08)	-32.50 (19.95)	-5.80 (8.39)	-6.03 (8.88)	-6.21 (7.56)	-5.76 (8.15)
Model 2C	2.14 (6.23)	-26.94 (18.34)	-27.67 (18.43)	-32.73 (19.77)	-32.29 (19.97)	-5.59 (8.79)	-5.82 (8.45)	-6.00 (8.05)	-5.55 (7.49)
Model 3	29.09 (17.10)	0	-0.72 (4.09)	-5.78 (8.47)	-5.34 (10.10)	21.35 (18.51)	21.12 (18.93)	20.95 (18.71)	21.40 (19.14)
Model 3A	29.81 (17.77)	0.72 (4.09)	0	-5.06 (9.48)	-4.62 (9.29)	22.07 (19.03)	21.84 (18.94)	21.67 (19.25)	22.12 (19.15)
Model 3B	34.87 (19.29)	5.78 (8.47)	5.06 (9.48)	0	0.44 (5.41)	27.13 (20.47)	26.90 (20.90)	26.73 (19.99)	27.18 (20.46)
Model 3C	34.43 (20.13)	5.34 (10.10)	4.62 (9.29)	-0.44 (5.41)	0	26.70 (21.17)	26.46 (21.12)	26.29 (20.73)	26.74 (20.60)
Model 4	5.06 (7.29)	-24.03 (18.44)	-24.75 (18.88)	-29.81 (20.42)	-29.37 (21.03)	-2.68 (6.28)	-2.91 (6.60)	-3.08 (7.12)	-2.63 (7.50)
Model 4A	4.65 (8.36)	-24.44 (19.02)	-25.16 (19.05)	-30.22 (21.06)	-29.78 (21.30)	-3.09 (7.43)	-3.32 (6.94)	-3.49 (8.30)	-3.04 (7.94)
Model 4B	5.52 (8.57)	-23.57 (18.79)	-24.29 (19.27)	-29.35 (20.21)	-28.91 (20.87)	-2.22 (7.63)	-2.45 (7.97)	-2.62 (6.98)	-2.17 (7.38)
Model 4C	5.53 (10.03)	-23.56 (19.57)	-24.28 (19.61)	-29.34 (20.95)	-28.90 (21.15)	-2.20 (9.08)	-2.44 (8.70)	-2.61 (8.49)	-2.16 (7.97)
Model 5	4.88 (6.28)	-24.21 (18.03)	-24.93 (18.62)	-29.99 (20.04)	-29.55 (20.80)	-2.86 (6.39)	-3.09 (6.97)	-3.26 (7.17)	-2.81 (7.76)
Model 5A	4.87 (6.87)	-24.22 (18.48)	-24.94 (18.52)	0	-29.57 (20.74)	-2.87 (6.82)	-3.10 (6.34)	-3.27 (7.63)	-2.82 (7.21)
Model 5B	4.81 (7.25)	-24.28 (18.30)	-25.00 (18.90)	-30.06 (19.66)	-29.62 (20.44)	-2.93 (7.27)	-3.16 (7.83)	-3.33 (6.46)	-2.88 (7.11)
Model 5C	5.42 (7.82)	-23.66 (18.77)	-24.39 (18.82)	-29.45 (20.17)	-29.01 (20.32)	-2.31 (7.74)	-2.54 (7.32)	-2.72 (6.98)	-2.27 (6.33)
Model 6	-0.33 (1.85)	-29.42 (17.22)	-30.14 (17.73)	-35.20 (19.42)	-34.76 (20.11)	-8.07 (7.36)	-8.30 (7.60)	-8.47 (8.16)	-8.02 (8.41)
Model 6A	0.85 (3.57)	-28.23 (17.66)	-28.95 (17.74)	-34.02 (19.81)	-33.58 (20.10)	-6.88 (7.38)	-7.11 (6.92)	-7.28 (8.20)	-6.84 (7.77)
Model 6B	1.01 (4.33)	-28.08 (17.53)	-28.80 (18.04)	-33.86 (18.93)	-33.42 (19.61)	-6.73 (8.17)	-6.96 (8.43)	-7.13 (7.27)	-6.68 (7.47)
Model 6C	1.95 (5.66)	-27.14 (18.07)	-27.86 (18.15)	-32.92 (19.44)	-32.48 (19.66)	-5.79 (8.31)	-6.02 (7.92)	-6.19 (7.41)	-5.74 (6.76)
Model 7	7.74 (7.01)	-21.35 (18.51)	-22.07 (19.03)	-27.13 (20.47)	-26.70 (21.17)	0	-0.23 (2.64)	-0.40 (3.47)	0.05 (4.55)
Model 7A	7.97 (7.56)	-21.12 (18.93)	-21.84 (18.94)	-26.90 (20.90)	-26.46 (21.12)	0.23 (2.64)	0	-0.17 (4.44)	0.28 (3.75)
Model 7B	8.14 (7.83)	-20.95 (18.71)	-21.67 (19.25)	-26.73 (19.99)	-26.29 (20.73)	0.40 (3.47)	0.17 (4.44)	0	0.45 (3.01)
Model 7C	7.69 (8.36)	-21.40 (19.14)	-22.12 (19.15)	-27.18 (20.46)	-26.74 (20.60)	-0.05 (4.55)	-0.28 (3.75)	-0.45 (3.01)	0

Note. The Δelpd is reported with 2×*Δsd* in paratheses. Positive Δelpd values indicate that the comparison group has better predictive accuracy than the reference group. Negative Δelpd values indicate that the reference group has better predictive accuracy than the comparison group. If 2×*Δsd* is greater than Δelpd, the models are considered similar in predictive accuracy. Shading was done to improve the readability of the table.

Summarize the Posterior of the Best Model

Again, in both model comparison methods, models that included perceived emotion were the best models, relative to models that included other emotions. But again, because perceived emotion and influence were measured simultaneously, I also examined the models that included maximum intensity, which were similarly well-supported by the data. Table 7 shows all regression coefficients for the models that included either perceived emotion or maximum intensity. Figure 15 shows the posterior distribution of each parameter within those models.

Table 7.

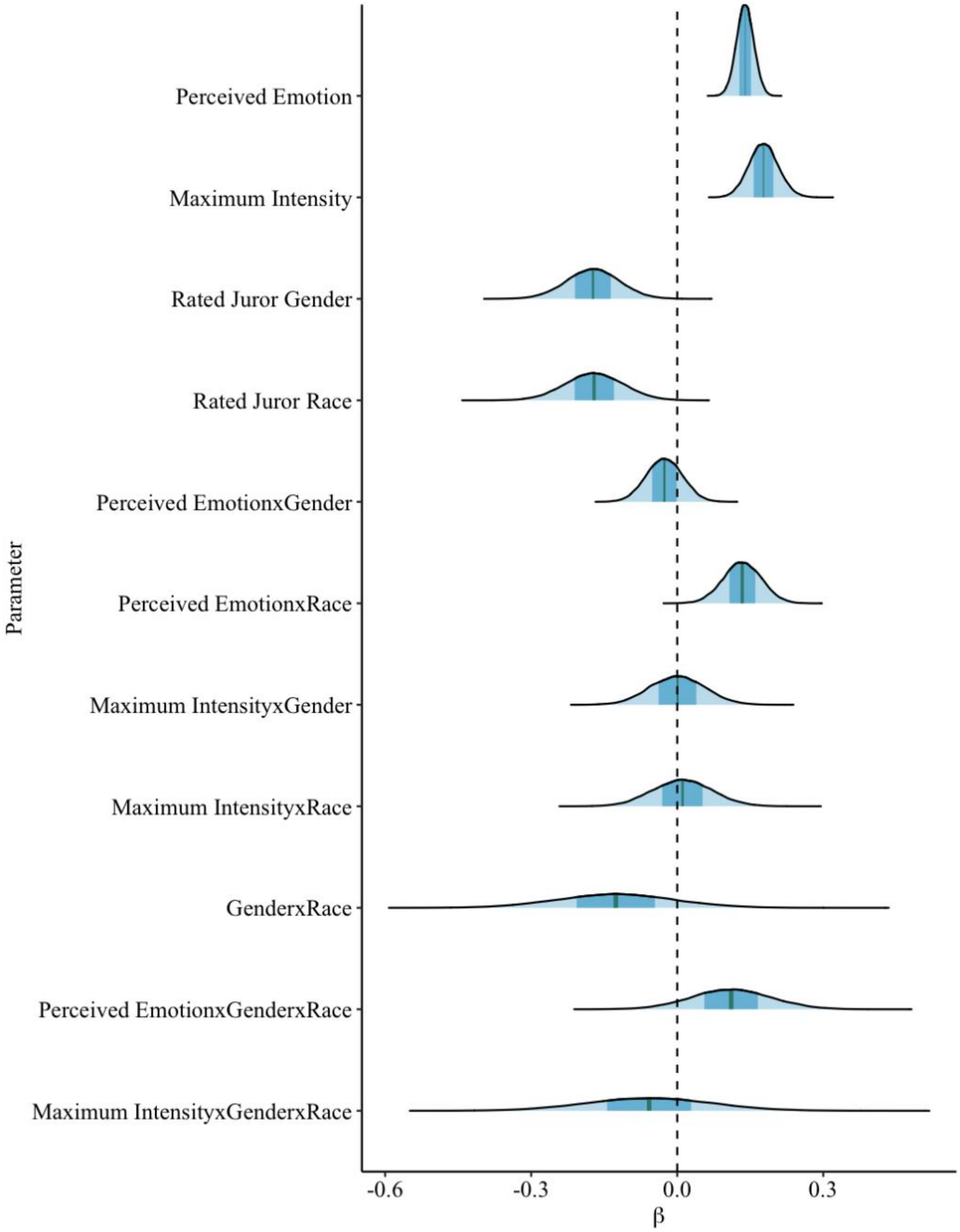
Parameter Estimates for All Parameters in Models Involving Perceived Emotion or Maximum Intensity

Parameter	Perceived Emotion				Maximum Intensity			
	b	β	*SE*	95% *CI*s	*b*	β	*SE*	95% *CI*s
Emotion Measure (A)	**0.12**	**.14**	**.02**	**[.10, .17]**	**0.03**	**.18**	**.03**	**[.12, .24]**
Gender (B)	**0.06**	**-.17**	**.05**	**[-.28, -.07]**	**0.07**	**-.11**	**.05**	**[-.22, -.004]**
Race (C)	**0.21**	**-.17**	**.06**	**[-.29, -.05]**	**0.08**	**-.15**	**.06**	**[-.27, -.04]**
AxB	0.01	-.03	.04	[-.10, .05]	-0.02	.008	.06	[-.11, .11]
AxC	**-0.06**	**.13**	**.04**	**[.06, .21]**	-0.003	.01	.06	[-.11, .13]
AxB	-0.05	-.13	.12	[-.36, .11]	0.02	-.08	.12	[-.32, .15]
AxBxC	0.02	.11	.08	[-.05, .27]	-0.08	-.06	.13	[-.31, .19]

Note. 95% credible intervals that do not cross zero are **bolded.**

Figure 15.

Posterior Distributions of Influence



Note. Light shading represent 95% credible intervals and dark shading represents 50% credible intervals. The thick middle line represents the median beta value. A value of 0 (the dotted line) on the x-axis means that there is no effect of that predictor. Each bell curve represents the distribution of parameter estimates for that parameter. As these are credible intervals, values that are closer to the median are more probable. Wider bell curves represent more uncertainty in the estimate.

Hypothesis 2a: The Interaction of Emotion and Gender. As shown in Table 7 and Figure 15, contrary to the hypothesis, there were no meaningful interactions between perceived emotion and gender or maximum intensity and gender.

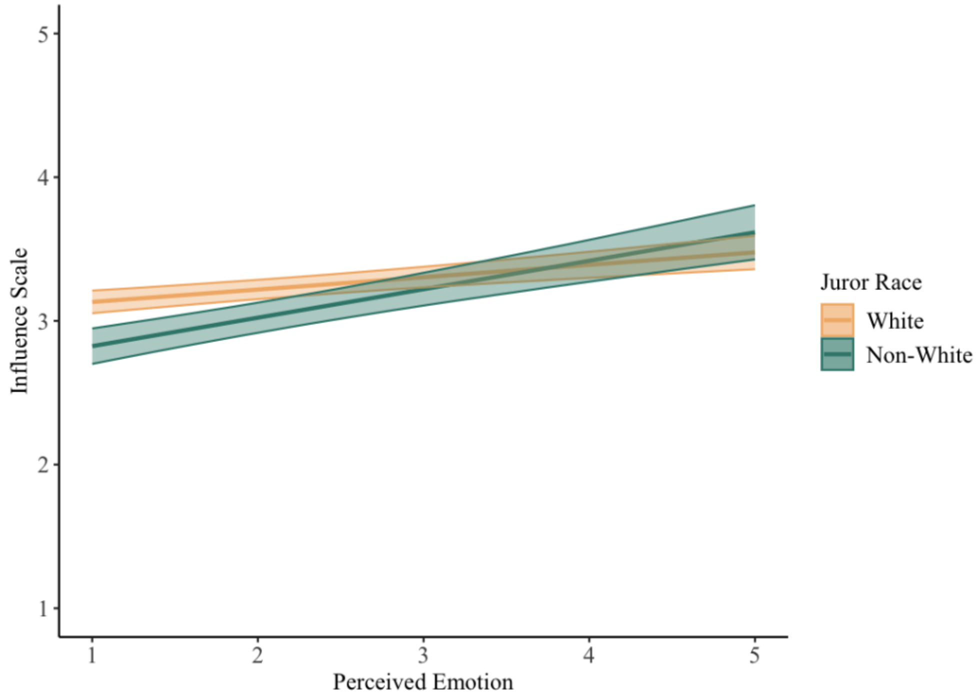
Main Effect of Gender. As shown in Figure 15, there were meaningful main effects of gender, such that men were seen as more influential (M = 3.24, SD = 1.01) than women (M = 3.11, SD = 1.09).

Hypothesis 2b: The Interaction of Emotion and Race. As shown in Table 7 and Figure 15, there was a meaningful interaction between perceived emotion and race. However, the pattern of the interaction was opposite of the hypothesis. As shown in Figure 16, both White and non-White mock jurors were perceived as more influential when as they expressed more emotion. However, the effect was stronger for non-White mock jurors, β = .24, 95% CI [.17, .30], than for White mock jurors, β = .10, 95% CI [.06, .14]. There was not a meaningful interaction between maximum intensity and race (Figure 17).

Main Effect of Race. As shown in Figure 15, there was also a main effect of race such that White jurors were more influential (M = 3.21, SD = 1.04) than non-White jurors (M = 3.03, SD = 1.10).

Figure 16.

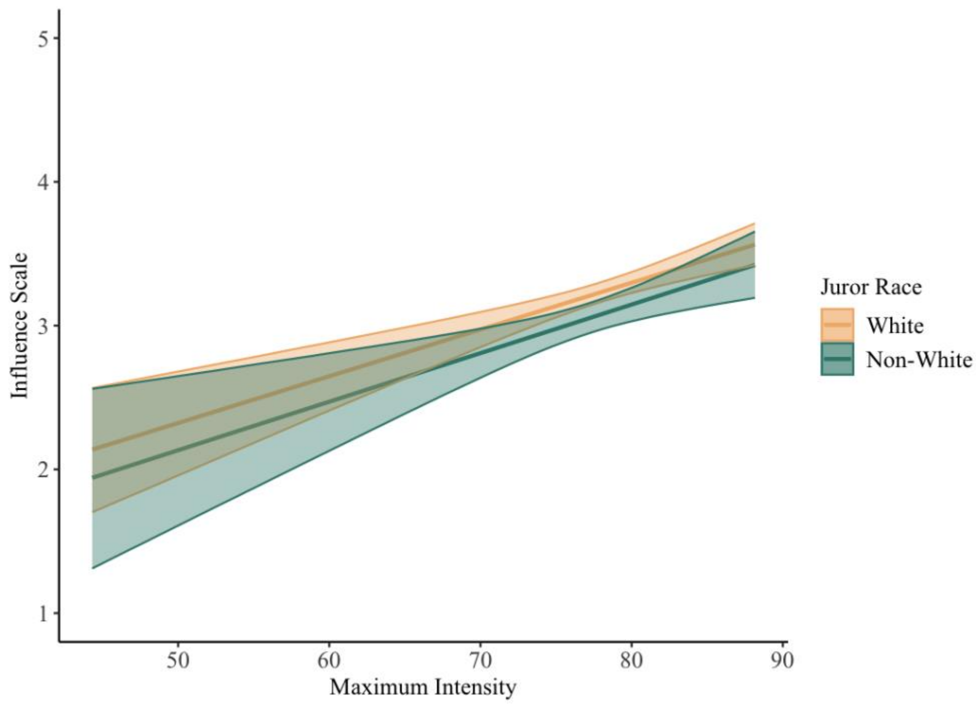
Influence Scale by Perceived Emotion and Rated Juror Race



Note. Shaded areas represent 95% credible intervals.

Figure 17.

Influence Scale by Maximum Intensity and Rated Juror Race



Note. Shaded areas represent 95% credible intervals.

Hypothesis 2c: The Interaction of Emotion, Gender, and Race. As shown in Table 7 and Figure 15, contrary to the hypothesis, there were no 3-way interactions between perceived emotion, gender, and race or maximum intensity, gender, and race. There were also no 2-way interaction between gender and race.

In sum, consistent with hypothesis 1, when jurors were perceived as more emotional and had a higher maximum intensity, they were seen as more influential. Additionally, White and male mock jurors were more influential than non-White and female mock jurors. Finally, there was an interaction between perceived emotion and race. However, the effect was in the opposite direction of hypothesis 2. At low and moderate levels of perceived emotion, White jurors were seen as more influential than non-White jurors. However, at high levels of perceived emotion, there were no race-based differences in perceived emotion.

Alternative Models

I've reported the results for the models that did not perform better than the Intercept-Only model in Appendix I. The model comparison approach to hypothesis testing (i.e., evaluating which model is most likely, given the data) that I have taken thus far dictates that I should conclude that there is not meaningful evidence for these models, given the data. In other words, a model comparison approach to hypothesis testing suggests that there is not meaningful support for the hypotheses that self-reported anger, self-reported emotion, expressed emotion, or mean pitch predict influence (all BF_{10}s < 1, all Δelpds were either within 2 standard errors of the difference or were negative). And there is only partial support for the hypothesis that expressing anger (compared to other emotions) predicts influence, because while the Bayes factor for that model, compared to the Intercept-Only Model, suggested that there was decisive evidence of that model (BF_{10}= 976.03), the predictive accuracy was low. However, following Box's admonition that "All models are wrong, but some are useful", it can also be helpful to examine these models to develop questions for future research about when the patterns observed here might be more pronounced or predictive of behavior.

Broadly speaking, increased self-reported anger predicted a decrease in influence; jurors who were classified as angry were rated as more influential than jurors who were classified as calm, disgusted, or surprised and when I compared jurors who were classified as angry to jurors who were classified as expressing any other emotion, angry jurors were more influential than jurors who expressed any other emotion. Consistent with hypothesis 2, perceived influence increased as White jurors rated themselves as more emotional but decreased as non-White jurors rated themselves as more emotional. Additionally, non-White jurors who were classified as fearful, were more influential than White jurors who were classified as fearful. All other credible intervals crossed zero, suggesting that all other hypotheses were not supported from both a model comparison perspective or a parameter estimation perspective.

I also conducted several exploratory models to examine the impact of three exploratory moderators (case evidence, jury instructions, and deliberation modality). The results of these exploratory models are reported in full in Appendix J. Because Model 3B (Perceived Emotion X Race) was the best performing model, I began by using that model as the basis for the exploratory analyses. However, because deliberation modality predicted perceived emotion (only participants were perceived as less emotional than in-person participants), I used Model 7B (Maximum Intensity X Race) as the basis for the third exploratory model. All of the models that included the exploratory predictors performed poorly, relative to models that did not include the exploratory predictors. All credible intervals for the exploratory three-way interactions crossed zero, suggesting that the exploratory moderators were not supported from both a model comparison perspective or a parameter estimation perspective.

Discussion

Jury deliberation is often considered a black box in the legal system. Deliberations are conducted in secret and, as a result, we know little about when and how jurors exert influence during deliberation. When experimental research attempts to investigate and understand the dynamics of deliberation, the research is often conducted within a tightly controlled environment and researchers have to extrapolate about the jury decision-making process from individual juror decisions. This research adds to the experimental literature about influence in jury deliberations by examining predictors of influence within mock jury deliberations. I found that, consistent with prior research, perceived emotionality and maximum intensity during deliberation predicted influence such that as mock jurors were seen as more emotional and as their intensity during deliberation increased, other jurors found them to be more influential. Additionally, men and White people were more influential on the jury than women and non-White people. However, contrary to the hypothesis, being perceived as more emotional actually predicted increased influence for jurors of color. Specifically, being perceived as more emotional predicted an increase in influence for both White and non-White mock jurors, but this increase in influence was stronger for non-White mock jurors. These findings have important implications both for psychological theory on the relationship between emotions and influence and for jury researchers who are seeking to understand more about jury deliberation.

Theoretical Contribution

This research adds to our understanding of how and when emotional expression impacts influence. Specifically, this research conceptually replicates research that suggests that expressing more emotion can increase influence (e.g., Steinel et al., 2008; Walter et al., 2019) in a more naturalistic setting. That is, rather than experimentally manipulating emotion like prior research, this research demonstrates that those findings can generalize to situations where people are naturally expressing emotion. This research also conceptually replicates research that suggests that men and White people are seen as more influential than women and non-White people.

Contrary to prior research and the hypothesis, being perceived as more emotional did not exacerbate racial differences in influence. In contrast, while being perceived as more emotional increased influence for all jurors, increased perceived emotionality was actually more beneficial for non-White mock jurors than it was for White mock jurors. This finding might be due to the emotionally evocative nature of the case. In other words, in cases like this where the external cause of emotion is clear, it is possible that people are less likely to rely on stereotypes. This is consistent with stereotyping literature, which suggests that people rely on their stereotypes when they lack other information to explain peoples' behavior (e.g., Allport, 1954). Additionally, it was relatively rare for jurors to be perceived highly emotional (only 17% of White participants and 15% of non-White participants were seen as very or extremely emotional by other jurors). It could be that, because it was very rare that participants were perceived to be highly emotional, those emotional jurors were taken more seriously, regardless of race. If we had had truly high levels of emotion, perhaps women or jurors of color would have been penalized at levels, similar to experiments that presented high emotion conditions.

Interestingly, although there was an interaction between perceived emotion and juror race, there was no interaction between maximum intensity and juror race. Instead, White jurors were always seen as more influential than non-White jurors, regardless of their maximum intensity. This suggests that racial differences in the impact of emotion on influence might not be driven by the expression of emotion but rather by the perceptions of the people who are being influenced.

This research also builds on our understanding of the relationship between emotion and influence by examining several measures of emotion. Consistent with past research (e.g., Mauss & Robinson, 2009), there was little correlation between the measures of emotion. So, in deliberation, mock jurors aren't necessarily expressing the emotions that they report experiencing and their acoustic indicators of expressed emotion do not line up with others' perceptions of their emotion. It is particularly interesting that acoustic indicators of expressed emotion are not correlated with perceived emotion because it begs the question: What do people rely on when forming their opinions about others' emotions if they are not relying on behavioral cues like acoustic indicators of emotion? Probing the relationships between these different measures of emotion is an important avenue for future research.

Further, this research examined which measures of emotion were the most predictive of influence. I found that the models that included perceived emotion and maximum intensity performed best and that the models that included perceived emotion, specifically, were the most probable given the data. This finding raises methodological questions for researchers who attempt to manipulate expressed emotion to investigate the effects of emotion and juror demographics on influence because perceptions of emotion were a far better predictor of emotion than the measures of expressed emotion.

This also might explain why the race effects that I found here are not consistent with experimental research. Where most experimental research manipulates expressed emotion, I found that expressed emotion was less predictive of influence than perceived emotion. While most of the research that experimentally manipulates expressed emotion uses perceived emotion as a manipulation check, one study used perceived emotion as a mediator between expressed emotion and perceived influence (Salerno et al., 2019). In that study, in contrast to what I found here, increased perceived emotion predicted a decrease in perceived influence for Black mock jurors but not for White mock jurors. However, in that study, participants rated the angry juror as substantially more emotional than the participants in this study rated each other, which might suggest that race effects occur at extreme amounts of emotion that might be uncommon in naturally occurring emotional expression. The differences in these two findings begs the question of whether the traditional racial bias we see in experiments examining the different impact of emotion on influence for White and non-White people is due to the use of highly emotional stimuli in experimental tests of the impact of emotion on influence.

Finally, examining the different patterns of influence among the emotion models might provide some information about how different types of emotion impact perceptions of influence. The finding that perceived emotion best predicts perceived influence is not particularly surprising. This is consistent with the Emotion as Social Information theory in that other mock jurors use their perceptions of someone's emotions to draw conclusions about the level of influence that the person should hold.

Further, when the speech-emotion recognition model classified a person as expressing high arousal, negative valence emotions (i.e., anger, fear, and disgust), their fellow mock jurors perceived them as more emotional. This suggests that mock jurors might be judging others as more emotional in this case when they express emotions that are appropriate and expected in the setting. In other words, perhaps a mock juror was perceived as more emotional when they expressed emotions that matched others' expectations for the situation and, because that emotion was considered appropriate to the situation, the other mock jurors rated that mock juror as more influential. This finding would be in line with the research that suggests that the impact of emotion depends on whether that emotion is appropriate (Rose et al., 2006; Van Kleef & Côté, 2007).

The only other emotion measure to consistently perform better than the Intercept-Only model was maximum intensity. Given that research has shown that intensity is associated with an increase in perceptions of confidence (Jiang & Pell, 2017), it makes sense that a higher maximum intensity predicted an increase in influence. That is, it is possible that speaking with increased intensity signifies increased confidence in a position, which in turn increases influence. Interestingly, unlike maximum intensity, mean pitch did not predict influence. Several papers have demonstrated that variations pitch might be indicative of someone's interest and engagement with a topic (Dietrich, et al., 2019a, 2019b) but these papers pay little attention to intensity. Given that this research suggests that 1) mean pitch and maximum intensity are not correlated and 2) that maximum intensity predicts influence, but mean pitch does not, it might be useful for future researchers to consider the two acoustic dimensions together. That is, perhaps mean pitch is predictive of a person's own interest in a topic, but maximum intensity is predictive of the influence that person holds over others.

Finally, neither self-report measures of emotion performed better than the Intercept-Only model. This suggests, broadly, that peoples' own evaluations of both the way that they felt and how emotional they were acting have little influence over others' perceptions of them. And yet, race did moderate the impact of self-reported emotion and influence in the hypothesized direction. While this effect should be interpreted with caution, given that the model performed poorly, relative to the Intercept-Only model, this finding merits further investigation, first as to whether this effect can be replicated and then into explanations for why there might be racial differences in how a person's own perception of their emotion predicts the level of influence that they hold over other people.

Legal Implications

This research examines which jurors might hold the most influence. Even when juries are diverse, juries cannot be truly representative if some jurors hold more influence than others. This research suggests that, overall, women and people of color hold less influence on juries than White men. And while the race differences might lessen when jurors are perceived as more emotional, White jurors were more influential at both low and moderate levels of perceived emotion. Given that very few jurors were perceived as highly emotional, this research suggests that overall, White people remain more influential on juries than people of color. This research provides some insight into what happens when juries are more diverse. Specifically, this research provides some indication that increasing diversity on juries might not be enough because when minoritized people are on those juries, they might not be able to achieve the same level of influence as their White counterparts.

Additionally, this research examines naturally occurring emotion in jury deliberation and the ways in which the expression of that emotion might impact how jurors are perceived. These naturally occurring expressions of emotion closely approximate how a jury might interact and will provide important information on how jurors might deliberate in real trials. Much of jury research focuses on individual juror decisions. Researchers have to make assumptions about how juries will function based on the actions of individual jurors. This research provides rare insight into the deliberation process itself. Trial attorneys and other legal actors are constantly trying to understand how jurors deliberate and which jurors hold influence in deliberation. This research provides those attorneys with additional information about the innerworkings of the deliberation process and the situations in which jurors' emotions might impact their deliberative process.

Strengths, Limitations, and Future Directions

Although mock jury studies can never capture every aspect of a real trial and real deliberation, this study was designed to mimic the process as closely as possible. Participants watched a trial video that used real attorneys and deliberated together just as they would in a real trial. We gave participants as much time as possible to come to a verdict and indeed, only 13 (8.50% of juries) failed to reach a verdict. We also recruited participants from a variety of sources (MTurk, Craigslist, flyers, word of mouth) in order to attempt to recruit a sample that was more diverse and representative of real jury service than a typical MTurk study. However, because we did not strategically recruit or schedule participants based on their race, it was relatively common for the mock juries to have more White participants than non-White participants (78.29% of juries had more White participants than non-White participants and 25.66% of juries were all White). These racial imbalances temper the conclusions that I can make about the impact of juror race on influence because it is possible that the racial composition of the jury might have impacted who held influence on each jury. That being said, it is the unfortunate reality that the majority of jurors are White and in this way, this research might capture the racial dynamics of real juries. Still, future research could examine the impact of the racial composition of the jury on the levels of influence exerted by White and non-White jurors.

Additionally, because the goal of this research was to explore the impact of naturally occurring emotion on influence, I am unable to make firm causal claims. While a strength of this research is that participants were rating each other, and therefore these effects are not due to something unique about a specific person, the correlational nature of the research means that there might be alternative causal explanations for the relationship between emotion and influence. For example, perhaps jurors are seen as more emotional when they participate more in the deliberation and more active jurors are more influential. While there are a number of potential third variables that might explain the relationship between emotion and influence, this research is consistent with experimental research on the impact of emotion on influence (e.g., Steinel et al., 2008). When considered in the context of that experimental research, this research adds additional evidence about the potential causal impact of emotion on influence.

These findings in this research might also be limited by the content of the case. The facts of this case were very emotionally evocative. It is possible that in a less emotionally charged case, emotional expression might be seen as inappropriate, and jurors might actually be penalized if they are seen as extremely emotional. Future research should examine emotions in deliberation in a variety of case types in order to determine the extent to which these findings replicate in cases with a less emotionally evocative fact pattern.

Moreover, the findings relating to expressed emotion, broadly, and expressed anger specifically, might be limited by the use of a machine-learning model. While using machine-learning to classify mock jurors' emotions is intended to provide a measure of expressed emotion that is based solely on the mock jurors' acoustic profiles, rather than any person's perceptions of their emotions, speech-emotion recognition remains very difficult (Magdin et al., 2019). This is in part because emotions are often fluid and can overlap, which makes categorization difficult. Although many researchers are attempting to develop speech-emotion recognition models, there is no clear consensus on which features and classifiers to use to categorize emotions (Koolagudi et al., 2018).

The model that I used had a balanced accuracy of a little over 75%, which is considered excellent model performance but there were important differences between the training and testing data and the study data. First, the audio quality in the RAVDESS dataset was of a higher quality than in the study data. Second, the actors in the RAVDESS dataset were told to speak using specific emotions, which means that, during training and testing, the model learned to detect clear emotions. In other words, the model was not trained on data where emotions fluctuated and overlapped. But across a deliberation, it is possible that a participant's emotional expression was more inconsistent and fluctuated. In this way, the speech-emotion recognition model might not be as accurate in predicting new, more messy and complex data, relative to the cleaner and clearer RAVDESS data.

That being said, inspection of the other emotion measures suggests that the speech-emotion recognition model was properly classifying emotions. First, the emotion classifications are consistent with the pitch and intensity data. When participants were classified as calm, their pitch and intensity were, on average, lower than when they were classified as any other emotion. Similarly, participants that were classified as happy and surprised spoke in the highest pitch and participants that were classified as happy and angry spoke with the most intensity. These findings are consistent with other research on speech-emotion recognition (Goudbeek & Scherer, 2010; Magdin et al., 2019). While the emotion classifications were not consistent with participants' self-reported anger and emotion, other measures of emotion were also not consistent with self-reported anger and emotion. For example, there was almost no correlation between perceived emotion and either self-reported emotion or self-reported anger. Further, other participants rated that participant who were classified as calm as less emotional than participants who were classified as angry, even though the participants who were classified as calm rated themselves as more angry and more emotional than the participants who were classified as angry. While speech-emotion recognition models are by no means perfect and future research should continue to investigate creative ways of measuring expressed emotion, this data suggests that this speech-emotion recognition model was, at the very least, more consistent with perceptions of others' emotion than people's own evaluations of their anger and general emotionality.

Finally, while perceived emotion was the most predictive of influence, perceived emotion and influence were measured simultaneously. Therefore, it is possible that the relationship between perceived emotion and influence were an artifact of the survey design and the correlation between the two being artificially inflated by the methodology. These concerns are somewhat alleviated because an increase in maximum intensity also predicted an increase in influence, but future research could measure perceptions of emotion and perceptions of influence independently.

Conclusion

Jury trials are often highly emotionally charged, and the emotional nature of a jury trial might bleed into deliberation. This research suggests that, indeed, as jurors are perceived to be more emotional and as their maximum intensity during deliberation increases, their fellow jurors rate them as more influential. This finding suggests that expressing emotions might be an important part of the deliberation process and adds to our understanding of how emotions might influence behavior in juries, specifically, and groups in general.

Further, courts across the country are working to increase diversity on juries. However, as juries become more diverse, it is important to understand how that increased diversity changes who holds influence on a jury. This research suggests that White people and men might hold more influence on juries than people of color and women. Importantly, the racial differences in influence might only disappear when the juror is perceived as extremely emotional and gender differences in influence were not impacted by juror emotions. The ideal American jury is a body where diverse jurors are all able to participate equally. However, despite efforts to make juries appear more diverse from the outside, the race and gender inequality on juries might have deeper, more insidious effects in that, even when the jury is diverse, people of color and women might not be able to participate to the same extent as White men.

REFERENCES

Adam, H., & Shirako, A. (2013). Not all anger is created equal: The impact of the expresser's culture on the social effects of anger in negotiations. *Journal of Applied Psychology*, *98*(5), 785–798. https://doi.org/10.1037/a0032387

Alhamid, M. (2022, May 12). *Ensemble models: What are they and when should you use them?*. Built In. https://builtin.com/machine-learning/ensemble-model

Allport, G. W. (1954). *The nature of prejudice.* Addison-Wesley.

Bachorowski, J. A., & Owren, M. J. (1995). Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychological Science*, *6*(4), 219–224. https://doi.org/10.1111/j.1467-9280.1995.tb00596.x

Barkved, K. (2022, March 9). *How to Know If Your Machine Learning Model Has Good Performance*. Obviously.ai. https://www.obviously.ai/post/machine-learning-model-performance

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Barrett, L. F. (1997). The relationships among momentary emotion experiences, personality descriptions, and retrospective ratings of emotion. *Personality and Social Psychology Bulletin*, *23*(10), 1100–1110. https://doi.org/10.1177/01461672972310010

Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and social psychology review*, *10*(1), 20–46. https://doi.org/10.1207/s15327957pspr1001_2

Barrett, L. F., & Bliss-Moreau, E. (2009). She's emotional. He's having a bad day: Attributional explanations for emotion stereotypes. *Emotion*, *9*(5), 649–658. https://doi.org/10.1037/a0016821

Bartz, A. E., Blume, N. E., & Rose, J. (1996). Gender differences in self-report measures of anger: The role of social desirability and negative affect. *Journal of Social Behavior and Personality*, *11*(5), 241. https://www.proquest.com/scholarly-journals/gender-differences-self-report-measures-anger/docview/1292244795/se-2

Bastiaansen, M., Lub, X. D., Mitas, O., Jung, T. H., Ascenção, M. P., Han, D. I., Moilanen, T., Smit, B., & Strijbosch, W. (2019). Emotions as core building blocks of an experience. *International Journal of Contemporary Hospitality Management, 31*(2) 651–668. https://doi.org/10.1108/IJCHM-11-2017-0761

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Berdahl, J. L. (2007). Harassment based on sex: Protecting social status in the context of gender hierarchy. *Academy of Management Review*, *32*(2), 641–658. https://doi.org/10.5465/amr.2007.24351879

Boersma, P., & Weenink, D. (2022). Praat: doing phonetics by computer [Computer program]. Version 6.3.03, retrieved January 24, 2023, from http://www.praat.org/

Brescoll, V. L., & Uhlmann, E. L. (2008). Can an angry woman get ahead? Status conferral, gender, and expression of emotion in the workplace. *Psychological Science*, *19*(3), 268–275. https://doi.org/10.1111/j.1467-9280.2008.02079.x

Bright, D. A., & Goodman-Delahunty, J. (2006). Gruesome evidence and emotion: anger, blame, and jury decision-making. *Law and Human Behavior*, *30*(2), 183–202. https://doi.org/10.1007/s10979-006-9027-y

Bürkner, P.C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software, 80*(1), 1-28. doi:10.18637/jss.v080.i01

Clyde, M. A., Ghosh, J., & Littman, M. L. (2011). Bayesian adaptive sampling for variable selection and model averaging. *Journal of Computational and Graphical Statistics*, *20*(1), 80–101. https://doi.org/10.1198/jcgs.2010.09049

Corley, C. (2021, September 6). *Arizona's Supreme Court eliminates peremptory challenges*. NPR. Retrieved January 24, 2023, from https://www.npr.org/2021/09/06/1034556234/arizonas-supreme-court-eliminates-peremptory-challenges

Cornwell, E. Y., & Hans, V. P. (2011). Representation through participation: A multilevel analysis of jury deliberations. *Law & Society Review*, *45*(3), 667-698. https://doi.org/10.1111/j.1540-5893.2011.00447.x

Cullen, A. C., & Frey, C. H. (1999). *Probabilistic techniques in exposure assessment: a handbook for dealing with variability and uncertainty in models and inputs*. Springer Science & Business.

DeSteno, D., Petty, R. E., Rucker, D. D., Wegener, D. T., & Braverman, J. (2004). Discrete emotions and persuasion: the role of emotion-induced expectancies. *Journal of Personality and Social Psychology*, *86*(1), 43–56. https://doi.org/10.1037/0022-3514.86.1.43

Devillers, L., Vidrascu, L., & Lamel, L. (2005). Challenges in real-life emotion annotation and machine learning based detection. *Neural Networks*, *18*(4), 407–422. https://doi.org/10.1016/j.neunet.2005.03.007

Dietrich, B. J., Enos, R. D., & Sen, M. (2019a). Emotional arousal predicts voting on the U.S Supreme Court. *Political Analysis*, *27*(2), 237–243. doi:10.1017/pan.2018.47

Dietrich, B. J., Hayes, M., & O'brien, D. Z. (2019b). Pitch perfect: Vocal pitch and the emotional intensity of congressional speech. *American Political Science Review*, *113*(4), 941–962. doi:10.1017/S0003055419000467

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3–4), 169–200.

Elfenbein, H. A., & Ambady, N. (2002). Is there an in-group advantage in emotion recognition? *Psychological Bulletin, 128*(2), 243–249. https://doi.org/10.1037/0033-2909.128.2.243

Fauville, G., Luo, M., Queiroz, A. C. M., Lee, A., Bailenson, J. N., & Hancock, J. (2023). Video-conferencing usage dynamics and nonverbal mechanisms exacerbate Zoom Fatigue, particularly for women. *Computers in Human Behavior Reports*, *10*, 100271. https://doi.org/10.1016/j.chbr.2023.100271

Feinberg, D. R. (2022, January 1). Parselmouth Praat Scripts in Python. https://doi.org/10.17605/OSF.IO/6DWR3

Freund, Y., & Schapire, R. E. (1996, July). Experiments with a new boosting algorithm. *International Conference on Machine Learning*, *96*, 148–156. http://dl.acm.org/citation.cfm?id=3091696.3091715

Gelman, A. (2020, April 17). *Prior choice recommendations*. GitHub. https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations

Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and other stories*. Cambridge University Press.

George, D., & Mallery, P. (2010). *SPSS for Windows step by step. A simple study guide and reference* (10th ed.). Pearson Education, Inc.

Gibson, D. E., Schweitzer, M. E., Callister, R. R., & Gray, B. (2009). The influence of anger expressions on outcomes in organizations. *Negotiation and Conflict Management Research*, *2*(3), 236–262. https://doi.org/10.1111/j.1750-4716.2009.00039.x

Gill, J., & Witko, C. (2013). Bayesian analytical methods: A methodological prescription for public administration. *Journal of Public Administration Research and Theory*, *23*(2), 457–494. https://doi.org/10.1093/jopart/mus091

Goudbeek, M., Goldman, J. P., & Scherer, K. R. (2009). Emotion dimensions and formant position. In *Tenth Annual Conference of the International Speech Communication Association*. https://doi.org/10.21437/interspeech.2009-469

Goudbeek, M., & Scherer, K. (2010). Beyond arousal: Valence and potency/control cues in the vocal expression of emotion. *The Journal of the Acoustical Society of America*, *128*(3), 1322–1336. https://doi.org/10.1121/1.3466853

Griskevicius, V., Shiota, M. N., & Neufeld, S. L. (2010). Influence of different positive emotions on persuasion processing: a functional evolutionary approach. *Emotion*, *10*(2), 190–206. https://doi.org/10.1037/a0018421

Hickerson, A., & Gastil, J. (2008). Assessing the difference critique of deliberation: Gender, emotion, and the jury experience. *Communication Theory*, *18*(2), 281–303. https://doi.org/10.1111/j.1468-2885.2008.00323.x

Hinne, M., Gronau, Q. F., van den Bergh, D., & Wagenmakers, E. J. (2020). A conceptual introduction to Bayesian model averaging. *Advances in Methods and Practices in Psychological Science*, *3*(2), 200–215. https://doi.org/10.1177/2515245919898657

The impact of covid-19 on Workplace Inclusion: Survey. Catalyst. (2022, April 19). https://www.catalyst.org/research/workplace-inclusion-covid-19/

Jeffreys, H. (1961). *Theory of probability*. Oxford: UK Oxford University Press.

Jiang, X., & Pell, M. D. (2017). The sound of confidence and doubt. *Speech Communication*, *88*, 106–126. https://doi.org/10.1016/j.specom.2017.01.011

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition & Emotion*, *13*(5), 505–521. https://doi.org/10.1080/026999399379168

Koolagudi, S. G., Murthy, Y. S., & Bhaskar, S. P. (2018). Choice of a classifier, based on properties of a dataset: case study-speech emotion recognition. *International Journal of Speech Technology*, *21*(1), 167–183. https://doi.org/10.1007/s10772-018-9495-8

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). Elsevier.

Lewis, K. M. (2000). When leaders display emotion: How followers respond to negative emotional expression of male and female leaders. *Journal of Organizational Behavior*, *21*(2), 221–234. https://doi.org/10.1002/(SICI)1099-1379(200003)21:2<221::AID-JOB36>3.0.CO;2-0

Lüdtke, O., Robitzsch, A., Kenny, D. A., & Trautwein, U. (2013). A general and flexible approach to estimating the social relations model using Bayesian methods. *Psychological Methods*, *18*(1), 101–119. https://doi.org/10.1037/a0029252

Lynch, M., & Haney, C. (2015). Emotion, Authority, and Death: (Raced) Negotiations in Mock Capital Jury Deliberations. *Law & Social Inquiry, 40*(2), 377–405. doi:10.1111/lsi.12099

Magdin, M., Sulka, T., Tomanová, J., & Vozár, M. (2019). Voice analysis using PRAAT software and classification of user emotional state. *International Journal of Interactive Multimedia and Artificial Intelligence*, *5*(6), 33–42. https://doi.org/10.9781/ijimai.2019.03.004

Magnusson, K. (2018). powerlmm: Power Analysis for Longitudinal Multilevel Models. R package version 0.4.0, <https://CRAN.R-project.org/package=powerlmm>.

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *Journal of Open Source Software, 4*(40), 1541–1549. doi:10.21105/joss.01541

Martin, O. A., Kumar, R., & Lao, J. (2021). *Bayesian Modeling and Computation in Python*. CRC Press.

Matsumoto, D. (2002). Methodological requirements to test a possible in-group advantage in judging emotions across cultures: Comment on Elfenbein and Ambady (2002) and evidence. *Psychological Bulletin, 128*(2), 236–242. https://doi.org/10.1037/0033-2909.128.2.236

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion*, *23*(2), 209–237. https://doi.org/10.1080/02699930802204677

Motro, D., Evans, J. B., Ellis, A. P., & Benson, L. III. (2022). Race and reactions to women's expressions of anger at work: Examining the effects of the "angry Black woman" stereotype. *Journal of Applied Psychology, 107*(1), 145–152. https://doi.org/10.1037/apl0000884

Pereira, C. (2000). Dimensions of emotional meaning in speech. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*.

Ratan, R., Miller, D. B., & Bailenson, J. N. (2022). Facial appearance dissatisfaction explains differences in zoom fatigue. *Cyberpsychology, Behavior, and Social Networking*, *25*(2), 124–129. https://doi.org/10.1089/cyber.2021.0112

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, *128*(6), 934–960, https://doi-org /10.1037/0033-2909.128.6.934

Robnett, R. D., Underwood, C. R., Nelson, P. A., & Anderson, K. J. (2016). "She might be afraid of commitment": Perceptions of women who retain their surname after marriage. *Sex Roles*, *75*, 500–513. https://doi.org/10.1007/s11199-016-0634-x

Rockikz, A. (2019, July 29). *How to make a speech emotion recognizer using Python and Scikit-Learn*. Python Code. https://www.thepythoncode.com/article/building-a-speech-emotion-recognizer-using-sklearn

Rose, M. R., Nadler, J., & Clark, J. (2006). Appropriately upset? Emotion norms and perceptions of crime victims. *Law and Human Behavior*, *30*(2), 203–219. https://doi.org/10.1007/s10979-006-9030-3

Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*, *57*(4), 743–762. https://doi.org/10.1111/0022-4537.00239

Salerno, J. M., & Peter-Hagene, L. C. (2015). One angry woman: Anger expression increases influence for men, but decreases influence for women, during group deliberation. *Law and Human Behavior*, *39*(6), 581–592. https://doi.org/10.1037/lhb0000147

Salerno, J. M., Peter-Hagene, L. C., & Jay, A. C. (2019). Women and African Americans are less influential when they express anger during group decision making. *Group Processes & Intergroup Relations*, *22*(1), 57–79. https://doi.org/10.1177/1368430217702967

Salerno, J. M., & Phalen, H. J. (2019). Traditional gender roles and backlash against female attorneys expressing anger in court. *Journal of Empirical Legal Studies*, *16*(4), 909–932. https://doi.org/10.1111/jels.12238

Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and emotion*, *15*(2), 123–148. https://doi.org/10.1007/BF00995674

Schuller, B., Batliner, A., Steidl, S., & Seppi, D. (2011). Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. *Speech communication*, *53*(9–10), 1062–1087. https://doi.org/10.1016/j.specom.2011.01.011

Stan Development Team (2023). RStan: The R interface to Stan. R package version 2.21.8, https://mc-stan.org/

Stegmueller, D. (2013). How many countries for multilevel modeling? A comparison of frequentist and Bayesian approaches. *American journal of political science*, *57*(3), 748–761, https://doi.org/10.1111/ajps.12001

Steinel, W., Van Kleef, G. A., & Harinck, F. (2008). Are you talking to me?! Separating the people from the problem when expressing emotions in negotiation. *Journal of Experimental Social Psychology*, *44*(2), 362–369. https://doi.org/10.1016/j.jesp.2006.12.002

Tannenbaum, M. B., Hepler, J., Zimmerman, R. S., Saul, L., Jacobs, S., Wilson, K., & Albarracín, D. (2015). Appealing to fear: A meta-analysis of fear appeal effectiveness and theories. *Psychological Bulletin*, *141*(6), 1178–1204. https://doi.org/10.1037/a0039729

Thomas, D. L., & Diener, E. (1990). Memory accuracy in the recall of emotions. *Journal of Personality and Social Psychology, 59*(2), 291–297. https://doi.org/10.1037/0022-3514.59.2.291

Tiedens, L. Z. (2001). Anger and advancement versus sadness and subjugation: the effect of negative emotion expressions on social status conferral. *Journal of Personality and Social Psychology*, *80*(1), 86–94. https://doi.org/10.1037/0022-3514.80.1.86

VanderPlas, J. (2015, August 7). *Frequentism and Bayesianism V: Model selection*. Pythonic Perambulations. http://jakevdp.github.io/blog/2015/08/07/frequentism-and-bayesianism-5-model-selection/

Van Dijk, E., van Kleef, G. A., Steinel, W., & van Beest, I. (2008). A social functional approach to emotions in bargaining: When communicating anger pays and when it backfires. *Journal of Personality and Social Psychology, 94*(4), 600–614. https://doi.org/10.1037/0022-3514.94.4.600

Van Doorn, E. A., Heerdink, M. W., & Van Kleef, G. A. (2012). Emotion and the construal of social situations: Inferences of cooperation versus competition from expressions of anger, happiness, and disappointment. *Cognition & emotion*, *26*(3), 442–461. https://doi.org/10.1080/02699931.2011.648174

Van Kleef, G. A. (2009). How emotions regulate social life: The emotions as social information (EASI) model. *Current directions in psychological science*, *18*(3), 184–188. https://doi.org/10.1111/j.1467-8721.2009.01633.x

Van Kleef, G. A., & Côté, S. (2007). Expressing anger in conflict: When it helps and when it hurts. *Journal of Applied Psychology, 92*(6), 1557–1569. https://doi.org/10.1037/0021-9010.92.6.1557

Vehtari, A., Simpson, D. P., Yao, Y., & Gelman, A. (2019). Limitations of "Limitations of Bayesian leave-one-out cross-validation for model selection". *Computational Brain & Behavior*, *2*(1), 22–27, https://doi.org/10.1007/s42113-018-0020-6

Walley-Jean, J. C. (2009). Debunking the myth of the "angry Black woman": An exploration of anger in young African American women. *Black Women, Gender & Families*, *3*(2), 68–86. https://doi.org/10.1353/bwg.0.0011

Walter, N., Tukachinsky, R., Pelled, A., & Nabi, R. (2019). Meta-analysis of anger and persuasion: An empirical integration of four models. *Journal of Communication*, *69*(1), 73–93. https://doi.org/10.1093/joc/jqy054

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science*, *6*(3), 291–298. https://doi.org/10.1177/1745691611406923

APPENDIX A

IRB APPROVAL DOCUMENTS

APPROVAL: EXPEDITED REVIEW

Jessica Salerno
Social and Behavioral Sciences, School of (SSBS)
-
Jessica.Salerno@asu.edu

Dear Jessica Salerno:

On 11/2/2018 the ASU IRB reviewed the following protocol:

Type of Review:	Initial Study
Title:	Juror Physiology and Deliberation Study
Investigator:	Jessica Salerno
IRB ID:	STUDY00009106
Category of review:	(6) Voice, video, digital, or image recordings, (4) Noninvasive procedures, (7)(a) Behavioral research
Funding:	Name: National Science Foundation (NSF), Funding Source ID: SES-1556612
Grant Title:	
Grant ID:	
Documents Reviewed:	• Physio Delib Jury Study Recruitment message, Category: Recruitment materials/advertisements /verbal scripts/phone scripts; • Physio Delib Recruit.pdf, Category: Recruitment Materials; • Physio Delib Study Measures.pdf, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • FP 4640 Salerno FP.pdf, Category: Sponsor Attachment; • Nicole Roberts' CITI Completion Report, Category: Off-site authorizations (school permission, other IRB approvals, Tribal permission etc); • Physio and Delib Jury Study original protocol.docx, Category: IRB Protocol; • Community members Consent , Category: Consent Form;
	• ASU Undergrad Consent Form, Category: Consent Form;

The IRB approved the protocol from 11/2/2018 to 11/1/2019 inclusive. Three weeks before 11/1/2019 you are to submit a completed Continuing Review application and required attachments to request continuing approval or closure.

If continuing review approval is not granted before the expiration date of 11/1/2019 approval of this protocol expires on that date. When consent is appropriate, you must use final, watermarked versions available under the “Documents” tab in ERA-IRB.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Sincerely,

IRB Administrator

APPROVAL:CONTINUATION

Jessica Salerno
NCIAS: Social and Behavioral Sciences, School of (SSBS)
-
Jessica.Salerno@asu.edu

Dear Jessica Salerno:

On 10/28/2020 the ASU IRB reviewed the following protocol:

Type of Review:	Continuing Review
Title:	Juror Physiology and Deliberation Study
Investigator:	Jessica Salerno
IRB ID:	STUDY00009106
Category of review:	
Funding:	Name: National Science Foundation (NSF), Funding Source ID: SES-1556612
Grant Title:	None
Grant ID:	None
Documents Reviewed:	• ASU Undergrad Consent - no deliberation, Category: Consent Form; • ASU No Delib Recruitment message.pdf, Category: Recruitment Materials; • Physio Delib Recruit messages, Category: Recruitment Materials; • ASU Undergrad Consent Form, Category: Consent Form; • Community members Consent , Category: Consent Form; • Community members pre-screen consent, Category: Consent Form;

The IRB approved the protocol from 10/28/2020 to 10/27/2021 inclusive. Three weeks before 10/27/2021 you are to submit a completed Continuing Review application and required attachments to request continuing approval or closure.

If continuing review approval is not granted before the expiration date of 10/27/2021 approval of this protocol expires on that date. When consent is appropriate, you must use final, watermarked versions available under the “Documents” tab in ERA-IRB.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Sincerely,

IRB Administrator

APPENDIX B

PRESCREENING SURVEY FOR IN-PERSON PARTICIPANTS

Introduction

Hello! Thank you for your interest in participating in a mock juror deliberation study at Arizona State University's West Campus!

What comes first is a survey that asks questions to determine whether you are eligible for this study. If you become concerned with the questions you do not have to complete the questionnaire. Your participation is voluntary.

If you are eligible for the in-person jury study, you will be given a link to sign up for a specific study session date. If you are eligible and agree to participate, you will come to our campus (at 4701 W Thunderbird Rd, Glendale, AZ 85306) and will be asked to play the role of a juror in a mock murder trial. You will watch selections of video from a real trial. You will give us your opinion on the case and then discuss the evidence with other people to try and reach agreement on a verdict.

If you participate in the mock jury study, you will be paid $20 per hour in Amazon credit with an additional $5 per hour in cash for parking. This will result in you being paid a total of between $20 and $40 in Amazon Credit for a session lasting 1-2 hours (plus $5 and $10 in cash for parking).

At the end of this survey, if you are eligible to participate and you agree to participate in the study, you will be directed to a link where you can schedule a time for you to participate.

Informed Consent

We are researchers at Arizona State University, and we would like to invite you to participate in this short survey to determine whether you are eligible to participate in an in-person jury deliberation study. This will involve answering some questions about yourself to determine whether you are eligible, as well as some health questions that will be helpful to us in interpreting your psychological monitoring results, if you are chosen to be hooked up to physiological monitoring.

Your participation in this survey is voluntary. You have the right not to answer any questions, and to stop participating at any time. If you choose not to participate or to withdraw from the study at any time, there will be no penalty.

If you decide to participate, we expect the survey to take you 5-10 minutes. Although there may be no other direct benefits to you, the possible benefits of your participation in the research include the opportunity to be involved in and learn about research. There are no foreseeable risks or discomforts to your participation. You must be 18 or older to participate in the study.

All information obtained in this study is strictly confidential and your responses will be anonymous. The anonymous data are stored on a password protected computer hard disk in a secure location so that only the study investigator may access it. The results of this research may be used in reports, presentations, and publications, but the researchers will not identify you. You will be asked to create a reproducible ID number by combining the last four digits of your phone number and your birthyear. You will be asked for this number again when you arrive for the in-person study, but the personal information you provide here in the Phase 1 screening survey are used only for screening purposes and will not be linked to the responses you give during the next phases of the research study.

If you have any questions concerning this study, please contact the research team / study investigator via email to: Dr. Jessica Salerno, at email: Jessica.salerno@asu.edu. If you have any questions about your rights as a subject/participant in this research, or if you feel you have been placed at risk, you can contact the Chair of the Human Subjects Institutional Review Board, through the ASU Office of Research Integrity and Assurance, at (480) 965-6788.

By advancing to the next page, you are consenting to participate in the first phase of this study, which is just the pre-screening survey. Please click the "NEXT" button to proceed with the survey. **Please DO NOT use the "back" button on your browser, as it will invalidate your responses.**

Relevant Measures

Participants were asked the following questions to determine their eligibility to participate in the study. They were also asked additional questions that are not included below to determine whether they were eligible to participate in an unrelated part of the study. As I did not use those measures in this study, they are not included here. I have added an asterisk to indicate the answers that were required in order for a person to be eligible to participate in the study.

Are you willing to attend an in-person session at Arizona State University, West Campus?

○ Yes*

○ No

If you participate in our study, you will watch a presentation of case facts and evidence from one of several real trials. Some of the cases include potentially disturbing evidence, for example, graphic photographs of the murder victims and crime scenes that include blood.

These are things that normal jurors would be exposed to, but if you do not want to see these photographs you do not have to participate.

Are you still interested in participating in the study knowing that you might be assigned a

case that includes crime scene photographs or videos?

○ Yes*

○ No

As part of this project, we will make a digital video recording of you and the other people on your jury while you discuss the case. These videos are for research purposes only and nobody will see the videos outside of the people who work in our research lab.

Are you ok with being video recorded during the study?

○ Yes*

○ No

Traditionally, to serve as a juror a person has to:

1) be a US citizen,
2) be over 18,
3) never been convicted of a felony,
4) and speak, read, and write fluent English.

Based on this list, would you qualify as a juror?

○ Yes, I am a US citizen, over 18, have never been convicted of a felony and speak/write/read fluent English*

○ No, at least one of these things is not true about me

Do you have a hearing or vision impairment that would need to be accommodated if you served
on a jury?

○ Yes

○ No*

Have you ever participated in an online study about a murder case before?

○ Yes

○ No*

APPENDIX C

PRESCREENING SURVEY FOR ONLINE PARTICIPANTS

Introduction

Thank you for your interest in participating in an online mock juror deliberation study!

What comes first is a very quick survey that asks questions to determine whether you are eligible for this study. If you become concerned with the questions you do not have to complete the questionnaire. Your participation is voluntary.

If you are eligible for the online jury study, you will be given a link to sign up for a specific study session date to complete the second part of the study. If you are eligible and agree to participate, you will receive a Zoom link and will be asked to play the role of a juror in a mock murder trial. You will watch selections of video from a real trial. You will give us your opinion on the case and then discuss the evidence with other people to try and reach agreement on a verdict.

If you participate in the mock jury study, you will be paid $20 per hour . This will result in you being paid a total of between $40 and $60 in Amazon Credit for a session lasting 2-3 hours.

At the end of this initial survey, if you are eligible to participate and you agree to participate in the study, you will be directed to a link where you can schedule a date and time for you to participate in the mock jury zoom study.

Informed Consent

We are researchers at Arizona State University, and we would like to invite you to participate in this short survey to determine whether you are eligible to participate in an online jury deliberation study. This will involve answering some questions about yourself to determine whether you are eligible.

Your participation in this survey is voluntary. You have the right not to answer any questions, and to stop participating at any time. If you choose not to participate or to withdraw from the study at any time, there will be no penalty.

If you decide to participate, we expect the survey to take you 5-10 minutes. Although there may be no other direct benefits to you, the possible benefits of your participation in the research include the opportunity to be involved in and learn about research. There are no foreseeable risks or discomforts to your participation. You must be 18 or older to participate in the study.

All information obtained in this study is strictly confidential and your responses will be anonymous. The anonymous data are stored on a password protected computer hard disk in a secure location so that only the study investigator may access it. The results of this research may be used in reports, presentations, and publications, but the researchers will not identify you. You will be asked to create a reproducible ID number by combining the last four digits of your phone number and your birthyear. You will be asked for this number again when you participate in the online study, but the personal information you provide here in the Phase 1 screening survey are used only for screening purposes and will not be linked to the responses you give during the next phases of the research study.

If you have any questions concerning this study, please contact the research team / study investigator via email to: Dr. Jessica Salerno, at email: Jessica.salerno@asu.edu.

If you have any questions about your rights as a subject/participant in this research, or if you feel you have been placed at risk, you can contact the Chair of the Human Subjects Institutional Review Board, through the ASU Office of Research Integrity and Assurance, at (480) 965-6788.

By advancing to the next page, you are consenting to participate in the first phase of this study, which is just the pre-screening survey. Please click the "NEXT" button to proceed with the survey. **Please DO NOT use the "back" button on your browser, as it will invalidate your responses.**

Measures

I have added an asterisk to indicate the answers that were required in order for a person to be eligible to participate in the study.

Are you willing to attend a 2–3 hour mock jury study online via a Zoom session?

○ Yes*

○ No

If you participate in our study, you will watch a presentation of case facts and evidence from one of several real trials. Some of the cases include potentially disturbing evidence, for example, graphic photographs of the murder victims and crime scenes that include blood.

These are things that normal jurors would be exposed to, but if you do not want to see these photographs you do not have to participate.

Are you still interested in participating in the study knowing that you might be assigned a

case that includes crime scene photographs or videos?

○ Yes*

○ No

As part of this project, we will make a digital video recording of you and the other people on your jury while you discuss the case. These videos are for research purposes only and nobody will see the videos outside of the people who work in our research lab.

Are you ok with being video recorded during the study?

○ Yes*

○ No

If the participant was completing the survey on a mobile phone or tablet:

You are currently completing this survey on a mobile phone or tablet. To participate in the online deliberation, you need to have access to a computer. Do you have access to a computer that you

could use to participate in the online deliberation?

- ○ Yes*
- ○ No
- ○ I don't know

To participate in the online deliberation, you need to be in a relatively quiet room (or wearing headphones so you are not distracted) for up to 3 hours. In other words, you will not be able to participate while in your car, walking, etc. You also need to be able to talk to others. Do you

have access to a quiet room, with a computer, where you can talk to others for up to 3 hours?

- ○ Yes*
- ○ No
- ○ I don't know

To participate in the online deliberation, you need to have a webcam on your computer. Do you

have access to a webcam?

- ○ Yes*
- ○ No
- ○ I don’t know

To participate in the online deliberation, you need to have your webcam on for the entire session.

Are you comfortable and able to keep your webcam on for the entire session?

- ○ Yes*
- ○ No

To participate in the online deliberation, you need to have a microphone on your computer. Do

you have access to a microphone on your computer?

- ○ Yes*
- ○ No
- ○ I don’t know

To participate in the online deliberation, you need internet upload speed of at least 1.5 mbps. Do

you have an internet upload speed of at least 1.5 mbps?

- Yes*
- No
- I don't know

Have you ever used Zoom?

- Yes*
- No
- I don't know

Are you familiar with how to use Zoom?

- Yes*
- No
- I don't know

Traditionally, to serve as a juror a person has to:

5) be a US citizen,

6) be over 18,

7) never been convicted of a felony,

8) and speak, read, and write fluent English.

Based on this list, would you qualify as a juror?

○ Yes, I am a US citizen, over 18, have never been convicted of a felony and speak/write/read fluent English*

○ No, at least one of these things is not true about me

Do you have a hearing or vision impairment that would need to be accommodated if you served
on a jury?

○ Yes

○ No*

Have you ever participated in an online study about a murder case before?

○ Yes

○ No*

APPENDIX D

STUDY MATERIALS

Prosecution Opening Statement

On the evening of June 17th 2012 , the defendant, Michael Stevens, had an intense argument with his wife, Stacy. At the end of the argument, Stacy told him she was going to leave him. After yelling “you’ll be sorry when I’m gone!”, she locked herself in the master bedroom. The defendant spent the next two hours talking to his sister, who lived with the couple, and at 10:00pm, he pretended to go to bed, saying he would sleep in the guest bedroom.

But the defendant didn’t go to bed. Instead, he broke into the master bedroom with a knife, and slit his wife’s throat all the way across. You will hear that the defendant brutally slit Stacy’s throat with two separate cuts, one so deep that it went through her larynx. Stacy did not survive the heinous attack.

The defendant and Stacy’s marital problems did not begin on June 17th. The defendant’s sister, Elle Stevens, will tell you about their marital problems. She will describe the fight they had on June 17th. She’ll also tell you that the defendant was acting strange the day after the murder.

The defense will argue that Stacy committed suicide. They will likely point to the fact that the bedroom door was found locked from the inside after she was dead. However, we will call Henry Gold to testify, who is an expert locksmith. He will explain that, although the bedroom door was locked from the inside, it is possible to maneuver such a lock from the outside. Additionally, you will hear from Dr. Christopher Oettle, the state pathologist. He examined Stacy’s body and will explain to you how the crime scene and body indicate that Stacy was killed by someone else and that she could not have committed suicide. The evidence you hear will conclusively prove beyond a reasonable doubt that the defendant murdered Stacy Stevens. Stacy threatened to leave, and he responded in the most brutal way possible. By the end of trial, when you consider all the evidence, I will ask you to find him guilty.

Defense Opening Statement

Stacy Stevens was very troubled. She had a history of depression, and her marriage was troubled. On June 17, 2012, after a difficult fight with her husband, Michael, she locked herself in her bedroom and slit her own throat.

During the argument, Stacy said “You will be sorry when I’m gone”, which was clearly her way of telling him that she was going to take her own life. Stacy locked herself in the master bedroom and Michael went to sleep in the guest bedroom. In the morning, Michael found the door still locked and Stacy refusing to respond to his pleas to talk to her. You will hear from his sister that before she left for work she witnessed her brother in great distress and very worried that his wife might have hurt herself. In his distress he went to talk to his neighbor, William Morgan, and ended up asking him to call the police. But it was too late. Stacy had already taken her own life.

The prosecutor will argue that Michael killed Stacy. But the evidence does not support this argument. There is no hard evidence that Michael was ever in the room. You will hear from Michael's sister and Mr. Morgan, their neighbor, that Stacy had a history of depression and about the actions Michael was taking to repair their marriage.

You will hear that there is a lack of forensic evidence indicating that Stacy was murdered that one would expect if this were a murder. Further, you will hear that the bodily evidence is consistent with Stacy committing suicide.

Michael Stevens loved his wife. On June 18, the day her death was discovered, he was trying to mend their marriage. The scientific evidence will demonstrate that Stacy was not murdered but rather committed suicide. After the trial, after you have heard all the evidence, you will have no choice but to find Michael Stevens not guilty.

<u>Elle Stevens, Defendant's Sister</u>
Examination by Prosecution Attorney
Q. Can you describe the relationship between the defendant and Stacy?
Defense. Objection, relevance.
Judge. Overruled.
A. Michael loved Stacy very much, but they had a difficult marriage. I lived with them when Stacy died and they were arguing constantly.

Q. You said that you lived with them in June 2012. Can you tell us what happened on June 17?
A. I overheard Michael and Stacy arguing. Stacy said "you'll be sorry when I'm gone" and locked herself in the bedroom.

Q. What happened next?
A. Michael came into my room and talked with me for about two hours about the problems they were experiencing and how everything was fine now, but Stacy was still upset. He confessed that he was worried Stacy would take the children and leave him. At around 10 pm, Michael told me he was going to leave Stacy alone for the night and sleep in the guest bedroom.

Q. Did you see the defendant or Stacy the next morning?
A. I didn't see Stacy. I saw Michael, who was trying to open the door, which was still locked. He was upset and trying to talk to her through the door right before I left for work at around 9. He was arrested later that day.

<u>Cross Examination by Defense Attorney</u>
Q. After Michael went to bed, you went to sleep, didn't you?
A. Yes, I did.

Q. Were you woken up at any point during the night?

A. No I wasn't.

Q. Did you hear a struggle at any point?
A. No.

Q. Did you hear anything during the night that would suggest Michael got up and left the guest room?
A. No.

Q. Did you hear anything during the night that would suggest Michael entered the room where Stacy was?
A. No.

Q. To your knowledge, did Michael and Stacy ever seek counseling for their marital problems?
A. I'm not sure if they ever attended counseling but on June 18th, before I left for work, Michael called a marriage counselor for advice.

<u>Henry Gold, Locksmith</u>
Examination by Prosecution Attorney

Q. Did you examine the Stevens' bedroom door lock?
A. Yes, and I took a picture of the lock.

Q. I'm showing you these photographs, previously admitted as State's exhibits E, F, G, and H. Can you describe the lock in these pictures?
A. Yes. The lock was a standard deadlock, or deadbolt. A deadlock cannot be moved to an open position except by rotating the lock cylinder, usually with a key. State's exhibit E is the lock from the outside of the bedroom. You can see the keyhole and the door handle. Exhibit F is the door from the side. This top part is the deadbolt. It is inserted into a strike plate, featured here in exhibit G, to lock the door when activated by a key. As you turn the door handle, the latch bolt, this bolt here, moves in and out of the faceplate. That's this part. When you close the door, the latch bolt catches in the strike plate, featured here in exhibit G. Finally, exhibit H is the lock from the inside of the bedroom. To lock the door, you turn this handle.
Q. Can a dead lock be picked?
A. Yes, a dead lock can be picked using two bobby pins, if you know how to do it.

Q. In this case, the police found the bedroom door locked from the inside. Does this mean the door could not have been locked from the outside?
A. Not exactly. It is possible to maneuver a lock from the outside to make it appear locked from the inside. You do this by reversing the process used to pick a lock.

Q. So it is possible to manipulate a lock from the outside to make it look locked from the inside, by reverse picking a lock, so to speak?
A. Exactly Cross Examination by Defense Attorney

Q. No evidence in this case indicates that Michael knows how to maneuver a lock in this way, correct?
A. The information is easily accessible online, but that is correct.

Q. Are you aware of any evidence that Michael searched for this information?
A. No.

<u>Christopher Oettle, Pathologist</u>
Examination by Prosecution Attorney
Q. What is a pathologist?
A. We're doctors who determine the cause and manner of a death by examining bodily organs, tissue, and fluids.

Q. What do you mean by cause of death?
A. Cause of death is the disease or injury that led directly to death.

Q. What do you mean by manner of death?
A. Manner of death is whether the death was an accident, was due to a self-inflicted act (like suicide), or whether it, was a homicide, which is when another person is responsible for the injuries, which resulted in the death.

Q. Did you form an opinion regarding the cause of death in this case?
A. Yes.

Q. Doctor, what is your opinion, to a reasonable medical certainty, on the cause of Stacy's death?
A. She died of a major wound across her throat, caused by two separate cuts and resulting in major loss of blood and death.

Q. Before you describe the wounds that lead you to that conclusion, is there anything that will help you to explain Stacy's wounds to the jury?
A. Yes, during the crime scene investigation and autopsy, I took photographs of the deceased.

Q. Alright, I'm showing you one of those photographs, marked State's exhibit A. Can you describe the photograph?
A. Yes, this is a photograph of the deceased. In this photograph, you can see the major wound across her throat and several blood smudges on the deceased's face and chest.

Q. And Exhibit C? Can you describe that photograph?

A. This is an autopsy photograph of the deceased. It shows the wound spanning from the left side of the neck all the way to the right side. You can see that the wound is gaping and has a half-moon shape.

Q. Can you tell us what Exhibit B shows?
A. This is an autopsy photograph from another angle. This one shows the wound from the front. As you can see, the wound is roughly 2 inches wide at the front of the throat. You can also see that the edges of the wound appear smooth everywhere.

Q. I'd like to move onto Exhibit D, what does this photograph show?
A. Again, this is an autopsy photograph from another angle. In this photograph, you can see the larynx because the skin was pulled back.

Q. What is significant about a photograph of the larynx?
A. In this photograph, you can see that one of the cuts was deep enough to go through the larynx.

Q. Does this photograph show anything else?
A. Yes, you can see that the internal jugular vein and the common carotid artery on each side of the throat were cut. Additionally, you can see superficial parallel cuts along the edge of the wound.

Q. Did you determine the manner of death in this case?
A. Yes.

Q. And what was the manner of death?
A. Homicide—her death was caused by another person causing her injuries.

Q. How did you come to that conclusion?
A. The crime scene and bodily evidence indicate that the deceased died by homicide and that she could not have committed suicide based on the angle of the wounds, the position of the body, the blood evidence, and her clothing.

Q. What about the angle of the wounds indicated that Stacy died by homicide?
A. The angle and depth of the wounds were more consistent with homicide than with suicide by a right-handed person. Stacy was right handed, which suggests she could not have done this.

Q. What about the position of the body indicated that Stacy died by homicide?
A. The deceased was found face down, but she had blood on her back, which means the body was turned after much of the bleeding took place. The suicide scenario would require her to be conscious enough to get up and change her position after cutting her throat the first time, which is unlikely.

Q. What about the blood evidence indicated that Stacy died by homicide?
A. The blood smears on the deceased's left leg and the position of her nightgown suggested her body might have been dragged.

Q. Finally, what about her clothing indicated that Stacy died by homicide?
A. The deceased was wearing several necklaces, which is uncommon because people usually remove "obstacles" such as jewelry before they commit suicide.

Q. Based on all of the evidence you collected and we've discussed today, what is your final conclusion?
A. The deceased died by homicide, not by suicide.

Cross Examination by Defense Attorney
Q. I'd like to talk to you about the victim's injuries. There was no evidence of a struggle or defense injuries to the victim's arms and hands, correct?
A. That's correct.

Q. Typically, victims of homicide struggle and attempt to protect themselves, correct?
Prosecution Attorney: Speculation.
Judge: Sustained.

Q. Next I'd like to talk to you about the bloodstain patterns. The pattern of bloodstains on the wall indicated that the victim was coughing blood for a period of time after the first cut, correct?
A. Yes. That is correct.

Q. This might mean that both of the cuts were not made at the same time.
A. That is one possible explanation.

Q. In a homicide scenario, the perpetrator would most likely make both cuts at once, correct?
A. Not necessarily. While that might be true most of the time, there are certainly times where a perpetrator does not make both cuts at once, particularly when the perpetrator is unsure of their actions.

<u>William Morgan, Neighbor</u>
Examination by Defense Attorney
Q. Where were you on June 18, 2012?
A. I was at home working in my yard.

Q. Did anything unusual happen that day?
A. Yes, at about 10:00am, Michael, my neighbor, came running up and asked if I had seen Stacy or heard from her.

Q. Why was Michael asking?
A. Well he was really worried because he and Stacy had a big fight and she had locked herself in her room—he knows that Stacy is close friends with my wife and thought maybe she had called her. He was worried that she had hurt herself. He was very upset and was having trouble holding it together.

Q. What did you do?
A. I called the police because Michael was too distressed to talk. They came and found Stacy in the bedroom.

Q. I'd like to shift gears and discuss your relationship with Michael and Stacy. Were you close?
Prosecution Attorney: Objection-Relevance.
Defense Attorney: This question is laying foundation to show that Mr. Morgan is qualified to discuss Stacy's mental health.
Judge: Overruled.
A. Yes. We, my wife and I, spent a lot of time with Michael and Stacy. We confided in each other.

Q. Are you aware of the Stevens' marital problems?
A. Yes. Michael and Stacy were having a lot of problems. Michael told me that June 18th that he had called a marriage counselor that morning for advice about their marriage.

Q. There has been a lot of discussion today about Stacy's mental state. Did Stacy ever confide in you about her mental health?
A. Yes. Stacy told my wife and I that she often suffered from severe depression.

Q. Are you aware of whether Stacy was being treated for her depression? A. She wasn't. We tried to convince her several times to find a therapist and discuss meditation with her doctor when it got really bad, but she was always really resistant.

Prosecution Closing Statement

Stacy Stevens and her husband, the defendant, had marital problems. That is not contested. On June 17, 2012, they argued, like they had so many times before. But this time was different because Stacy threatened to leave the defendant. And the defendant, fearing that Stacy would leave him and take their children, did the unthinkable. He took a knife and slit Stacy's throat in two places. He then left her to bleed to death in their bedroom.

The evidence proves that the defendant killed Stacy Stevens. He had the motive, the means, and the opportunity to kill Stacy—no one else did.

He had the motive. You heard about their marital problems from the defendant's sister, Elle Stevens. You heard that the defendant was afraid that Stacy was going to take his children and leave.

He had the means. You heard from Dr. Oettle that Stacy was killed in the master bedroom with a knife. The defendant had access to the master bedroom and access to their kitchen knives. Henry Gold, a locksmith, testified that it is easy to pick deadlocks, the type of lock that the Stevens had on their bedroom door. Further, Mr. Gold testified that anybody with access to the internet could learn how to maneuver a lock from the outside to make it appear locked from the inside. The defendant had the means to kill his wife.

He had the opportunity. You heard from Elle Stevens that she went to bed immediately after the defendant left their room. The defendant had plenty of time from when his parents when to sleep from when they woke up to kill his wife. And he took advantage of that time.

But that isn't all of the evidence against the defendant—the crime scene and pathology evidence gives him away. You heard Dr. Oettle testify that Stacy's wounds, these wounds were consistent with homicide, not with suicide. Do these looks like something a woman would do to herself? Just because of a fight—a fight that was routine in this marriage? No. The defendant ran a knife from the right side of Stacy's throat across to the left in two cuts. He cut her throat so deeply that he severed both internal jugular veins and both common carotid arteries. He left her lying in a pool of her own blood.

The defense wants you to believe that after losing a large amount of blood, Stacy stood up and slashed her neck a second time. But you heard Dr. Oettle testify that that scenario is unlikely. Further, you heard Dr. Oettle testify that the blood evidence indicated that the body might have been dragged. All of this evidence paints a picture of a brutal, heinous murder—not a suicide.

The defendant did not want his wife to leave him. But rather than resolving their conflict, he took an irreversible, dramatic step: he killed her. Look at the evidence, look at the photographs of Stacy Stevens. When you do, you will reach the conclusion that the defendant is guilty.

Defense Closing Statement

Michael Stevens loved his wife. He helped her for years while she struggled with depression. And that depression put a strain on their marriage, yes. But Michael wanted nothing more than to fix their marriage. Unfortunately, on June 17, after a difficult fight, Stacy took her own life.

The State has now charged Michael with her murder. But they must prove beyond a reasonable doubt that he is guilty. They have not done that. The forensic evidence is not conclusive of Michael killing his wife—not by a long shot.

The State argues that Michael was acting suspicious on June 17th and 18th and that this demonstrates that he killed Stacy. But on June 18th, Michael was trying to talk to Stacy through the bedroom door. He called a marriage counselor about how to save his marriage. Are those the actions of someone who killed his wife? No. Why would Michael try to talk to Stacy through the door if he knew that Stacy was already dead? Why would he call a marriage counselor and seek advice on how to fix his marriage if he knew that Stacy was dead? He wouldn't. All of the actions that Michael took on June 18th were designed to fix his marriage. They were not the actions of a man who had just murdered his wife.

Further, you heard today that Stacy had a history of depression. And you heard that during the fight on June 17, she made a suicidal threat.

And even if you do not believe Michael, and even if you do not think that Stacy made a suicidal threat, the forensic evidence is clear and raises reasonable doubt about Michael killing Stacy. The forensic evidence indicated that Stacy committed suicide. The bloodstains indicate that the two cuts could not have been made at the same time, a narrative consistent with suicide, not homicide. The position of the knife under Stacy's body is consistent with her falling while holding the knife, not with someone cutting her throat. And finally, there was no evidence that Stacy struggled or that someone attempted to clean up the crime scene. These two things also indicate that Stacy committed suicide and raise doubt that Michael did this.

Stacy Stevens death was a tragedy. Her family was ripped apart and their lives will never be the same. But that does not mean Michael killed her. Michael Stevens is a grieving husband, not a murderer. The prosecution has failed to meet their burden of proof and you must return a verdict of not guilty.

Control Instructions

THE JUDGE'S INSTRUCTIONS TO YOU, THE JURY

Members of the jury, before the evidence and arguments in this case are completed, I will instruct you as to the law.

The law that applies to this case is stated in these instructions, and it is your duty to follow all of them. You must not single out certain instructions and disregard others.

It is your duty to determine the facts and to determine them only from the evidence in this case. You are to apply the law to the facts and in this way decide the case. Neither sympathy nor prejudice should influence you.

The evidence which you should consider consists only of the testimony of the witnesses and the information they provide.

You should consider all the evidence in the light of your own observations and experience in life. By these instructions I do not mean to indicate any opinion as to the facts or as to what your verdict should be.

Faithful performance by you of your duties as jurors is vital to the administration of justice.

The defendant is presumed to be innocent of the charge against him of first degree murder. This presumption remains with him throughout every stage of the trial and during your deliberations on the verdict.

This presumption is not overcome unless, from all the evidence in this case, you are convinced beyond a reasonable doubt that the defendant is guilty. The State has the burden of proving that the defendant is guilty of first degree murder, and this burden remains on the State throughout the case. The defendant is not required to prove his innocence.

Only you are the judges of the believability of the witnesses and of the weight to be given to the testimony of each of them.

In considering the testimony of any witness, you may take into account his or her ability and opportunity to observe, age, memory, manner while testifying, any interest, bias, or prejudice he or she may have, and the reasonableness of his or her testimony considered in the light of all the evidence in the case.

You should judge the testimony of the defendant in the same manner as you judge the testimony of any other witness.

YOU HAVE TWO VERDICT OPTIONS IN THIS CASE:

- FIND THE DEFENDANT, MICHAEL STEVENS, **GUILTY** OF FIRST-DEGREE MURDER.
- FIND THE DEFENDANT, MICHAEL STEVENS, **NOT GUILTY**.

To sustain the charge of first degree murder, the State (the Prosecution) must prove the following Propositions:

1. **First Proposition**: That the defendant, Michael Stevens, performed the acts which caused the death of Stacy Stevens.

AND

2. **Second Proposition**: That when the defendant, Michael Stevens, did so, he intended to kill Stacy Stevens.

Choose NOT GUILTY if:
If you find from your consideration of all the evidence that **any** **one** of these propositions has **not** been proved beyond a reasonable doubt, you should return a verdict of **Not Guilty.** In other words, if you think that either the **First Proposition** OR the **Second Proposition** described above was not proved, you should vote Not Guilty.

Choose GUILTY if:
If you find from your consideration of all the evidence that **each** **one** of these propositions has been proved beyond a reasonable doubt, you should return a verdict of **Guilty.** In other words, if you think that BOTH **First Proposition** AND **the Second Proposition** described above was proved, you should vote Guilty.

<u>Emotion Awareness Instructions (Additional instructions are italicized)</u>

THE JUDGE'S INSTRUCTIONS TO YOU, THE JURY

Members of the jury, before the evidence and arguments in this case are completed, I will instruct you as to the law.

The law that applies to this case is stated in these instructions, and it is your duty to follow all of them. You must not single out certain instructions and disregard others.

It is your duty to determine the facts and to determine them only from the evidence in this case. You are to apply the law to the facts and in this way decide the case. Neither sympathy nor prejudice should influence you.

The evidence which you should consider consists only of the testimony of the witnesses and the information they provide.

In this case, photographs of the deceased might be admitted in evidence. If so, these photographs have been admitted to provide details about the victim's injuries and to help you visualize issues relevant to the case.

You may, understandably, find the photographs upsetting. Be aware that in addition to helping you resolve the issues in the case, the photographs may also influence your decision in inappropriate ways. Being upset about the disturbing events depicted in the photograph might make you want to convict someone for the crime.

This desire to convict someone might lower your threshold for how much proof you need to believe that the State has met the "beyond a reasonable doubt" standard and convict.

The desire to convict might also influence you to pay more attention to evidence that supports a guilty verdict than you pay to evidence that supports a not guilty verdict.

You should consider all the evidence in the light of your own observations and experience in life. By these instructions I do not mean to indicate any opinion as to the facts or as to what your verdict should be.

Faithful performance by you of your duties as jurors is vital to the administration of justice.

The defendant is presumed to be innocent of the charge against him of first degree murder. This presumption remains with him throughout every stage of the trial and during your deliberations on the verdict.

This presumption is not overcome unless, from all the evidence in this case, you are convinced beyond a reasonable doubt that the defendant is guilty. The State has the burden of proving that the defendant is guilty of first degree murder, and this burden remains on the State throughout the case. The defendant is not required to prove his innocence.

Only you are the judges of the believability of the witnesses and of the weight to be given to the testimony of each of them.

In considering the testimony of any witness, you may take into account his or her ability and opportunity to observe, age, memory, manner while testifying, any interest, bias, or prejudice he or she may have, and the reasonableness of his or her testimony considered in the light of all the evidence in the case.

You should judge the testimony of the defendant in the same manner as you judge the testimony of any other witness.

YOU HAVE TWO VERDICT OPTIONS IN THIS CASE:

- FIND THE DEFENDANT, MICHAEL STEVENS, **GUILTY** OF FIRST-DEGREE MURDER.
- FIND THE DEFENDANT, MICHAEL STEVENS, **NOT GUILTY**.

To sustain the charge of first degree murder, the State (the Prosecution) must prove the following Propositions:

1. **First Proposition**: That the defendant, Michael Stevens, performed the acts which caused the death of Stacy Stevens.

<u>AND</u>

2. **Second Proposition**: That when the defendant, Michael Stevens, did so, he intended to kill Stacy Stevens.

Choose NOT GUILTY if:
If you find from your consideration of all the evidence that **<u>any</u> <u>one</u>** of these propositions has **<u>not</u>** been proved beyond a reasonable doubt, you should return a verdict of **Not Guilty.** In other words, if you think that either the **First Proposition** <u>OR</u> the **Second Proposition** described above was not proved, you should vote Not Guilty.

Choose GUILTY if:
If you find from your consideration of all the evidence that **<u>each</u> <u>one</u>** of these propositions has been proved beyond a reasonable doubt, you should return a verdict of **Guilty.** In other words, if you think that BOTH **First Proposition** AND **the Second Proposition** described above was proved, you should vote Guilty.

Using the scale below, please indicate how much you were feeling each of the following emotions **when you heard the evidence of the victim's injuries.**

	Not at all	Slightly	Moderately	Much	Very Much
I felt anxiety	○	○	○	○	○
I felt contempt	○	○	○	○	○
I felt grossed-out	○	○	○	○	○
I felt outrage	○	○	○	○	○
I felt sadness	○	○	○	○	○
I felt unhappiness	○	○	○	○	○
I felt empathy for the victim	○	○	○	○	○
I felt sympathy for the victim	○	○	○	○	○
I felt pity	○	○	○	○	○
I felt anger	○	○	○	○	○
I felt disgust	○	○	○	○	○
I felt interest	○	○	○	○	○
I felt repulsed	○	○	○	○	○
I felt fear	○	○	○	○	○
I felt compassion for the victim	○	○	○	○	○
I felt depression	○	○	○	○	○
I felt happiness	○	○	○	○	○
I felt infuriated	○	○	○	○	○
I felt pleasure	○	○	○	○	○
I felt sickened	○	○	○	○	○

Next we are going to ask your impressions of the other jurors who you deliberated with today.

You might have noticed that everyone has a juror number in front of them on the table. Please tell us your impressions about each of the jurors you deliberated with, corresponding to the juror number.

You will be asked about eight juror numbers. When you are asked about your own Juror number, please answer the questions about your own contributions to the deliberation. For example, if you are Juror 3, below answer the Juror 3 questions about yourself.

If we ask about a juror number that did not exist on your jury (e.g., if we ask about 8 jurors, but your jury only had 6 jurors) just leave the ones that do not exist on your jury blank.

Please think about **Juror XX** when answering the following questions. Please think about each word and rate Juror 1 on each quality.

	Not at all	Slightly	Somewhat	Very Much	Extremely
Influential	○	○	○	○	○
Persuasive	○	○	○	○	○
Presented high quality arguments	○	○	○	○	○
Likeable	○	○	○	○	○
Competent	○	○	○	○	○
Warm	○	○	○	○	○
Emotional	○	○	○	○	○

What is your gender?

- ○ Male
- ○ Female
- ○ Other ______________________________________

What is your ethnicity?

- ○ White
- ○ Black or African American
- ○ American Indian or Alaska Native
- ○ Asian
- ○ Native Hawaiian or Pacific Islander
- ○ Other ______________________________________

APPENDIX E

MODEL BUILDING PROCESS

Building the Models

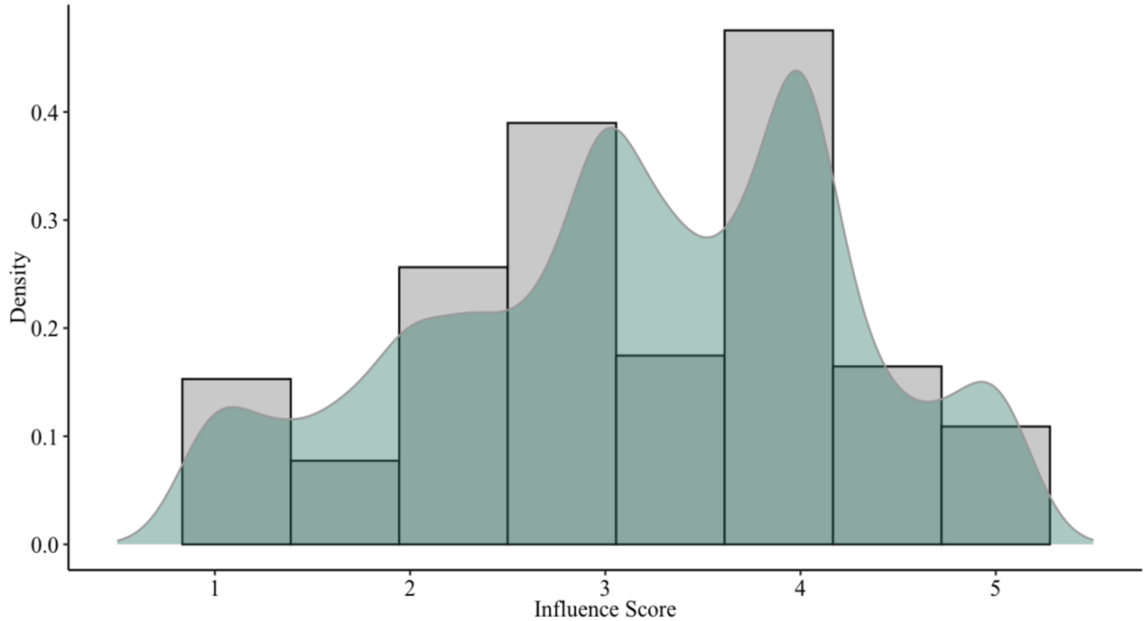
Martin et al. (2021) recommend considering three steps when constructing models: 1) determine the distribution that describes the observed data; 2) determining how to model the data collection process; and 3) set priors.

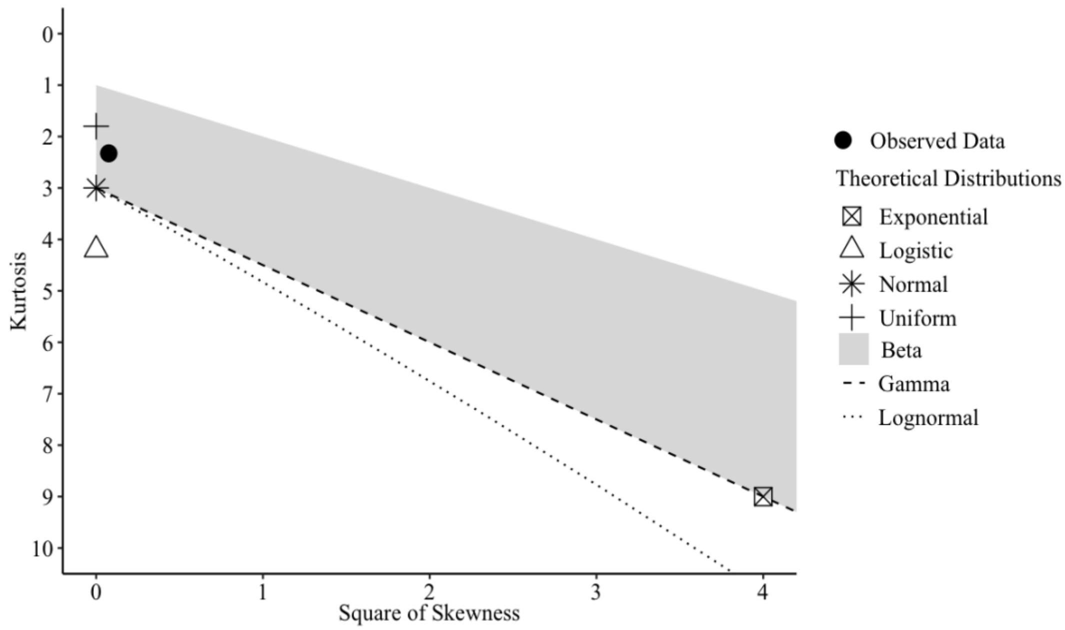
Determine the Distribution of the Dependent Variable

I began by examining the distribution of the dependent variable. As shown in Figure A1, influence was slightly left-skewed (skewness = -0.27) and platykurtic (kurtosis = -0.68). However, these values for skew and kurtosis fall within the recommended range to conclude that data is normally distributed (George & Mallery, 2010) and an examination of a Cullen and Frey (1999) graph suggests that the normal distribution is the best fit for the data.

Figure A1.

Distribution of Ratings of Perceived Influence of Other Jurors





Note. The top plot shows a histogram, and density plot for influence. The bottom plot shows a skewness-kurtosis graph for a variety of distributions (Cullen & Frey, 1999). The black dot represents the observed influence scores. The other symbols and lines represent the skew and kurtosis of the theoretical distributions. The observed data falls between the uniform and normal distributions, suggesting that modeling the data with a normal distribution is appropriate.

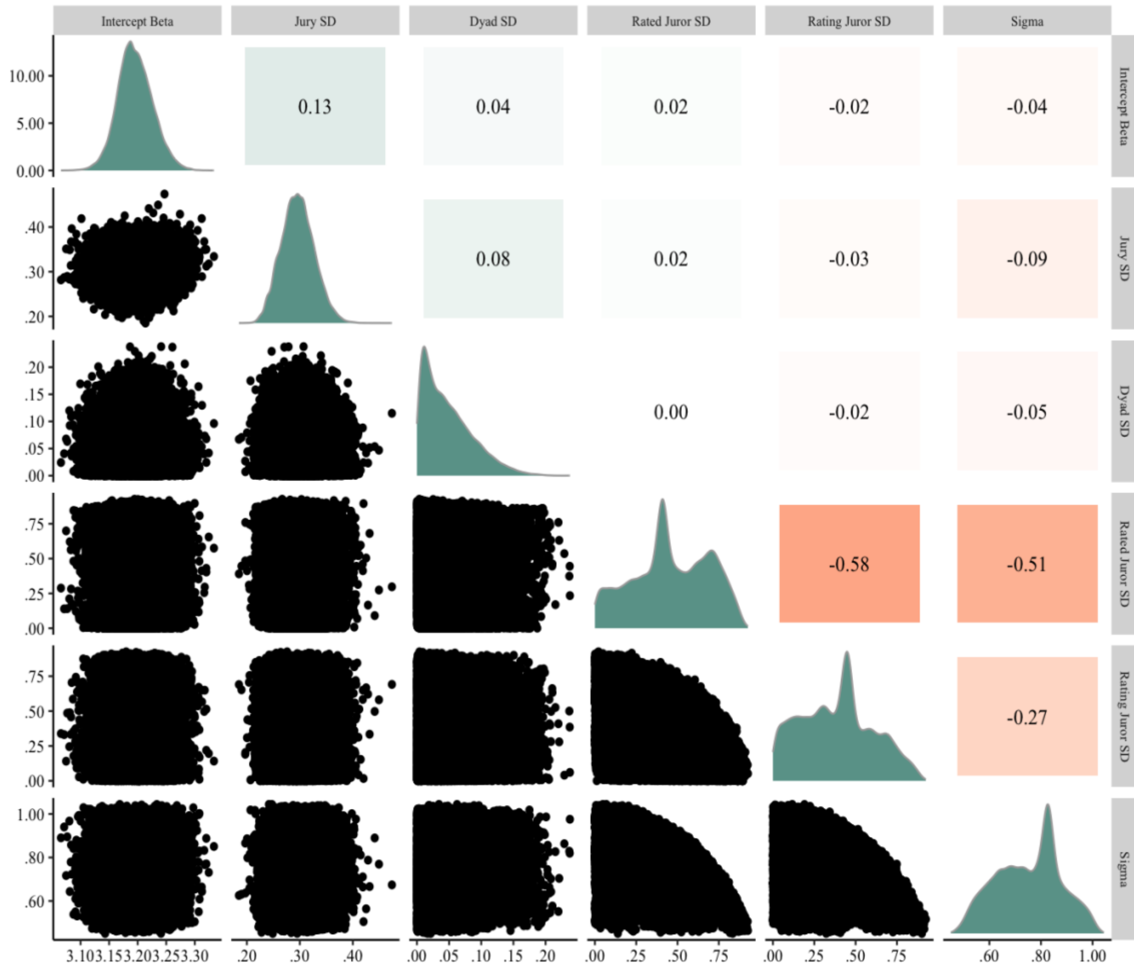
Model the Data Collection Process

After determining that a normal distribution was the most appropriate distribution for the observed data, I began testing several random effects structures to determine the best way to model the data collection process. In line with Barr et al. (2013) recommendation that researchers should use the maximal random effects structure that is justified by the design, I began by testing a model with the maximal random effects structure such that rating and rated juror were crossed random effects that were nested within dyad and dyad was, in turn, nested within jury. The model syntax for that model was:

$$Influence \sim 1 + (1 \,|\, Jury) + (1 \,|\, Jury{:}DyadID) + (1 \,|\, Jury{:}DyadID{:}RatingJuror) \\ + (1|Jury{:}DyadID{:}RatedJuror)$$

However, there were significant convergence problems with this model. First, there were 5,213 divergent transitions while running the model. Divergent transitions can indicate that part of the posterior is not being properly estimated and, therefore, estimates are uncertain. Second, the estimated Bayesian Fraction of Missing Information (e-BFMI) was less than or equal to 0.06 across all four chains. The rule of thumb is that e-BFMI should be above 0.2. A low e-BFMI can also indicate that part of the posterior is not being properly estimated and that estimates are uncertain. Third, the largest R-hat was 1.23, which suggests that the four separate Markov chains that were used to estimate the posterior did not converge. Again, this suggests that the estimates are uncertain. Finally, both the bulk effective sample size and tail effective sample size were too low. This can indicate that both the parameter estimates, and that the credible interval estimates are unstable. I investigated the pairs plot (Figure), which is a scatterplot matrix of the parameter estimates. The pairs plot suggested that the convergence problems were driven by some combination of the Dyad, Rated Juror, and Rating Juror random effects.

Figure A2.
Pairs Plot for the Most Complex Model



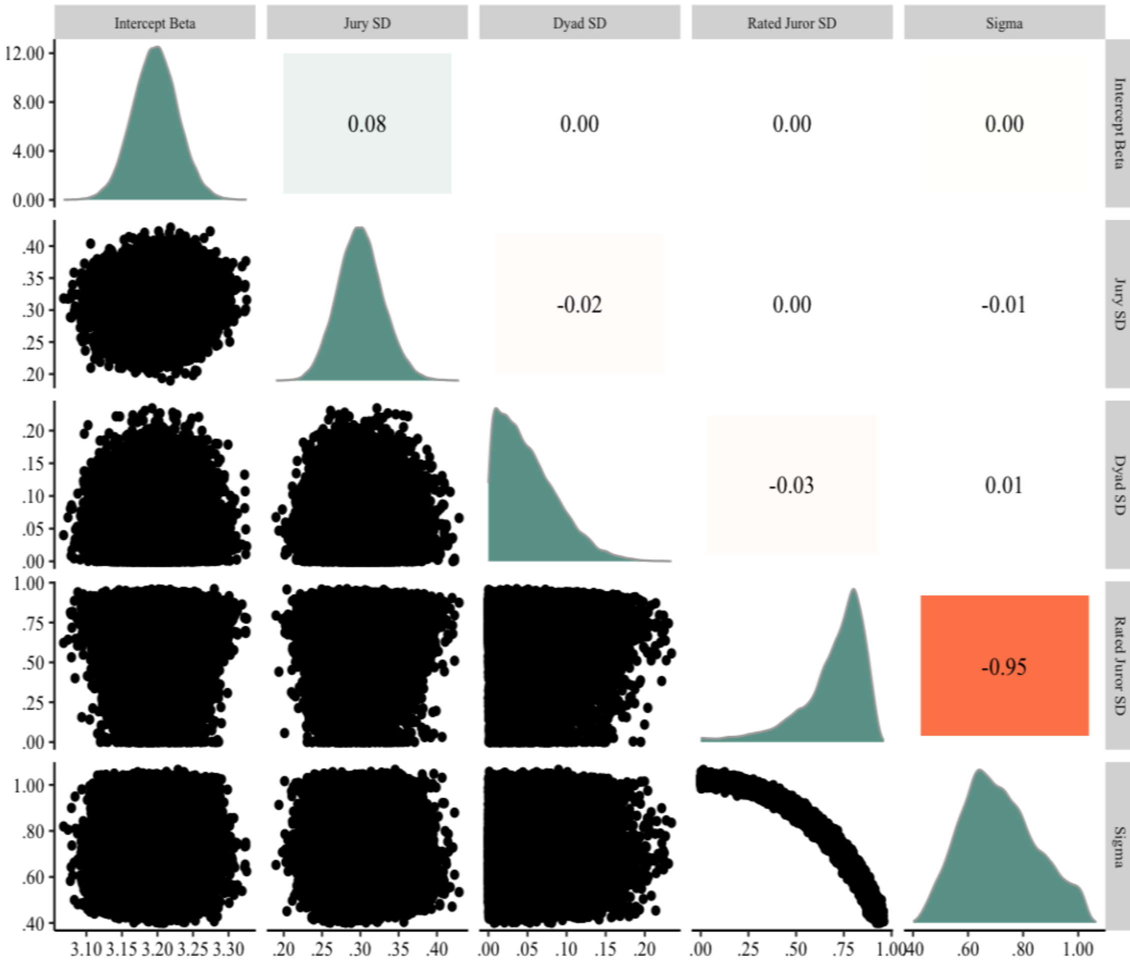
Note. A pairs plot shows the relationship between all parameters in the model. The plots on the diagonal are density plots for each of the variables and should be relatively normally distributed. The upper half of the plot shows the correlations between all variables in the model. High correlations can be an indicator of poor model fit. The lower half of the plot shows scatterplots of each relationship. Funneling, or when a scatterplot is more spread out in one portion of the graph compared to the other, can indicate that the sampler was not able to explore the entire parameter space.

While using the maximal random effects structure can reduce Type I errors and is the most reflective of the data collection process, when there are significant convergence problems, the posterior parameter estimates can be unstable, which can also increase the risk of Type I errors. Therefore, I followed the recommendations of Bates et al. (2015) and I removed random effects until I found the most parsimonious model. Based on visual inspection of the pairs plot from the most complex model and because the research questions are primarily focused on perceived influence, rather than susceptibility to influence, I first removed the random effect relating to Rating Juror. The model syntax for the new model was:

$$Influence \sim 1 + (1 \mid Jury) + (1 \mid Jury\!:\!DyadID) + (1 \mid Jury\!:\!DyadID\!:\!RatedJuror)$$

While the convergence improved, there were still significant convergence problems including 247 divergent transitions, an estimated Bayesian Fraction of Missing Information (e-BFMI) was less than or equal to 0.02 across all four chains, and 377 transitions that exceeded the maximum treedepth, suggesting that, again, part of the posterior is not being properly estimated and that estimates are uncertain. Again, visual inspection of the pairs plot (Figure) suggested that the convergence problems were driven by some combination of Dyad and Rated Juror.

Figure A3.
Pairs Plot for the Model without Rating Juror

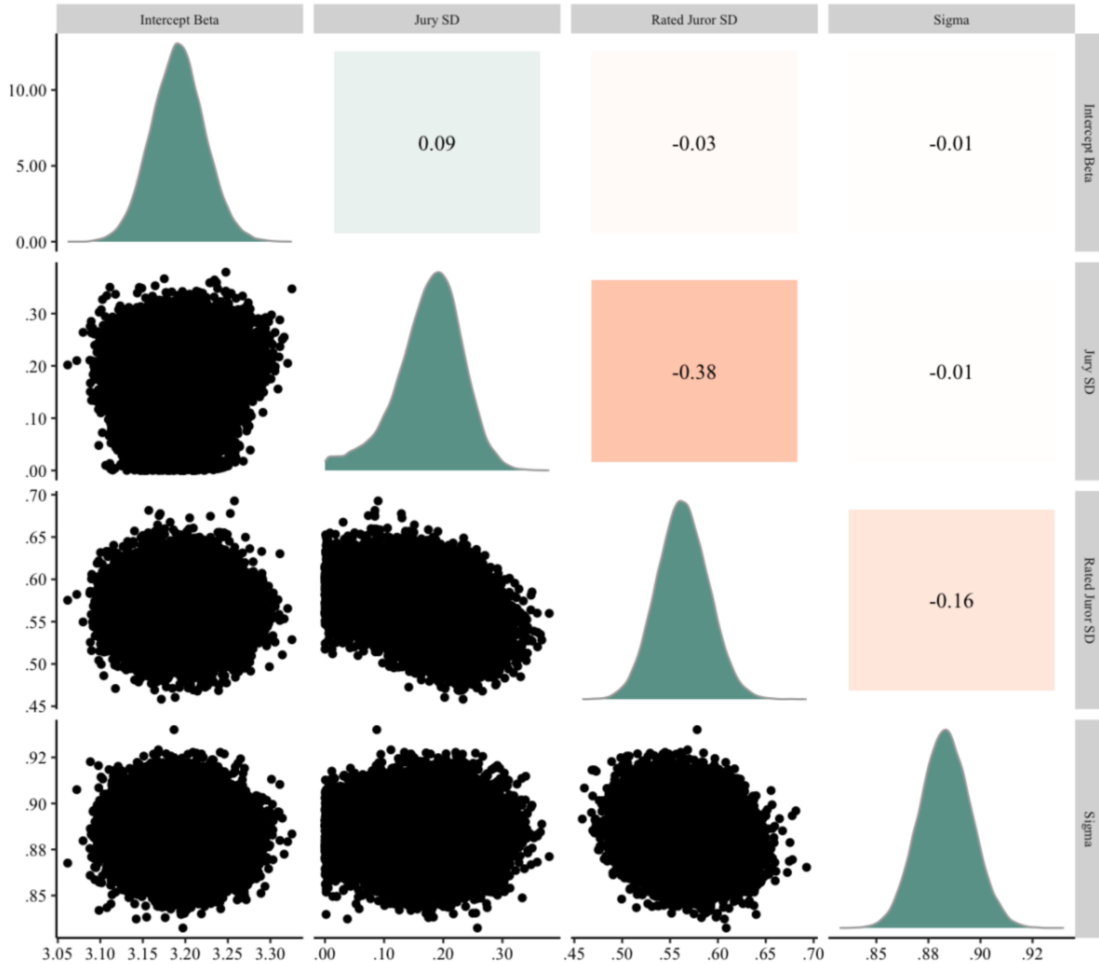


Note. A pairs plot shows the relationship between all parameters in the model. The plots on the diagonal are density plots for each of the variables and should be relatively normally distributed. The upper half of the plot shows the correlations between all variables in the model. High correlations can be an indicator of poor model fit. The lower half of the plot shows scatterplots of each relationship. Funneling, or when a scatterplot is more spread out in one portion of the graph compared to the other, can indicate that the sampler was not able to explore the entire parameter space.

Because the new model continued to have convergence problems, I removed the random effects relating to Dyad. The new model converged and the pairs plot (Figure) suggested that the model was functioning as expected. The final random effects structure was:

$$Influence \sim 1 + (1 \mid Jury) + (1 \mid Jury{:}\,RatedJuror)$$

Figure A4.
Pairs Plot for the Final Model



Note. A pairs plot shows the relationship between all parameters in the model. The plots on the diagonal are density plots for each of the variables and should be relatively normally distributed. The upper half of the plot shows the correlations between all variables in the model. High correlations can be an indicator of poor model fit. The lower half of the plot shows scatterplots of each relationship. Funneling, or when a scatterplot is more spread out in one portion of the graph compared to the other, can indicate that the sampler was not able to explore the entire parameter space.

Set Priors

I used weakly informative, regularizing priors for all models (Gill & Witko, 2013; Stan Development Team, 2023). Weakly informative, regularizing priors provide some information about the expected scale of the dependent variable while containing enough uncertainty to prevent the prior from having undue weight in the model. These priors are useful because they can prevent some over-fitting to the data and make the model more robust to outliers. I began by setting the prior for the intercept. The intercept prior specifies the expected structure of the data when there are no predictors in the model. Because influence was relatively normally distributed and has a lower-bound of one and an upper-bound of five, I set intercept priors that reflected a normal distribution where 99% of the values fall between one and five (see Figure A). This prior assumes that influence will have a mean of 3 and a standard deviation of .75 and is written as follows:

$$\beta_0 \sim N(3, .75)$$

Then, I set the priors of the predictors. When setting priors for the predictors, weakly informative, regularizing priors assume that there will be no effect of the predictors on the response variable. However, these priors should also allow for a high amount of uncertainty, to ensure that the prior does not unreasonably influence the model. Because predictor coefficients represent the change in the response variable for a one-unit change in the predictor, I used a prior at assumes a normal distribution, that there will be no change in influence for every one-unit change in any predictor, and that there is significant uncertainty in the prior (see Figure B). This prior is a standard weakly informative prior (Gill & Witko, 2013), assumes a mean of 0 and a standard deviation of 1, and is written as follows:

$$\beta_n \sim N(0, 1)$$

Finally, I set the priors for the standard deviation of the random effects (σ) and the residual standard deviation (∑). Gelman (2020) recommends using a half-t distribution that restricts large values for the random effects and residual standard deviation priors. I began by using the same prior for both the standard deviations of the random effects and the residual standard deviation. These priors assumed a half-t distribution (see Figure D) with a mean of 0, 3 degrees of freedom, and a scale parameter of 2.5 and is written as follows:

$$\sigma \sim HalfT(3,0,2.5)$$

$$\Sigma \sim HalfT(3,0,2.5)$$

However, when I examined the prior predictive distribution, given these priors, I found that the prior prediction of influence ranged from approximately -40 to 40, rather than from 1 to 5 (**Error! Reference source not found.**A). When a prior results in values that are not reasonable, a more informative prior should be used to limit the occurrence of those values. Thus, I reduced the scale parameter for the priors for the standard deviations of the random effects from 2.5 to 0.1. With this new prior, the expected posterior prediction of influence fell between 1 and 5 (**Error! Reference source not found.**B). The new prior (see Figure C) for the standard deviation of the random effects was written as follows:

$$\sigma \sim HalfT(3,0,.1)$$

The final priors that I used in all models were:

$$\beta_0 \sim N(3,.75)$$

$$\beta_n \sim N(0,1)$$

$$\sigma \sim HalfT(3,0,.1)$$

$$\Sigma \sim HalfT(3,0,2.5)$$

β_0 represents the intercept prior for every model; β_n represents the prior for all predictors in the model; σ represents the prior for the standard deviation of all random effects; and Σ represents the prior for the residual standard deviation.

Figure A5.

The Prior Distributions

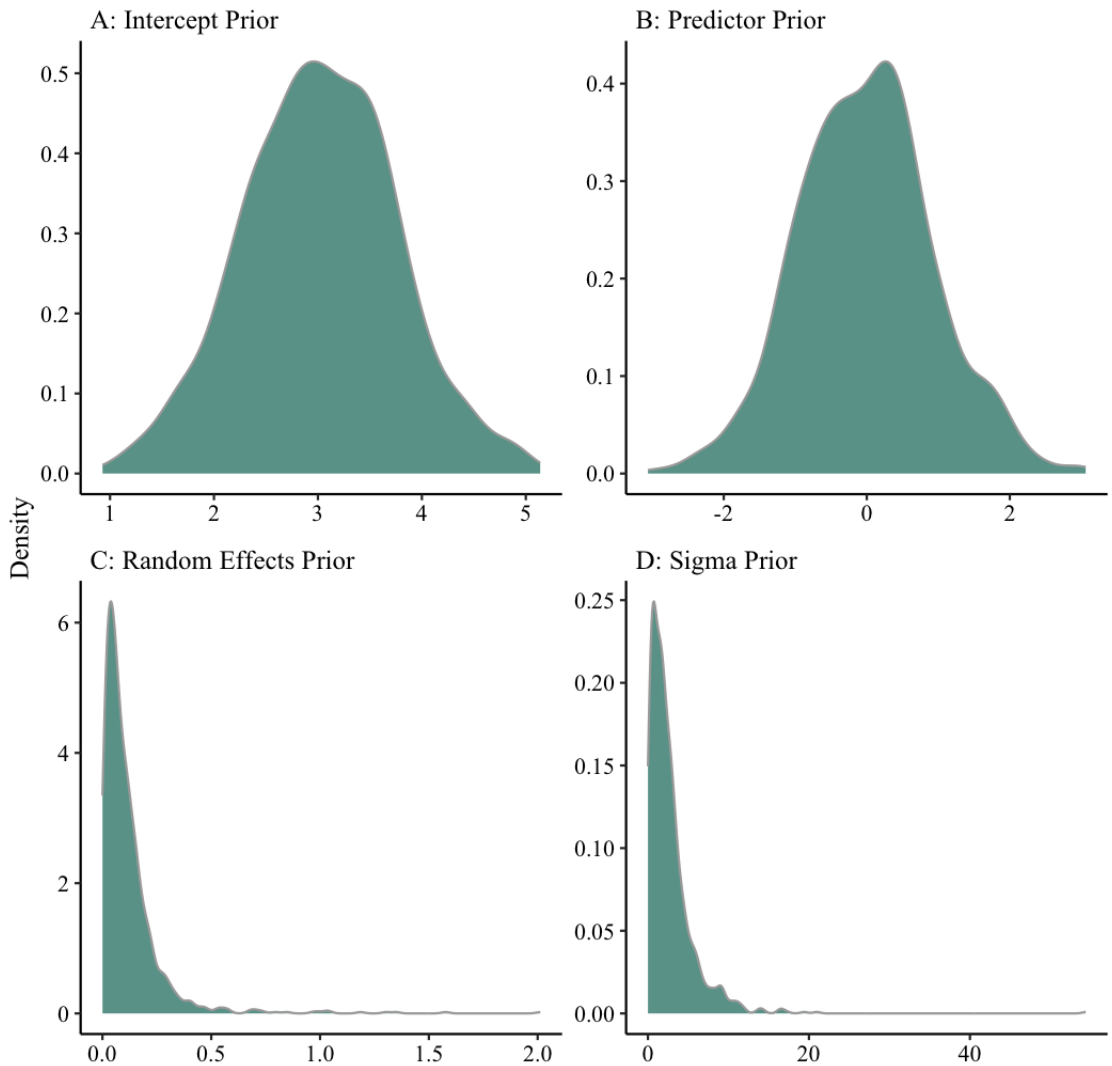
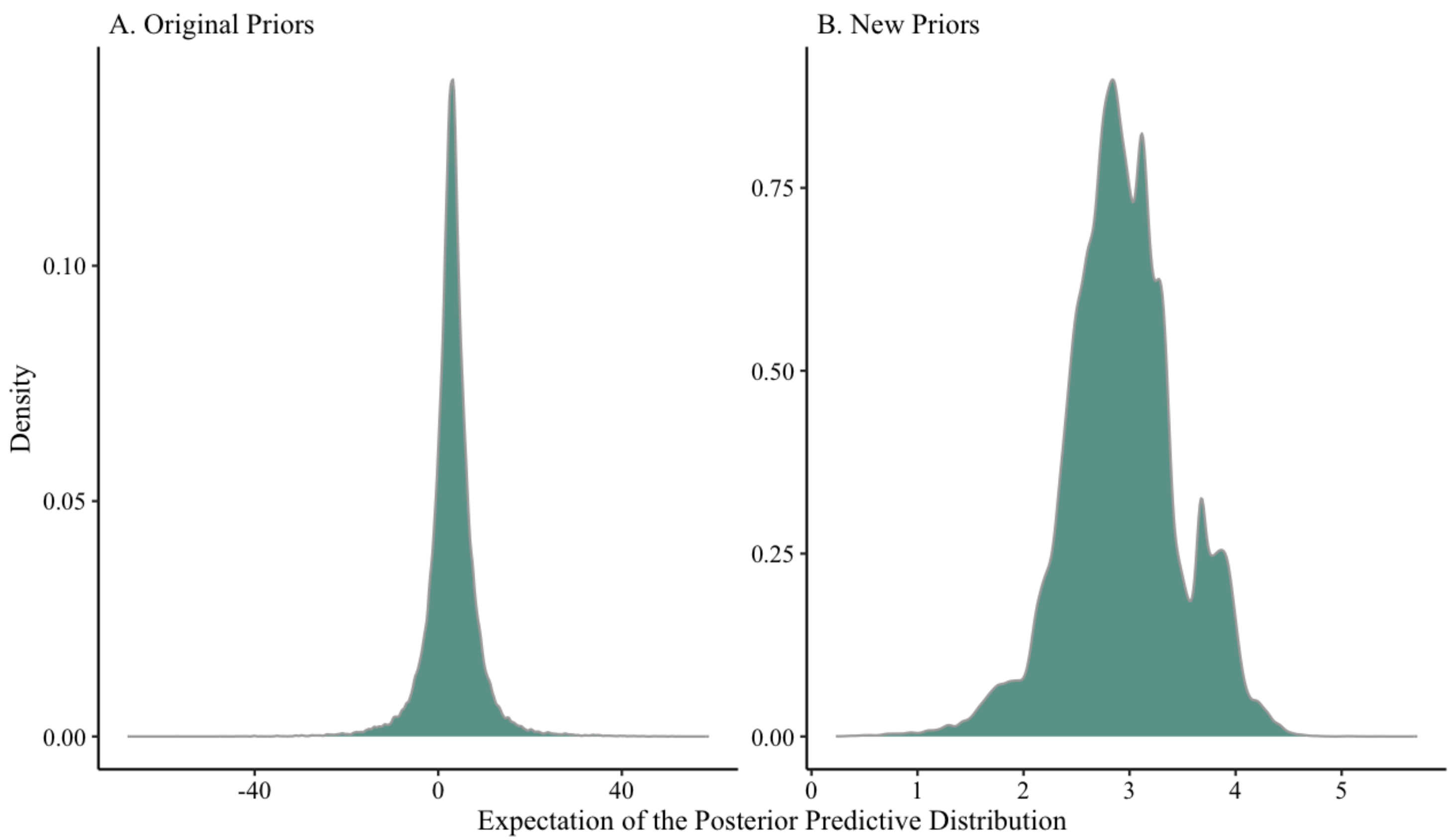


Figure A6.

The Expectation of the Prior Predictive Distribution with the Original Priors and with the Final Priors.



Note. These graphs show the expectation of the prior predictive distribution for the original priors (which uses a scale parameter of 2.5 for the standard deviation of the random effects) and the expectation of the prior predictive distribution for the final priors (which uses a scale parameter of .1 for the standard deviation of the random effects). The prior predictive distribution shows the distribution of possible posterior values, given the prior only. These graphs represent what we expect the data will look like before we observe the data.

APPENDIX F

ADDITIONAL INFORMATION ABOUT THE RAVDESS DATASET

The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) is a database of emotional speech (Livingstone & Russo, 2018). The database contains calm, happy, sad, angry, fearful, surprised, disgusted, and neutral speech. RAVDESS contains audio files from 24 English-speaking actors who recorded two emotionally neutral statements with one of the eight emotions with normal or strong intensity (except neutral speech, which was not recorded at strong intensity). Each actor said each statement twice. Thus, each actor created 60 audio files (2 statements x 2 repetitions x 2 intensities x 7 emotions and 2 statements x 2 repetitions for neutral statements). The emotional statements were then validated by 297 independent coders who correctly classified speech, on average 72% (*SD* = 27%) of the time. Table A1 contains descriptive statistics for the RAVDESS dataset.

Table A8.
Descriptive Statistics of all Continuous Predictors and Influence by all Categorical Predictors

Categorical Predictor	Category	*n* (%)	Pitch	Intensity
Gender	Female	672 (50.00%)	213.69 (17.92)	66.70 (9.51)
	Male	672 (50.00%)	157.21 (41.54)	65.45 (9.29)
Acoustic Indicators of Emotion	Calm	192 (13.33%)	159.14 (47.23)	55.7 (5.59)
	Neutral	96 (6.67%)	162.41 (46.02)	58.40 (4.63)
	Angry	192 (13.33%)	199.8 (35.12)	75.71 (8.12)
	Disgust	192 (13.33%)	175.68 (43.95)	64.33 (6.22)
	Fearful	192 (13.33%)	195.99 (39.21)	70.45 (9.19)
	Happy	192 (13.33%)	196.05 (37.37)	68.98 (7.1)
	Sad	192 (13.33%)	175.49 (45.16)	60.99 (7.84)
	Surprised	192 (13.33%)	196.01 (31.58)	66.32 (5.73)
	Not Angry	1248 (86.67)	181.47 (43.89)	64.00 (8.54)
Total		1440 (100.00%)	183.92 (43.27)	65.56 (9.37)

Note. For each continuous measure, the means are reported with the standard deviation in parentheses. Shading was done to improve the readability of the table. Not angry represents all emotions represented in the dataset apart from anger.

APPENDIX G

ANALYSIS OF ALTERNATIVE PRIORS

In order to ensure that the results were not being driven by a specific set of priors, I replicated the findings of the 7 emotion models using three additional sets of priors. One set of priors was more vague than the priors that I used in all analyses, one set of priors was stronger than the priors that I used in all analyses, and one set of priors was weakly biased towards the alternative hypothesis. The following are each set of priors:

<u>Vague Priors</u>

$$\beta_0 \sim N(3, .75)$$

$$\beta_n \sim N(0, 100)$$

$$\sigma \sim HalfT(3,0, .1)$$

$$\Sigma \sim HalfT(3,0, 2.5)$$

<u>Strong Priors</u>

$$\beta_0 \sim N(3, .75)$$

$$\beta_n \sim N(0, .25)$$

$$\sigma \sim HalfT(3,0, .1)$$

$$\Sigma \sim HalfT(3,0, 2.5)$$

<u>Alternative Priors</u>

$$\beta_0 \sim N(3, .75)$$

$$\beta_n \sim N(1, 1)$$

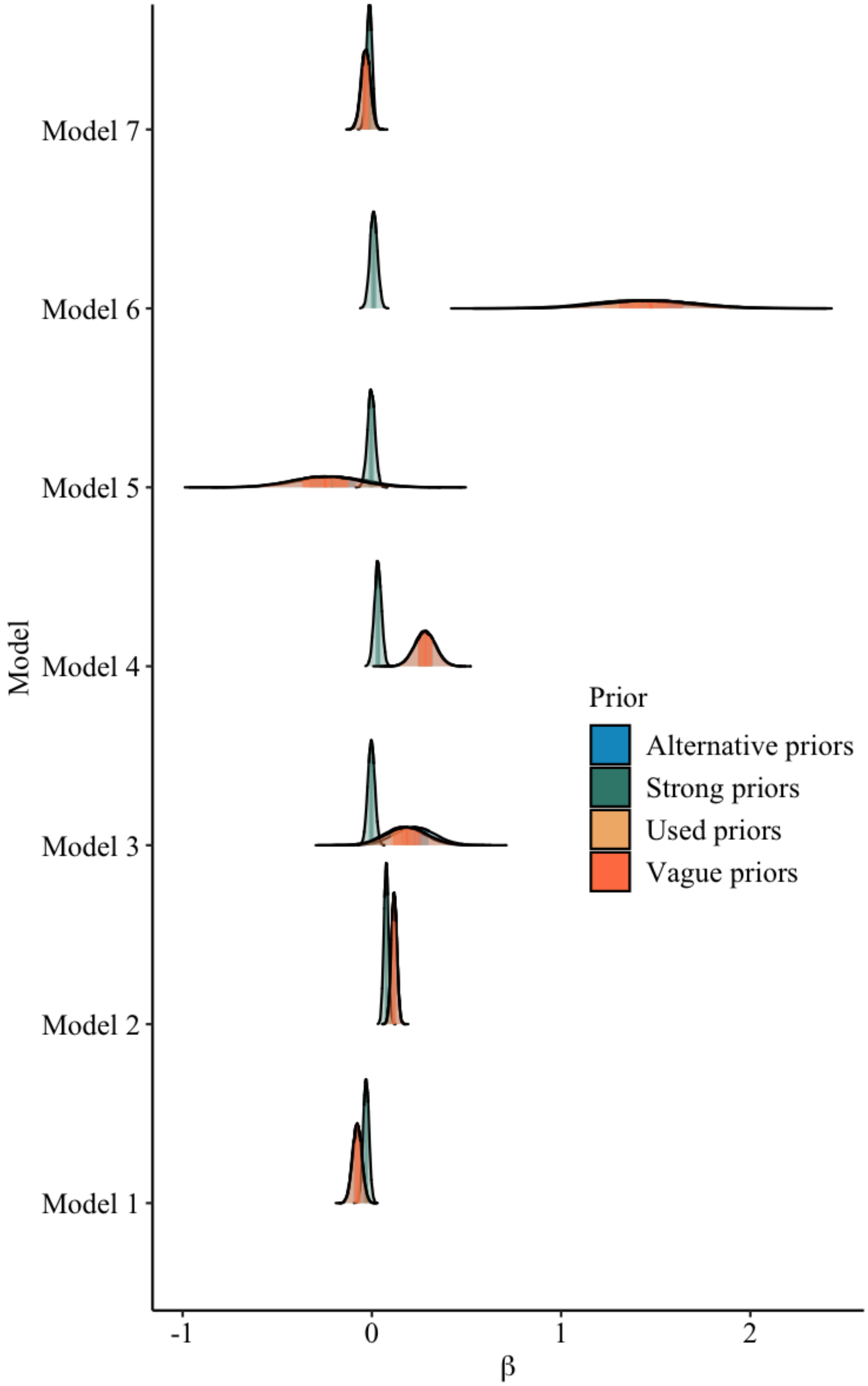
$$\sigma \sim HalfT(3,0, .1)$$

$$\Sigma \sim HalfT(3,0, 2.5)$$

The posterior distributions of the models were generally similar regardless of the priors that were used, and variations generally occurred under the vague or strong priors (Figure A7).

Figure A7.

Posterior Distributions of All Priors



Note. This graph shows the posterior distributions for all seven emotions models across all priors. If the density graphs do not overlap, that is an indication that the model is being unduly influenced by the priors. All models that used the priors used in the main analyses did not differ from analyses using other priors.

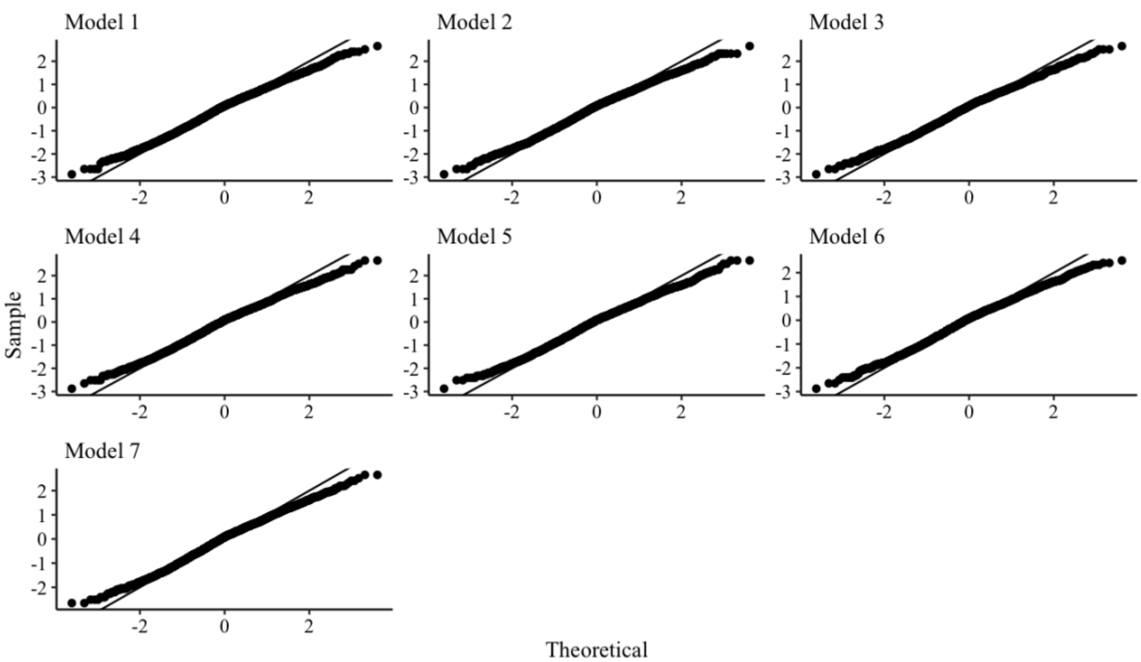
APPENDIX H

ADDITIONAL ANALYSES RELATING TO MODEL PERFORMANCE

After running the models that relate to Hypothesis 1, I examined whether the models were functioning properly by examining the residual plots and ensuring that all chains mixed. All of the residuals were slightly light-tailed but overall, reasonably consistent with normality (Figure A8). Trace plots of parameters (Figure A9) indicate that all chains were mixed.

Figure A8.

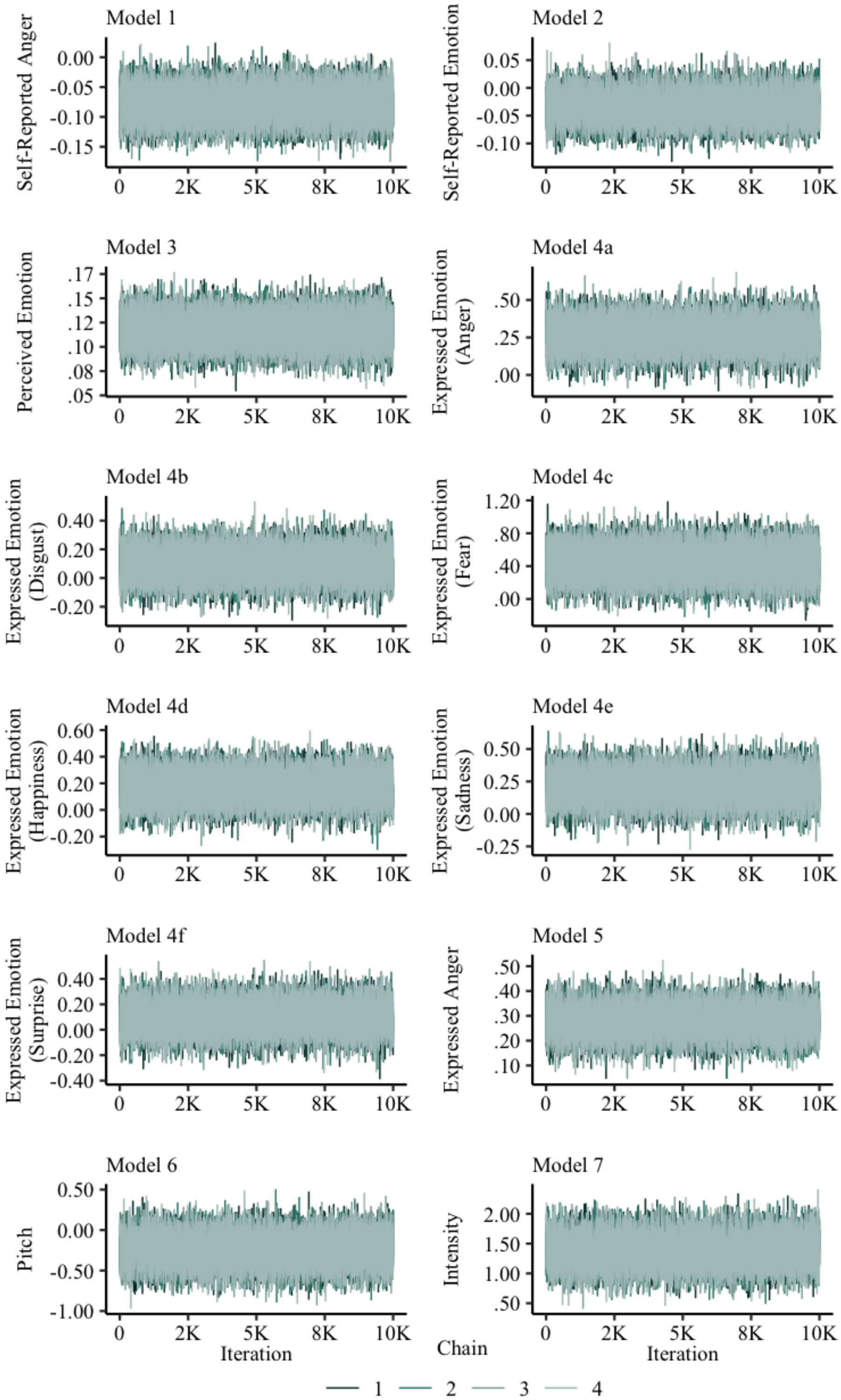
Q–Q Plots for All Models



Note. These graphs show Q-Q plots of the residuals for each model. The y-axis is quantiles from the sample and the x-axis is quantiles from the theoretical normal distribution. The line shows where points should fall if data is normally distributed. Because most points fall along that line, we can conclude that the residuals are normally distributed.

Figure A9.

Trace Plots of Each Hypothesis 1 Models



Note. These graphs show the trace plots of the four parallel Markov chains for each predictor in all seven models. Chains are mixing properly and stable when trace plots look like white noise with no discernible pattern.

I conducted the same examination of model fit for the models that I ran for Hypothesis 2. Again, all of the residuals were slightly light-tailed but overall, reasonably consistent with normality (Figures A10–A12). Trace plots of parameters (Figures A13–A15) indicate that all chains were mixed.

Figure A10.
Q–Q Plots for All Models that Include 2-Way Interactions with Gender

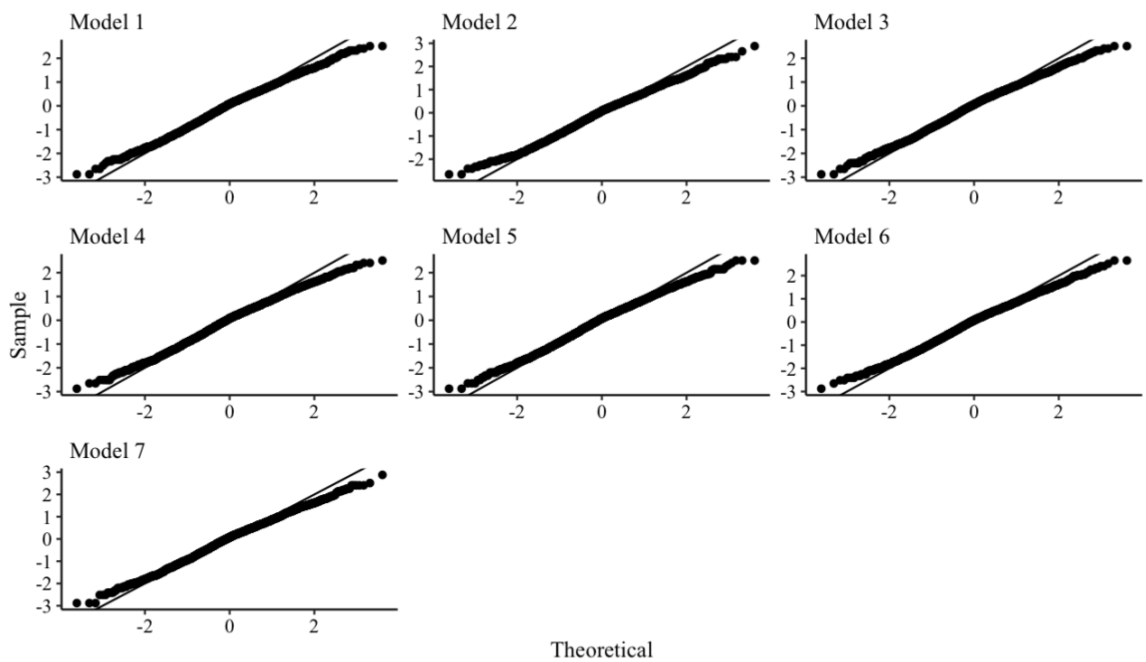


Figure A11.

Q–Q Plots for All Models that Include 2-Way Interactions with Race

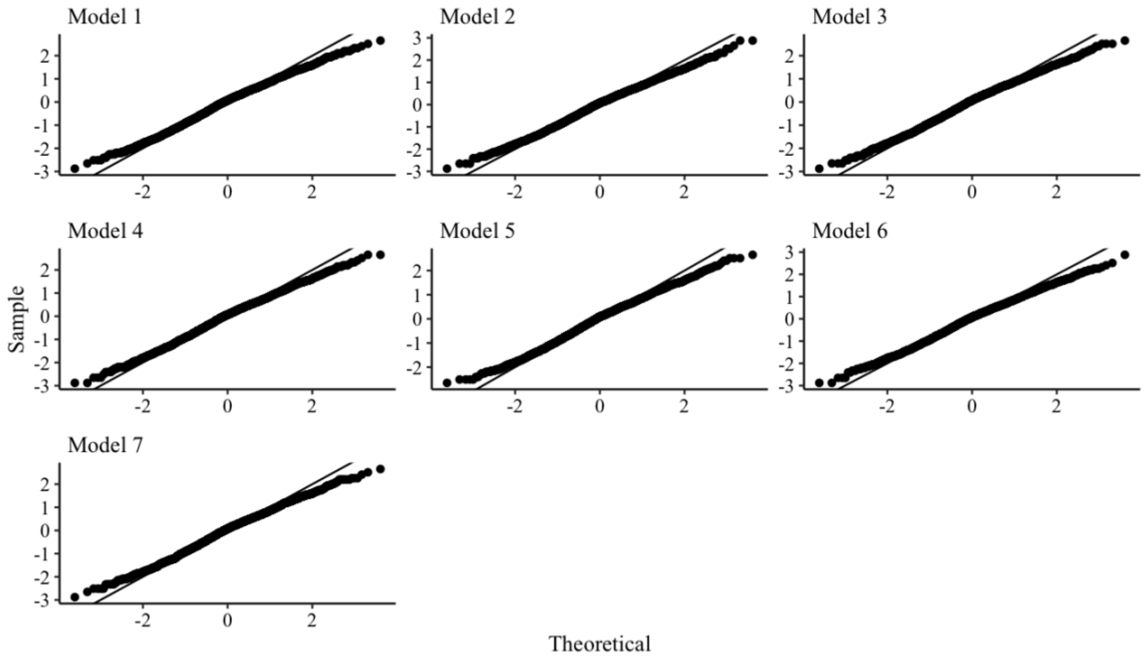


Figure A12.

Q–Q Plots for All Models that Include 3-Way Interactions with Gender and Race

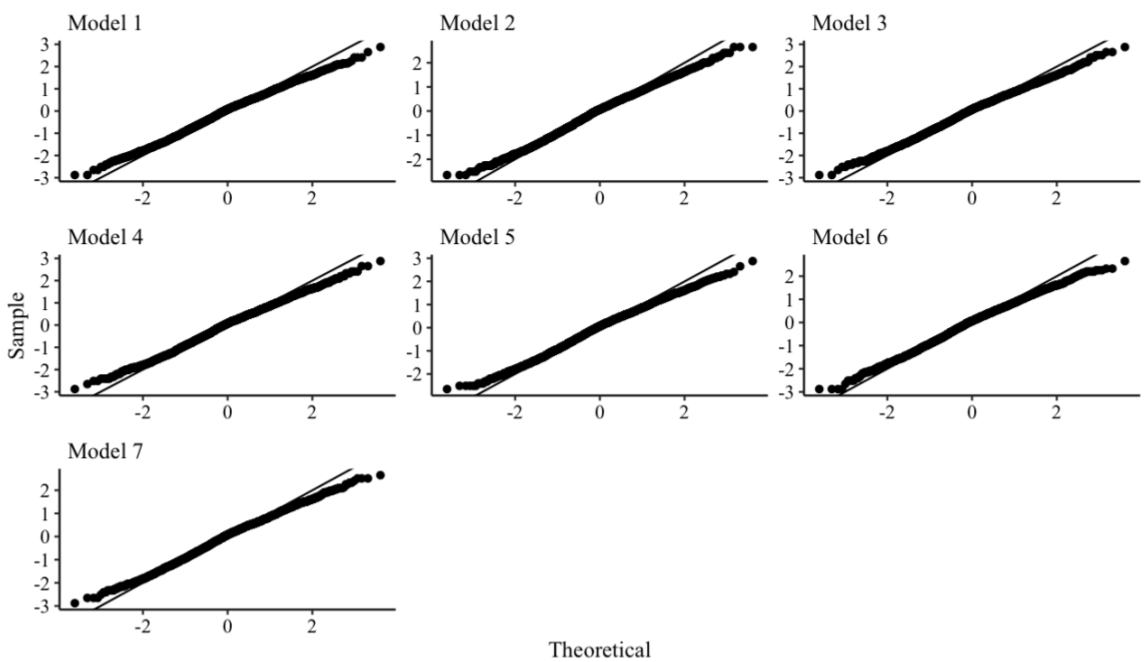


Figure A13.

Trace Plots of the Emotion Predictors in Models that Involve 2-Way Interactions with Gender

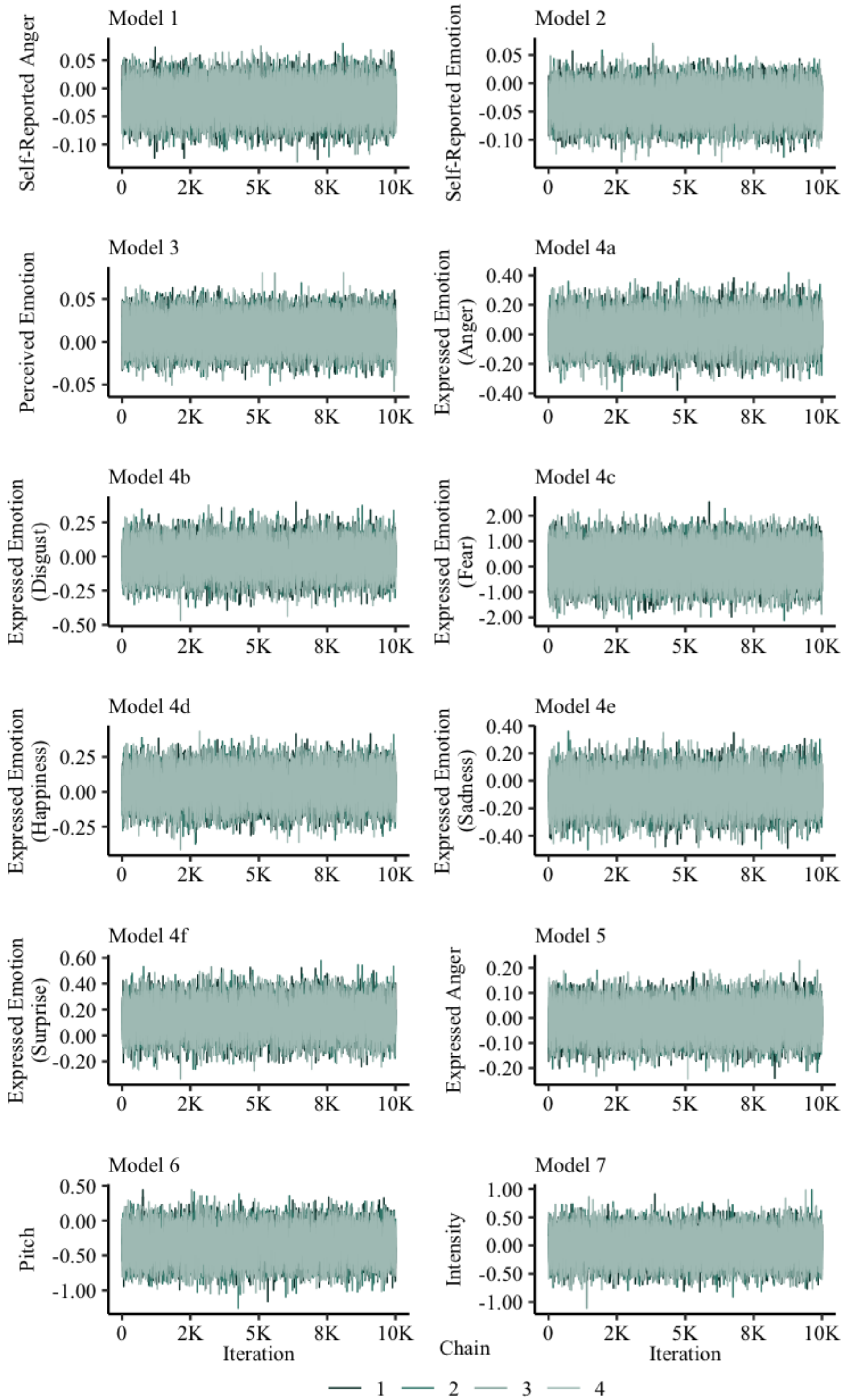


Figure A14.

Trace Plots of the Emotion Predictors in Models that Involve 2-Way Interactions with Race

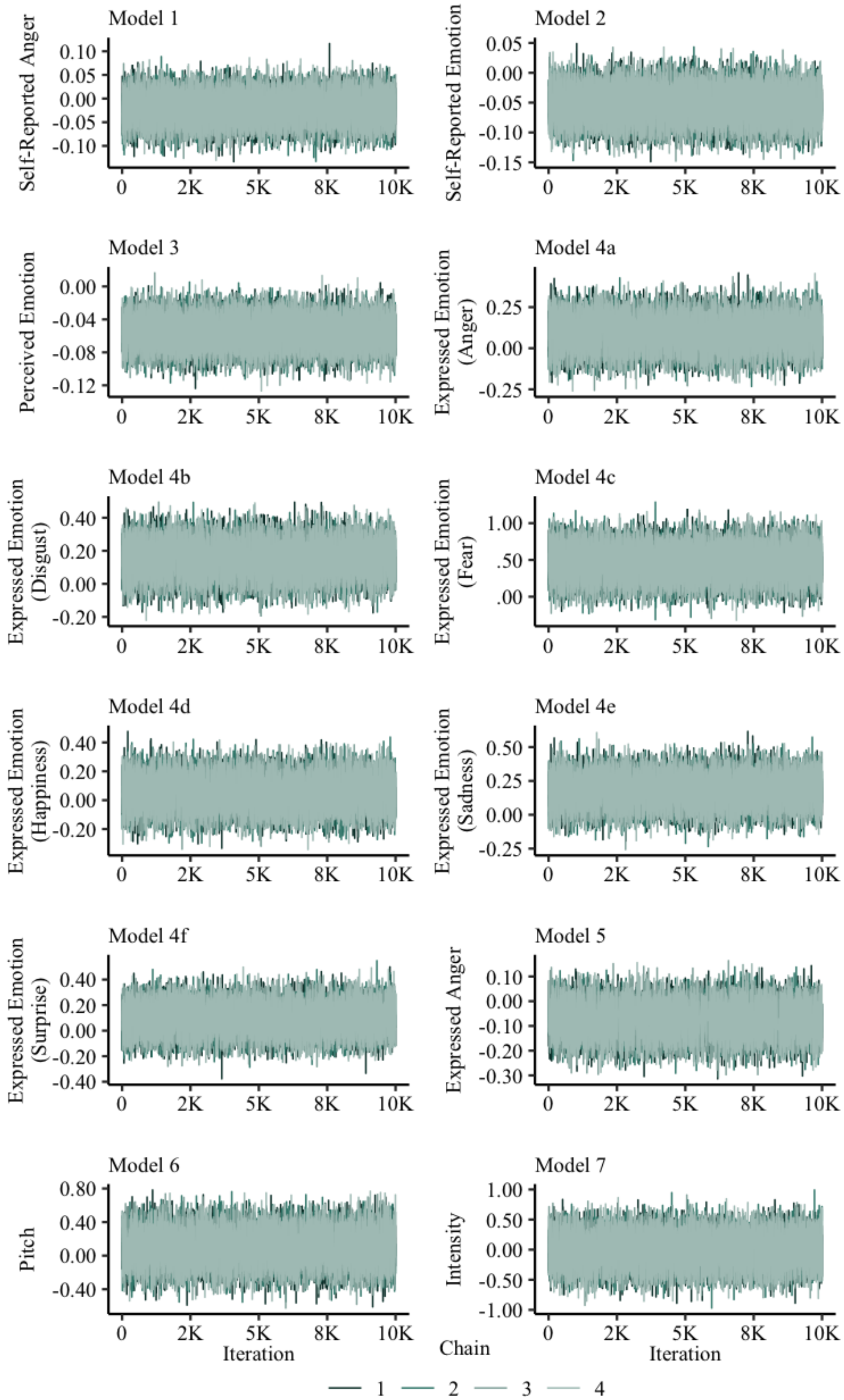
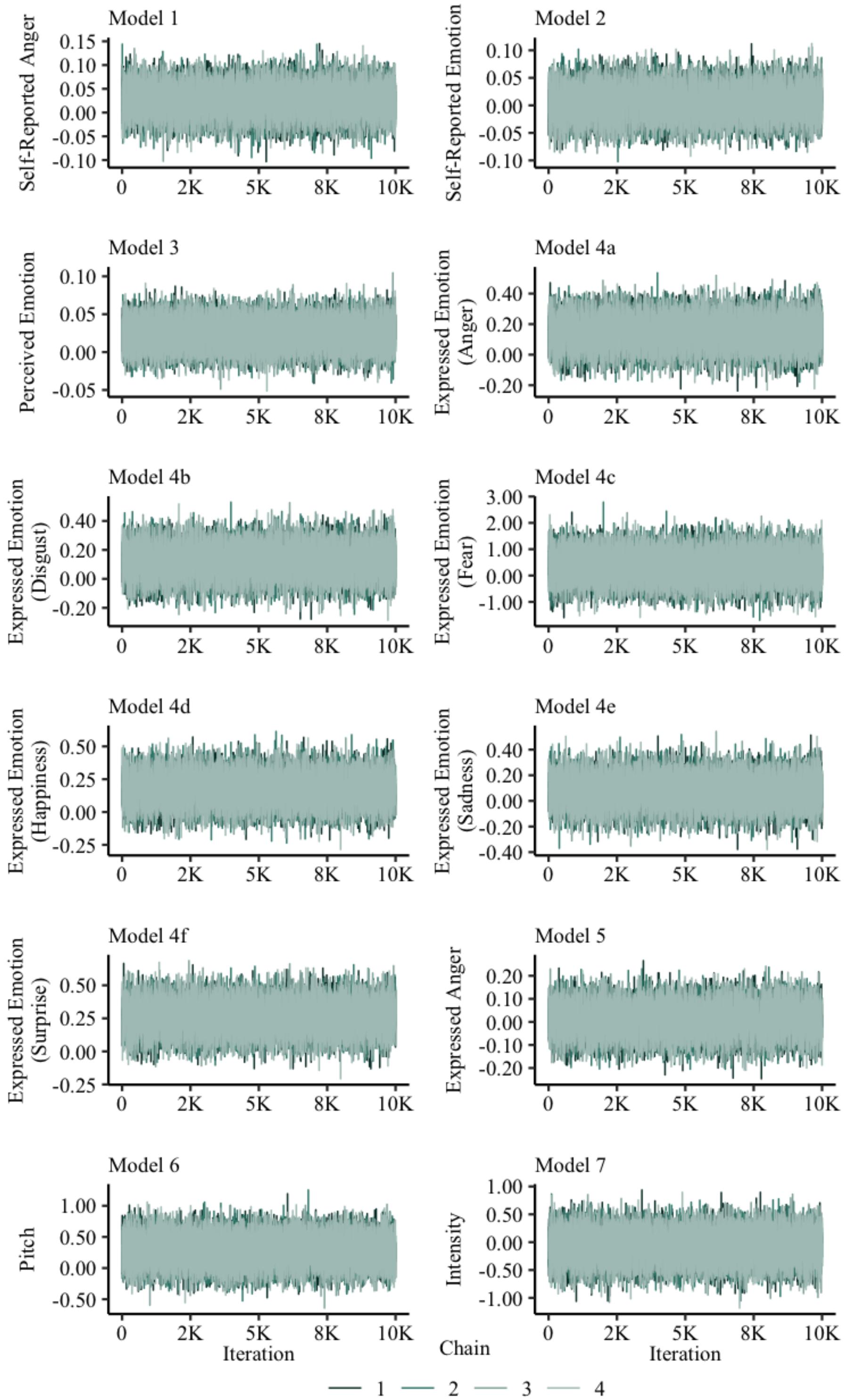


Figure A15.

Trace Plots of the Emotion Predictors in Models that Involve 3-Way Interactions with Gender and Race



APPENDIX I

RESULTS FROM THE LESS OPTIMAL MODELS

The following tables (A2–A6) are the Bayes factors and change in elpd for all models that included self-reported anger, self-reported emotion, expressed emotion, expressed anger, and mean pitch. The models that included expressed anger were more supported by the data than the Intercept-Only model, but these models had poor out-of-sample prediction. All other models performed worse than the Intercept-Only model on both metrics, suggesting that the models were 1) not supported, given the data and 2) not predictive on new data. Bayes factors that are greater than one indicate that there is more support for the numerator model than the denominator model. Positive Δelpd values indicate that the comparison group has better predictive accuracy than the reference group. Shading was done to improve the readability of the tables.

Table A2.

Bayes Factors and Δelpd and Δsd for All Model 1 Pairwise Comparisons

	Bayes Factors				Δelpd (2×*Δsd*)			
	1	1A	1B	1C	1	1A	1B	1C
0	0.50	0.001	0.04	$2.76×10^{-7}$	-3.31(4.23)	-2.13 (4.83)	-3.85 (5.74)	-3.85 (6.59)
1	1.00	0.003	0.08	$5.57×10^{-7}$	0	1.18 (2.59)	-0.54 (4.08)	-0.54 (5.28)
1A	336.69	1.00	26.70	0.0002	-1.18 (2.59)	0	-1.72 (4.90)	-1.71 (4.63)
1B	12.61	0.04	1.00	$7.03×10^{-6}$	0.54 (4.08)	1.72 (4.90)	0	0
1C	$1.79×10^{6}$	5327.90	$1.42×10^{5}$	1.00	0.54 (5.28)	1.71 (4.63)	0	0
2	168.91	0.50	13.39	$9.42×10^{-5}$	-3.00 (3.88)	-1.82 (4.48)	-3.54 (5.54)	-3.53 (6.35)
2A	8747.32	25.98	693.56	0.005	-2.54 (4.86)	-1.36 (3.89)	-3.08 (6.42)	-3.08 (6.04)
2B	101.16	0.30	8.02	$5.64×10^{-5}$	-1.38 (6.17)	-0.20 (6.65)	-1.92 (4.67)	-1.91 (5.71)
2C	$5.75×10^{6}$	17082.28	$4.56×10^{5}$	3.21	-1.17 (7.11)	0.01 (6.49)	-1.71 (5.91)	-1.70 (5.00)
3	$3.55×10^{-11}$	$1.05×10^{-13}$	$2.81×10^{-12}$	$1.98×10^{-17}$	25.78 (17.64)	26.96 (18.00)	25.24 (17.89)	25.24 (18.44)
3A	$2.04×10^{-10}$	$6.05×10^{-13}$	$1.62×10^{-11}$	$1.14×10^{-16}$	26.50 (18.17)	27.68 (18.06)	25.96 (18.44)	25.96 (18.51)
3B	$1.88×10^{-12}$	$5.60×10^{-15}$	$1.49×10^{-13}$	$1.05×10^{-18}$	31.56 (19.87)	32.74 (20.22)	31.02 (19.29)	31.02 (19.89)
3C	$2.99×10^{-8}$	$8.87×10^{-11}$	$2.37×10^{-9}$	$1.66×10^{-14}$	31.12 (20.57)	32.30 (20.48)	30.58 (20.03)	30.59 (20.04)
4	2.94	0.009	0.23	$1.64×10^{-6}$	1.75 (8.25)	2.93 (8.39)	1.21 (8.92)	1.22 (9.23)
4A	$6.04×10^{5}$	1793.38	47874.79	0.34	1.34 (9.04)	2.52 (8.62)	0.80 (9.80)	0.80 (9.59)
4B	$1.20×10^{5}$	357.51	9543.73	0.07	2.21 (9.44)	3.39 (9.60)	1.67 (8.78)	1.67 (9.09)
4C	$8.47×10^{15}$	$2.52×10^{13}$	$6.72×10^{14}$	$4.72×10^{9}$	2.22 (10.74)	3.40 (10.38)	1.68 (10.13)	1.69 (9.73)
5	$5.07×10^{-5}$	$1.51×10^{-6}$	$4.02 ×10^{-5}$	$2.83×10^{-10}$	1.57 (7.46)	2.75 (7.80)	1.03 (8.18)	1.03 (8.72)
5A	0.02	$4.96 ×10^{-5}$	0.001	$9.32×10^{-9}$	1.56 (7.73)	2.73 (7.27)	1.02 (8.50)	1.02 (8.26)
5B	0.004	$1.26 ×10^{-5}$	0.0003	$2.36×10^{-9}$	1.50 (8.45)	2.68 (8.75)	0.96 (7.61)	0.96 (8.12)
5C	34.38	0.10	2.73	$1.92×10^{-5}$	2.11 (8.77)	3.29 (8.33)	1.57 (7.97)	1.58 (7.39)
6	25.24	0.07	2.00	$1.41×10^{-5}$	-3.64 (4.38)	-2.46 (4.55)	-4.18 (5.91)	-4.18 (6.44)
6A	99.73	0.30	7.91	$5.56×10^{-5}$	-2.45 (5.10)	-1.28 (4.33)	-2.99 (6.45)	-2.99 (6.20)
6B	32.73	0.10	2.59	$1.82×10^{-5}$	-2.30 (6.09)	-1.12 (6.22)	-2.84 (4.72)	-2.84 (5.21)
6C	5890.90	17.50	467.08	0.003	-1.36 (6.90)	-0.18 (6.36)	-1.90 (5.74)	-1.90 (5.06)
7	$8.04×10^{-7}$	$2.39×10^{-9}$	$6.38×10^{-8}$	$4.48×10^{-13}$	4.43 (8.02)	5.61 (8.27)	3.89 (8.74)	3.89 (9.14)
7A	$7.74×10^{-6}$	$2.30×10^{-8}$	$6.14×10^{-7}$	$4.32×10^{-12}$	4.66 (8.32)	5.84 (7.85)	4.12 (9.08)	4.12 (8.78)
7B	$4.29×10^{-6}$	$1.27×10^{-8}$	$3.40×10^{-7}$	$2.39×10^{-12}$	4.83 (8.87)	6.01 (9.13)	4.29 (7.99)	4.29 (8.42)
7C	0.006	$1.90×10^{-5}$	0.0005	$3.56×10^{-9}$	4.38 (9.15)	5.56 (8.72)	3.84 (8.34)	3.84 (7.77)

Table A3.

Bayes Factors and Δelpd and Δsd for All Model 2 Pairwise Comparisons

	Bayes Factors				Δelpd (2×*Δsd)*			
	2	2A	2B	2C	2	2A	2B	2C
0	0.003	5.66×10^{-5}	0.005	8.61×10^{-8}	-0.31 (1.86)	-0.85 (3.57)	-0.77 (3.74)	-2.14 (6.23)
1	0.006	0.0001	0.01	1.74×10^{-7}	3.00 (3.88)	2.45 (5.10)	2.54 (4.86)	1.17 (7.11)
1A	1.99	0.04	3.33	5.85×10^{-5}	1.82 (4.48)	1.28 (4.33)	1.36 (3.89)	-0.01 (6.49)
1B	0.07	0.001	0.12	2.19×10^{-6}	3.54 (5.54)	2.99 (6.45)	3.08 (6.42)	1.71 (5.91)
1C	10620.06	205.07	17733.41	0.31	3.53 (6.35)	2.99 (6.20)	3.08 (6.04)	1.70 (5.00)
2	1.00	0.02	1.67	2.94×10^{-5}	0	-0.54 (3.83)	-0.46 (3.29)	-1.83 (6.01)
2A	51.79	1.00	86.47	0.002	0.46 (3.29)	-0.09 (3.15)	0	-1.37 (5.16)
2B	0.60	0.01	1.00	1.76×10^{-5}	1.62 (4.73)	1.08 (6.11)	1.16 (5.99)	-0.21 (3.90)
2C	34049.97	657.50	56856.76	1.00	1.83 (6.01)	1.29 (5.90)	1.37 (5.16)	0
3	2.10×10^{-13}	4.06×10^{-15}	3.51×10^{-13}	6.17×10^{-18}	28.77 (17.21)	28.23 (17.66)	28.32 (17.71)	26.94 (18.34)
3A	1.21×10^{-12}	2.33×10^{-14}	2.01×10^{-12}	3.54×10^{-17}	29.50 (17.84)	28.95 (17.74)	29.04 (17.78)	27.67 (18.43)
3B	1.12×10^{-14}	2.15×10^{-16}	1.86×10^{-14}	3.28×10^{-19}	34.56 (19.41)	34.02 (19.81)	34.10 (19.95)	32.73 (19.77)
3C	1.77×10^{-10}	3.41×10^{-12}	2.95×10^{-10}	5.19×10^{-15}	34.12 (20.19)	33.58 (20.10)	33.66 (20.20)	32.29 (19.97)
4	0.02	0.0003	0.03	5.11×10^{-7}	4.75 (7.55)	4.21 (7.48)	4.29 (7.79)	2.92 (8.86)
4A	3574.74	69.03	5969.11	0.10	4.33 (8.54)	3.79 (7.88)	3.88 (8.10)	2.50 (9.35)
4B	712.62	13.76	1189.93	0.02	5.21 (8.84)	4.66 (8.74)	4.75 (9.14)	3.38 (8.73)
4C	5.02×10^{13}	9.69×10^{11}	8.38×10^{13}	1.47×10^{9}	5.22 (10.20)	4.68 (9.54)	4.76 (9.88)	3.39 (9.43)
5	3.00×10^{-6}	5.80×10^{-8}	5.01×10^{-6}	8.82×10^{-11}	4.56 (6.53)	4.02 (6.80)	4.11 (7.13)	2.73 (8.37)
5A	9.89×10^{-5}	1.91×10^{-6}	0.0002	2.91×10^{-9}	4.55 (7.01)	4.01 (6.23)	4.10 (6.56)	2.72 (7.95)
5B	2.50×10^{-5}	4.83×10^{-7}	4.18×10^{-5}	7.35×10^{-10}	4.49 (7.53)	3.95 (7.70)	4.04 (8.12)	2.66 (7.69)
5C	0.20	0.004	0.34	5.98×10^{-6}	5.11 (7.99)	4.57 (7.22)	4.65 (7.63)	3.28 (6.99)
6	0.15	0.003	0.25	4.39×10^{-6}	-0.64 (2.30)	-1.18 (3.26)	-1.10 (3.42)	-2.47 (6.12)
6A	0.59	0.01	0.99	1.73×10^{-5}	0.54 (3.83)	0	0.09 (3.15)	-1.29 (5.90)
6B	0.19	0.004	0.32	5.69×10^{-6}	0.69 (4.59)	0.15 (5.07)	0.24 (5.39)	-1.14 (4.84)
6C	34.88	0.67	58.24	0.001	1.63 (5.87)	1.09 (4.43)	1.18 (5.56)	-0.20 (4.72)
7	4.76×10^{-9}	9.20×10^{-11}	7.95×10^{-9}	1.40×10^{-13}	7.42 (7.24)	6.88 (7.38)	6.97 (7.61)	5.59 (8.79)
7A	4.58×10^{-8}	8.85×10^{-10}	7.65×10^{-8}	1.35×10^{-12}	7.65 (7.69)	7.11 (6.92)	7.20 (7.15)	5.82 (8.45)
7B	2.54×10^{-8}	4.90×10^{-10}	4.24×10^{-8}	7.45×10^{-13}	7.83 (8.06)	7.28 (8.20)	7.37 (8.51)	6.00 (8.05)
7C	3.79×10^{-5}	7.31×10^{-7}	6.32×10^{-5}	1.11×10^{-9}	7.38 (8.48)	6.84 (7.77)	6.92 (8.07)	5.55 (7.49)

Table A4.

Bayes Factors and Δelpd and Δsd for All Model 4 Pairwise Comparisons

	Bayes Factors				Δelpd (2×*Δsd)*			
	4	4A	4B	4C	4	4A	4B	4C
0	0.17	8.20×10^{-7}	4.11×10^{-6}	5.84×10^{-17}	0.17	-5.06 (7.29)	-4.65 (8.36)	-5.52 (8.57)
1	0.34	1.66×10^{-6}	8.31×10^{-6}	1.18×10^{-16}	0.34	-1.75 (8.25)	-1.34 (9.04)	-2.21 (9.44)
1A	114.60	0.0006	0.003	3.97×10^{-14}	114.60	-2.93 (8.39)	-2.52 (8.62)	-3.39 (9.60)
1B	4.29	2.09×10^{-5}	0.0001	1.49×10^{-15}	4.29	-1.21 (8.92)	-0.80 (9.80)	-1.67 (8.78)
1C	6.11×10^{5}	2.97	14.90	2.12×10^{-10}	6.11×10^{5}	-1.22 (9.23)	-0.80 (9.59)	-1.67 (9.09)
2	57.49	0.0003	0.001	1.99×10^{-14}	57.49	-4.75 (7.55)	-4.33 (8.54)	-5.21 (8.84)
2A	2977.48	0.01	0.07	1.03×10^{-12}	2977.48	-4.29 (7.79)	-3.88 (8.10)	-4.75 (9.14)
2B	34.43	0.0002	0.0008	1.19×10^{-14}	34.43	-3.13 (8.60)	-2.71 (9.67)	-3.59 (8.44)
2C	1.96×10^{6}	9.53	47.78	6.79×10^{-10}	1.96×10^{6}	-2.92 (8.86)	-2.50 (9.35)	-3.38 (8.73)
3	1.21×10^{-11}	5.88×10^{-17}	2.95×10^{-16}	4.19×10^{-27}	1.21×10^{-11}	24.03 (18.44)	24.44 (19.02)	23.57 (18.79)
3A	6.94×10^{-11}	3.37×10^{-16}	1.69×10^{-15}	2.41×10^{-26}	6.94×10^{-11}	24.75 (18.88)	25.16 (19.05)	24.29 (19.27)
3B	6.42×10^{-13}	3.12×10^{-18}	1.57×10^{-17}	2.22×10^{-28}	6.42×10^{-13}	29.81 (20.42)	30.22 (21.06)	29.35 (20.21)
3C	1.02×10^{-8}	4.95×10^{-14}	2.48×10^{-13}	3.52×10^{-24}	1.02×10^{-8}	29.37 (21.03)	29.78 (21.30)	28.91 (20.87)
4	1.00	4.87×10^{-6}	2.44×10^{-5}	3.47×10^{-16}	1.00	0	0.41 (4.23)	-0.46 (4.73)
4A	2.06×10^{5}	1.00	5.02	7.13×10^{-11}	2.06×10^{5}	-0.41 (4.23)	0	-0.87 (6.49)
4B	40971.73	0.20	1.00	1.42×10^{-11}	40971.73	0.46 (4.73)	0.87 (6.49)	0
4C	2.88×10^{15}	1.40×10^{10}	7.04×10^{10}	1.00	2.88×10^{15}	0.47 (7.03)	0.88 (6.05)	0.01 (5.42)
5	0002	8.40×10^{-10}	4.21×10^{-9}	5.99×10^{-20}	0002	-0.18 (3.46)	0.23 (5.45)	-0.64 (5.56)
5A	0.006	2.77×10^{-8}	1.39×10^{-7}	1.97×10^{-18}	0.006	-0.20 (3.90)	0.22 (4.65)	-0.65 (5.95)
5B	0.001	7.00×10^{-9}	3.51×10^{-8}	4.99×10^{-19}	0.001	-0.25 (4.99)	0.16 (6.74)	-0.71 (4.49)
5C	11.70	5.69×10^{-5}	0.0003	4.06×10^{-15}	11.70	0.36 (5.45)	0.77 (6.24)	-0.10 (5.03)
6	8.59	4.18×10^{-5}	0.0002	2.98×10^{-15}	8.59	-5.39 (7.69)	-4.98 (8.45)	-5.85 (8.89)
6A	33.95	0.0002	0.0008	1.18×10^{-14}	33.95	-4.21 (7.48)	-3.79 (7.88)	-4.66 (8.74)
6B	11.14	5.42×10^{-5}	0.0003	3.86×10^{-15}	11.14	-4.05 (8.45)	-3.64 (9.31)	-4.51 (8.24)
6C	2005.19	0.01	0.05	6.95×10^{-13}	2005.19	-3.11 (8.40)	-2.70 (8.97)	-3.57 (8.19)
7	2.74×10^{-7}	1.33×10^{-12}	6.68×10^{-12}	9.49×10^{-23}	2.74×10^{-7}	2.68 (6.28)	3.09 (7.43)	2.22 (7.63)
7A	2.63×10^{-6}	1.28×10^{-11}	6.43×10^{-11}	9.14×10^{-22}	2.63×10^{-6}	2.91 (6.60)	3.32 (6.94)	2.45 (7.97)
7B	1.46×10^{-6}	7.10×10^{-12}	3.56×10^{-11}	5.06×10^{-22}	1.46×10^{-6}	3.08 (7.12)	3.49 (8.30)	2.62 (6.98)
7C	0.002	1.06×10^{-8}	5.31×10^{-8}	7.55×10^{-19}	0.002	2.63 (7.50)	3.04 (7.94)	2.17 (7.38)

Table A5.

Bayes Factors and Δelpd and Δsd for All Model 5 Pairwise Comparisons

	Bayes Factors				Δelpd (2×*Δsd)*			
	5	5A	5B	5C	5	5A	5B	5C
0	976.03	29.62	117.05	0.01	-4.88 (6.28)	-4.87 (6.87)	-4.81 (7.25)	-5.42 (7.82)
1	1971.73	59.83	236.46	0.03	-1.57 (7.46)	-1.56 (7.73)	-1.50 (8.45)	-2.11 (8.77)
1A	6.64×10^{5}	20144.91	79614.52	9.79	-2.75 (7.80)	-2.73 (7.27)	-2.68 (8.75)	-3.29 (8.33)
1B	24867.95	754.63	2982.35	0.37	-1.03 (8.18)	-1.02 (8.50)	-0.96 (7.61)	-1.57 (7.97)
1C	3.54×10^{9}	1.07×10^{8}	4.24×10^{8}	52183.29	-1.03 (8.72)	-1.02 (8.26)	-0.96 (8.12)	-1.58 (7.39)
2	3.33×10^{5}	10106.35	39941.22	4.91	-4.56 (6.53)	-4.55 (7.01)	-4.49 (7.53)	-5.11 (7.99)
2A	1.72×10^{7}	5.23×10^{5}	2.07×10^{6}	254.46	-4.11 (7.13)	-4.10 (6.56)	-4.04 (8.12)	-4.65 (7.63)
2B	1.99×10^{5}	6052.42	23919.72	2.94	-2.94 (7.76)	-2.93 (8.27)	-2.87 (7.06)	-3.49 (7.65)
2C	1.13×10^{10}	3.44×10^{8}	1.36×10^{9}	1.67×10^{5}	-2.73 (8.37)	-2.72 (7.95)	-2.66 (7.69)	-3.28 (6.99)
3	6.99×10^{-8}	2.12×10^{-9}	8.39×10^{-9}	1.03×10^{-12}	24.21 (18.03)	24.22 (18.48)	24.28 (18.30)	23.66 (18.77)
3A	4.02×10^{-7}	1.22×10^{-8}	4.82×10^{-8}	5.93×10^{-12}	24.93 (18.62)	24.94 (18.52)	25.00 (18.90)	24.39 (18.82)
3B	3.72×10^{-9}	1.13×10^{-10}	4.46×10^{-10}	5.48×10^{-14}	29.99 (20.04)	0	30.06 (19.66)	29.45 (20.17)
3C	5.89×10^{-5}	1.79×10^{-6}	7.06×10^{-6}	8.69×10^{-10}	29.55 (20.80)	29.57 (20.74)	29.62 (20.44)	29.01 (20.32)
4	5792.60	175.78	694.69	0.09	0.18 (3.46)	0.20 (3.90)	0.25 (4.99)	-0.36 (5.45)
4A	1.19×10^{9}	3.61×10^{7}	1.43×10^{8}	17565.02	-0.23 (5.45)	-0.22 (4.65)	-0.16 (6.74)	-0.77 (6.24)
4B	2.37×10^{8}	7.20×10^{6}	2.85×10^{7}	3501.55	0.64 (5.56)	0.65 (5.95)	0.71 (4.49)	0.10 (5.03)
4C	1.67×10^{19}	5.07×10^{17}	2.00×10^{18}	2.46×10^{14}	0.65 (7.57)	0.67 (7.02)	0.72 (6.70)	0.11 (5.96)
5	1.00	0.03	0.12	1.48×10^{-5}	0	0.01 (2.77)	0.07 (3.60)	-0.55 (4.71)
5A	32.95	1.00	3.95	0.0005	-0.01 (2.77)	0	0.06 (4.61)	-0.56 (3.86)
5B	8.34	0.25	1.00	0.0001	-0.07 (3.60)	-0.06 (4.61)	0	-0.61 (3.06)
5C	67779.47	2056.79	8128.62	1.00	0.55 (4.71)	0.56 (3.86)	0.61 (3.06)	0
6	49758.22	1509.93	5967.38	0.73	-5.21 (6.61)	-5.20 (6.87)	-5.14 (7.56)	-5.75 (7.84)
6A	1.97×10^{5}	5967.22	23582.97	2.90	-4.02 (6.80)	-4.01 (6.23)	-3.95 (7.70)	-4.57 (7.22)
6B	64532.21	1958.25	7739.18	0.95	-3.87 (7.47)	-3.86 (7.74)	-3.80 (6.70)	-4.41 (6.92)
6C	1.16×10^{7}	3.52×10^{5}	1.39×10^{6}	171.37	-2.93 (7.85)	-2.92 (7.39)	-2.86 (6.95)	-3.48 (6.19)
7	0.002	4.81×10^{-5}	0.0002	2.34×10^{-8}	2.86 (6.39)	2.87 (6.82)	2.93 (7.27)	2.31 (7.74)
7A	0.02	0.0005	0.002	2.25×10^{-7}	3.09 (6.97)	3.10 (6.34)	3.16 (7.83)	2.54 (7.32)
7B	0.008	0.0003	0.001	1.25×10^{-7}	3.26 (7.17)	3.27 (7.63)	3.33 (6.46)	2.72 (6.98)
7C	12.61	0.38	1.51	0.0002	2.81 (7.76)	2.82 (7.21)	2.88 (7.11)	2.27 (6.33)

Table A6.

Bayes Factors and Δelpd and Δsd for All Model 6 Pairwise Comparisons

	Bayes Factors				Δelpd (2×*Δsd)*			
	6	6A	6B	6C	6	6A	6B	6C
0	0.02	0.005	0.02	8.40×10^{-5}	0.33 (1.85)	-1.01 (4.33)	-1.93 (5.01)	-1.95 (5.66)
1	0.04	0.01	0.03	0.0002	3.64 (4.38)	2.30 (6.09)	1.38 (6.17)	1.36 (6.90)
1A	13.34	3.38	10.29	0.06	2.46 (4.55)	1.12 (6.22)	0.20 (6.65)	0.18 (6.36)
1B	0.50	0.13	0.39	0.002	4.18 (5.91)	2.84 (4.72)	1.92 (4.67)	1.90 (5.74)
1C	71082.85	17986.63	54809.16	304.51	4.18 (6.44)	2.84 (5.21)	1.91 (5.71)	1.90 (5.06)
2	6.69	1.69	5.16	0.03	0.64 (2.30)	-0.69 (4.59)	-1.62 (4.73)	-1.63 (5.87)
2A	346.62	87.71	267.27	1.48	1.10 (3.42)	-0.24 (5.39)	-1.16 (5.99)	-1.18 (5.56)
2B	4.01	1.01	3.09	0.02	2.26 (5.28)	0.93 (3.86)	0	-0.01 (5.33)
2C	2.28×10^{5}	57668.62	1.76×10^{5}	976.31	2.47 (6.12)	1.14 (4.84)	0.21 (3.90)	0.20 (4.72)
3	1.41×10^{-12}	3.56×10^{-13}	1.08×10^{-12}	6.02×10^{-15}	29.42 (17.22)	28.08 (17.53)	27.15 (17.73)	27.14 (18.07)
3A	8.08×10^{-12}	2.04×10^{-12}	6.23×10^{-12}	3.46×10^{-14}	30.14 (17.73)	28.80 (18.04)	27.88 (18.38)	27.86 (18.15)
3B	7.47×10^{-14}	1.89×10^{-14}	5.76×10^{-14}	3.20×10^{-16}	35.20 (19.42)	33.86 (18.93)	32.94 (19.08)	32.92 (19.44)
3C	1.18×10^{-9}	2.99×10^{-10}	9.12×10^{-10}	5.07×10^{-12}	34.76 (20.11)	33.42 (19.61)	32.50 (19.95)	32.48 (19.66)
4	0.12	0.03	0.09	0.0005	5.39 (7.69)	4.05 (8.45)	3.13 (8.60)	3.11 (8.40)
4A	23926.66	6054.34	18448.89	102.50	4.98 (8.45)	3.64 (9.31)	2.71 (9.67)	2.70 (8.97)
4B	4769.72	1206.92	3677.74	20.43	5.85 (8.89)	4.51 (8.24)	3.59 (8.44)	3.57 (8.19)
4C	3.36×10^{14}	8.50×10^{13}	2.59×10^{14}	1.44×10^{12}	5.86 (10.04)	4.52 (9.38)	3.60 (9.93)	3.58 (8.72)
5	2.01×10^{-5}	5.09×10^{-6}	1.55×10^{-5}	8.61×10^{-8}	5.21 (6.61)	3.87 (7.47)	2.94 (7.76)	2.93 (7.85)
5A	0.0007	0.0002	0.0005	2.84×10^{-6}	5.20 (6.87)	3.86 (7.74)	2.93 (8.27)	2.92 (7.39)
5B	0.0002	4.24×10^{-5}	0.0001	7.18×10^{-7}	5.14 (7.56)	3.80 (6.70)	2.87 (7.06)	2.86 (6.95)
5C	1.36	0.34	1.05	0.006	5.75 (7.84)	4.41 (6.92)	3.49 (7.65)	3.48 (6.19)
6	1.00	0.25	0.77	0.004	0	-1.34 (3.96)	-2.26 (5.28)	-2.28 (5.53)
6A	3.95	1.00	3.05	0.02	1.18 (3.26)	-0.15 (5.07)	-1.08 (6.11)	-1.09 (4.43)
6B	1.30	0.33	1.00	0.006	1.34 (3.96)	0	-0.93 (3.86)	-0.94 (3.85)
6C	233.43	59.07	179.99	1.00	2.28 (5.53)	0.94 (3.85)	0.01 (5.33)	0
7	3.19×10^{-8}	8.07×10^{-9}	2.46×10^{-8}	1.37×10^{-10}	8.07 (7.36)	6.73 (8.17)	5.80 (8.39)	5.79 (8.31)
7A	3.07×10^{-7}	7.76×10^{-8}	2.37×10^{-7}	1.31×10^{-9}	8.30 (7.60)	6.96 (8.43)	6.03 (8.88)	6.02 (7.92)
7B	1.70×10^{-7}	4.30×10^{-8}	1.31×10^{-7}	7.28×10^{-10}	8.47 (8.16)	7.13 (7.27)	6.21 (7.56)	6.19 (7.41)
7C	0.0003	6.41×10^{-5}	0.0002	1.09×10^{-6}	8.02 (8.41)	6.68 (7.47)	5.76 (8.15)	5.74 (6.76)

Table A7 shows the regression results for all models that included self-report anger, self-report emotion, expressed emotion, expressed anger, and mean pitch. Increased self-reported anger predicted a decrease in influence, increased expressed anger and expressed fear predicted an increase in influence and no other emotions meaningfully predicted influence. However, these models did not perform better than the Intercept-Only model, which suggests that these variables are poor predictors of influence.

Table A7.
Regression Results for All Models

Parameter	*b*	β	*SE*	95% *CI*s
	Self-Report Anger			
Self-Report Anger	**-0.08**	**-0.09**	**0.03**	**[-0.14, -0.03]**
Gender	0.10	-0.10	0.06	[-0.20, 0.02]
Race	**0.14**	**-0.18**	**0.06**	**[-0.30, -0.07]**
Self-Report Anger by Gender	-0.02	0.05	0.06	[-0.06, 0.17]
Self-Report Anger by Race	-0.02	0.05	0.07	[-0.08, 0.17]
Gender by Race	-0.02	-0.13	0.12	[-0.38, 0.11]
Self-Report Anger by Gender by Race	0.03	0.12	0.13	[-0.15, 0.38]
	Self-Report Emotion			
Self-Report Emotion	-0.03	-0.04	0.03	[-0.09, 0.02]
Gender	**0.14**	**-0.12**	**0.06**	**[-0.22, -0.01]**
Race	**0.21**	**-0.18**	**0.06**	**[-0.30, -0.06]**
Self-Report Emotion by Gender	-0.04	0.08	0.06	[-0.03, 0.19]
Self-Report Emotion by Race	**-0.05**	**0.12**	**0.06**	**[0.01, 0.24]**
Gender by Race	-0.06	-0.15	0.12	[-0.39, 0.09]
Self-Report Emotion by Gender by Race	0.00	0.02	0.12	[-0.21, 0.26]
	Expressed Emotion			
Calm v. Angry	**-.27**		**.10**	**[-.46, -.08]**
Disgust v. Angry	**-.17**		**.04**	**[-.24, -.10]**
Fear v. Angry	.17		.15	[-.13, .47]
Happiness v. Angry	-.10		.05	[-.20, .0004]
Sadness v. Angry	-.06		.06	[-.18, .06]
Surprise v. Angry	**-.17**		**.06**	**[-.28, -.06]**

Parameter	*b*	β	*SE*	95% *CI*s
	Expressed Emotion			
Gender	-.09		.05	[-.24, .06]
Race	-.14		.06	[-.30, .01]
Calm v. Angry x Gender	-.02		.02	[-.21, .16]
Disgust v. Angry x Gender	-.04		.04	[-.12, .04]
Fear v. Angry x Gender	.09		.60	[-1.08, 1.26]
Happiness v. Angry x Gender	.01		.05	[-.09, .11]
Sadness v. Angry x Gender	-.10		.06	[.22, .02]
Surprise v. Angry x Gender	.11		.06	[-.003, .22]
Calm v. Angry x Race	-.09		.02	[-.27, .09]
Disgust v. Angry x Race	.06		.04	[-.02, .13]
Fear v. Angry x Race	**.37**		**.19**	**[.0008, .74]**
Happiness v. Angry x Race	-.04		.06	[-.15, .07]
Sadness v. Angry x Race	.09		.06	[-.03, .21]
Surprise v. Angry x Race	.003		.06	[-.12, .13]
Calm v. Angry x Gender x Race	-.14		.09	[-.33, .04]
Disgust v. Angry x Gender x Race	-.02		.04	[-.11, .06]
Fear v. Angry x Gender x Race	.20		.54	[-.86, 1.24]
Happiness v. Angry x Gender x Race	.03		.06	[-.09, .15]
Sadness v. Angry x Gender x Race	-.07		.06	[-.19, .06]
Surprise v. Angry x Gender x Race	.12		.06	[-.01, .24]

Parameter	*b*	β	*SE*	95% *CI*s
	Expressed Anger			
Expressed Anger	**0.28**		**0. 05**	**[0.17, 0.39]**
Gender	0.07		0.04	[-0.02, 0.15]
Race	-0.12		0.04	[-0.26, 0.02]
Expressed Anger by Gender	-0.15		0.06	[-0.30, 0.01]
Expressed Anger by Race	-0.08		0.06	[-0.19, 0.04]
Gender by Race	-0.07		0.06	[-0.19, 0.05]
Expressed Anger by Gender by Race	0.01		0.06	[-0.11, 0.13]
	Mean Pitch			
Mean Pitch	-0.24	-0.04	0.03	[-0.10, 0.02]
Gender	0.20	-0.09	0.06	[-0.22, 0.03]
Race	**0.05**	**-0.18**	**0.06**	**[-0.30, -0.06]**
Mean Pitch by Gender	-0.35	0.12	0.06	[-0.01, 0.24]
Mean Pitch by Race	0.09	-0.03	0.06	[-0.15, 0.09]
Gender by Race	0.03	-0.08	0.14	[-0.36, 0.19]
Mean Pitch by Gender by Race	0.28	0.20	0.14	[-0.07, 0.47]

APPENDIX J

RESULTS FROM THE EXPLORATORY MODELS

Exploratory Moderators

Because Model 3B (Perceived Emotion by Race) was the best model, I used that model as the basis for the exploratory moderators. First, I examined whether the three exploratory moderators (photograph condition, instruction condition, and deliberation sample) predicted perceived emotion. Mock jurors were not perceived as more emotional when they saw color photographs, β = .02, *SE* = .04, 95% *CI* [-.06, .10], or black-and-white photographs, β = .01, *SE* = .05, 95% *CI* [–.09, .11], compared to no photographs. Similarly, instruction condition did not predict perceived emotion, β = .03, *SE* = .04, 95% *CI* [–.04, .10]. However, deliberation sample did predict perceived emotion, such that online participants were perceived as less emotional than in-person participants, β = –.09, *SE* = .04, 95% *CI* [–.16, –.02]. However, deliberation sample did not predict maximum intensity. Because the Model 7B was also well supported in the data, I used Model 7B as the basis for the third exploratory model.

Therefore, I ran 3 additional models to examine 1) the interaction between photograph condition, juror race, and perceived emotion; 2) the interaction between instruction condition, juror race, and perceived emotion; and 3) the interaction between survey, juror race, and maximum intensity. The following are the model syntax for the three exploratory models:

- Model 3D:

$$Influence \sim Perceived\ Emotion * Race * Photograph + (1 \mid Jury) + (1 \mid Jury{:}\,RatedJuror)$$

- Model 3E:

$$Influence \sim Perceived\ Emotion * Race * Instruction + (1 \mid Jury) + (1 \mid Jury{:}\,RatedJuror)$$

- Model 7F:

$$Influence \sim Perceived\ Emotion * Race * Survey + (1 \mid Jury) + (1 \mid Jury{:}RatedJuror)$$

Model 3B was still the most likely, given the data (Table A8), and had the best out-of-sample prediction (Table A9). While there was extreme support for Model 3B over all of the exploratory models, Models 3B, 3D, and 3E were all relatively similar in out-of-sample prediction (all elpd differences were within ~ 2 standard error of the difference, indicating no large differences). However, as shown in Table A10, none of the three-way interactions were supported by the data.

Table A8.

Bayes Factors for All Hypothesis 1 Pairwise Comparisons

Denominator	Numerator							
	Model 0	Model 3	Model 3B	Model 3D	Model 3E	Model 7	Model 7B	Model 7F
Model 0	1.00	5.61×10^{9}	6.12×10^{10}	0.91	71372.90	1.35×10^{5}	56266.75	8.70
Model 3	1.78×10^{-10}	1.00	10.90	1.62×10^{-10}	1.00×10^{-5}	2.00×10^{-5}	1.00×10^{-5}	1.55×10^{-9}
Model 3B	1.63×10^{-11}	0.09	1.00	1.49×10^{-11}	1.17×10^{-6}	2.21×10^{-6}	9.20×10^{-7}	1.63×10^{-11}
Model 3D	1.10	6.17×10^{9}	6.72×10^{10}	1.00	78413.18	1.48×10^{5}	61816.95	1.10
Model 3E	1.00×10^{-5}	78629.81	8.57×10^{5}	1.00×10^{-5}	1.00	1.89	0.79	1.00×10^{-5}
Model 7	7.40×10^{-6}	41547.80	4.53×10^{5}	6.74×10^{-6}	0.53	1.00	0.42	7.40×10^{-6}
Model 7B	2.00×10^{-5}	99739.86	$1{,}09\times10^{6}$	2.00×10^{-5}	1.27	2.40	1.00	2.00×10^{-5}
Model 7F	0.11	6.45×10^{9}	7.03×10^{9}	0.10	8205.61	15529.24	6468.89	0.11

Note. Values that are greater than one indicate that there is more support for the numerator model than the denominator model. Values less than one indicate that there is more support for the denominator model than the numerator model. Shading was done to improve the readability of the table.

Table A9.

Δelpd and Δsd for All Hypothesis 1 Pairwise Comparisons

Comparison	Reference						
	Model 0	Model 3	Model 3B	Model 3D	Model 3E	Model 7	Model 7B
Model 0	0						
Model 3	29.09 (17.10)	0					
Model 3B	34.87 (19.29)	5.78 (8.47)	0				
Model 3D	31.81 (19.93)	2.72 (9.73)	-3.06 (4.49)	0			
Model 3E	32.62 (19.49)	3.53 (8.93)	-2.25 (2.76)	0.81 (5.07)	0		
Model 7	7.74 (7.91)	-21.35 (18.51)	-27.13 (20.47)	-24.07 (18.51)	-24.88 (20.66)	0	
Model 7B	8.14 (7.83)	-20.95 (18.71)	-26.73 (19.99)	-23.67 (18.71)	-24.48 (20.16)	0.40 (3.47)	0
Model 7F	7.21 (7.93)	-21.88 (18.81)	-27.66 (20.11)	-24.60 (18.81)	-25.41 (20.29)	-0.53 (3.84)	-0.93 (1.68)

Note. The Δelpd is reported with 2×*Δsd* in paratheses. Positive Δelpd values indicate that the comparison group has better predictive accuracy than the reference group. Negative Δelpd values indicate that the reference group has better predictive accuracy than the comparison group. If 2×*Δsd* is greater than Δelpd, the models are considered similar in predictive accuracy.

Table A10.

Regression Results for Exploratory Three-Way Interactions

Parameter	*b*	β	*SE*	95% *CI*s
B&W Photograph v. No Photographs	-0.03	.06	.04	[-.03, .15]
Color Photograph v. No Photographs	-0.07	.01	.04	[-.07, .08]
Perceived Emotion by Race by B&W Photograph v. No Photographs	-0.02	-.12	.06	[-.07, .15]
Perceived Emotion by Race by Color Photograph v. No Photographs	0.01	.04	.06	[-.11, .06]
Instruction Condition	-0.06	,04	.07	[-.10, .17]
Perceived Emotion by Race by Instruction Condition	-0.01	-.06	.08	[.22, .07]
Survey	0.15	.04	.08	[-.12, .19]
Maximum Intensity by Race by Survey	-0.06	.02	.14	[-.29, .24]