THE FACE OF JUDGMENT:
MORAL EVALUATIONS IN THE CONTEXT OF UNCERTAINTY

By

Andrew Marcel Defever

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Psychology – Doctor of Philosophy

2018

ABSTRACT

THE FACE OF JUDGMENT:
MORAL JUDGMENT IN THE CONTEXT OF UNCERTAINTY

By

Andrew Marcel Defever

Moral judgment is an integral part of social living; a ubiquitous phenomenon across time and human cultures. Inherent to this is an uncertainty factor: inferential processes necessitate a component of situational ambiguity, requiring the perceiver to make a best-guess considering the available evidence. While research has demonstrated variability exists in judgments both between situations and within, less is known about how judgments occur under explicit uncertainty. Across a series of 4 studies, I examined whether third party emotional expressivity influences severity judgments of moral transgressions, targeting situations that manifest uncertainty. A pilot study was first implemented to identify stimuli appropriate for moral judgment research. Study 2 and Study 3 examined the influence of facial expressions (angry, disgust, neutral, happy) on moral severity judgments in both clear and uncertain contexts but did not find any effect. Study 4 implemented auditory expressions (angry/neutral crowd sounds) instead of facial expression and found that auditory expressions did influence severity judgments in two instances. Implications and future directions on the function of communicative expressions under situational uncertainty are discussed.

To my wife Rachel,
my parents Lynda and Guido,
and my sisters Erin and Mahrie.
Your lessons of love and kindness are my greatest gifts in life.
Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

Morality is a ubiquitous social phenomenon that encompasses 'oughts' within the domain of human behavior. That is, morality dictates how people should and should not behave toward others or toward sacred items in their lives (e.g., shrines, books, people, deities, etc.). Most, if not all, human cultures have a moral system in place, and even some non-human ape societies may have moral dictates as well (e.g., fairness; Brosnan & de Waal, 2014). However, precisely what morality is, its conceptual boundaries, and subsequent evaluative components, have not been thoroughly investigated and clarified. Indeed, the prolific nature of morality in our everyday life underscores the need for a better understanding of its constituent properties.

Here I outline a series of studies designed to examine one component related to morality: judgment and decision-making. Inherent in moral systems is an evaluative component; a way to assess whether an action is right or wrong, or whether it does or does not violate a moral dictate. Evaluative judgments about morality are multifaceted, including both qualitative evaluations of whether an act is or is not moral, and quantitative evaluations of how commendable or severe the act is (Bartels, Bauman, Cushman, Pizarro, & McGraw, 2014). When we view behavior that is potentially moral in nature, we may progress through a number of steps: gathering of evidence, emotional reactions, interpersonal judgment, and sometimes counter-action. Making morally relevant judgments, then, is an integral part of the operation of moral pre- and proscriptions.

I focus here on judgment of moral violations of others, such as the assessment of another transgressing a proscriptive moral norm. That is, judging the severity of an act when a person is seen as doing something that is morally prohibited. While there is evidence that internal characteristics (e.g., motivation, intent; Young & Saxe, 2011; Cushman, 2008) and external

characteristics inherent to a situation (context, means; Mikhail, 2007; McDonald, Defever, & Navarrete, 2017) can modify the perceived severity of an action, these typically involve judgments of clear and unambiguous situations. Much less is known about how moral judgments occur with imperfect knowledge or ambiguity, a situational characteristic necessarily present in judgments of others.

Across a series of 4 studies, I examined how judgments may be influenced by social cues in the assessment of ambiguous situations. Specifically, I investigated whether judgments are influenced by the non-verbal behavior of others in the proximal situation. As emotions are reliably linked to moral reactivity (Prinz, 2006; Avramova & Inbar, 2013), and produce distinct facial and somatic expressions theorized to have evolved as communicative devices (Shariff & Tracy, 2011), I examined whether the external signaling components of emotional expression convey meaningful information during the moral judgment process.

I hypothesized that (1) emotional expressions conveyed by others influenced moral severity judgments by informing the target of relevant contextual evidence. Because judgments of others are inferential, uncertain, and involve an appraisal of available evidence, I further hypothesized that (2) this relationship would be strongest when the situation is ambiguous, as perceivers with imperfect knowledge should glean and subsequently utilize the information contained in the emotion to inform their judgment. Finally, I explored how a number of individual difference characteristics might affect the degree to which people use emotional expressions to inform moral judgments under situational uncertainty.

Moral judgment and decision-making

Morality as a construct encapsulates a wide variety of behaviors and responses. Here I focus specifically on transgressions: acts assessed as immoral or 'wrong' that typically evoke a

negative affective response. In this case, moral judgment of a person's behavior is an assessment of wrong-ness; of the severity of the transgression. Further, moral judgment is a relative appraisal process on a continuous spectrum, such that some acts are more wrong than others. For example, theft and murder are both morally wrong, but an individual would not judge theft and murder as *equally* severe moral violations. From a sociological perspective, this continuous spectrum is also evident in the institutionalization of moral proscriptions. Murder is consistently punished more severely than theft across all modern punitive institutions, with the level of punishment scaling with the severity of the transgression.

However, it's not simply the action or the outcome that influence the assessment of a moral transgression. Extant literature suggests a number of situational and individual factors that influence the judged severity of an action. For example, a classic finding from trolley dilemma research (i.e., Thomson, 1985) shows that when deciding whether to kill one person to save five others, the action is judged as more wrong when it involves interpersonal contact (e.g., a push) versus when it does not (e.g., flipping a switch; Mikhail, 2007; Cushman, Young, & Hauser, 2006). Intention is another salient factor: if harm comes to another through an intended, volitional act (e.g., deciding to poison another person), it is judged as more wrong than if it occurred unintentionally or indirectly as a side effect (e.g., accidentally poisoning them; Cushman, 2008). While in both of these examples the outcome is the same across both courses of action (i.e., a person is killed), additional facets of the situation and/or the actors involved alters the judgment as more or less severe.

Why would the same outcomes be judged as more or less severe depending on conditional factors of the situation? Logically, judgment of another person's action is an inferential process (DeScioli & Kurzban, 2009). It involves a gathering of evidence to inform the

judgment and appraise the action appropriately, including information about both the actor's internal state (e.g., motivation, intention, and psychological state) and external information (e.g., the means implemented in the act and the socioecological context of the act).

When making an inferential judgment there is an inherent and unavoidable uncertainty to the judgment, derived from the subjective nature of the human mind: we are unable to peer into the 'contents' of another person's brain. When judging another person's actions, we must infer (through abductive reasoning) the most likely conclusion on what evidence is available. This applies *necessarily* to assessing the internal state of the violator: even if they verbalize their intention and motivation for their act, one cannot have epistemologically perfect certainty of their mental state.

Concerning external and contextual information, there is more variation in the level of knowledge a person might have in regards to judgment-relevant evidence. For example, if someone witnesses a person strike another person in the face at a boxing match, the contextual information is salient, readily available, and informs the judgment of the act. Here the act is likely judged as a minimally severe violation, since the context of a boxing match calls for interpersonal harm as normative and expected. However, if someone witnesses the same act on a busy public street, between two people in business attire on their way to work, much less contextual information is readily available. Likely this same act would be judged as more severe than in the boxing match, since the context of a morning commute to work does not call for interpersonal harm as normative or expected.

Uncertainty

Another way to construe the nature of inferential judgments is that they're fraught with uncertainty. That is, there are facets of the situation that are not reliably known, or potentially

unknowable. While uncertainty can be broken down into different epistemological types, I focus here on uncertainty as a lack of clarity in information visibly or readily available in a situation (Smithson, 2008).

Uncertainty of this type is a ubiquitous, powerful, and psychologically salient feature of the environment. From an evolutionary perspective, uncertainty has been so prolific in the environment as to shape cognitive mechanisms over long periods of time. For example, error management theory (EMT) describes how uncertainty in the environment has led to a biased evaluative mechanism. This biased mechanism arises from a consistent asymmetry in outcomes for false-positive and false-negative conclusions. For instance, men are thought to be biased in predicting women's sexual interest in them because it was costlier for men to miss a potentially interested mate than to perceive them as uninterested. That is, when the perceptions of a woman's intent are uncertain, men are thought to be biased toward perceiving women as interested (false alarm) as opposed to not interested (missed detection; Haselton & Buss, 2000).

Proximally, uncertainty makes people uncomfortable and anxious, and motivates individuals to alleviate their discomfort (Reiman, Fusselman, Fox, & Raichle, 1989). In uncertain situations individuals are likely to utilize cues from the environment in order to disambiguate the situation. For example, imagine you come across a crowd of people while walking down the street. The crowd is surrounding what appears to be two individuals, and you hear noises that sound like a verbal conflict. Are the two fighting? Are they in danger? Are *you* in danger? As you walk closer, you notice the crowd; their faces are not angry and their posture does not indicate they are threatened. As you merge into the crowd to get a better look, you notice the individuals are dressed in costumes, and there's a small bucket on the sidewalk full of change. Indeed, based on the available evidence you infer they are street performers and it is not

a threatening situation. In this way, you have disambiguated the uncertain situation by accumulating evidence about other people's emotional states, the environment, and assess the likelihood of given scenarios (a street performer versus an altercation).

Uncertainty in moral judgment. EMT highlights the power of uncertainty in shaping our perceptions of the world over evolutionary time. Uncertainty has been a powerful and salient feature of the environment that led to costly outcomes and shaped cognitive mechanisms accordingly. Proximally, there are challenges unique to observers of moral acts (as opposed to perpetrators or victims, for instance), with one primary challenge being the acquisition of evidence without certain knowledge (DeScioli & Kurzban, 2009). Additionally, concerning moral judgments it can also be costly to fail to detect a moral violation. The capacity for cooperation and peaceful exchange is a foundational characteristic of extended interactions among individuals, as well as a prominent feature of person perception (Kurzban & Leary, 2001). A failure to detect social free riders, or to detect that a person may be deceptive and likely to take advantage of someone (e.g., in the instance of a moral transgression), can lead to costly outcomes for the person in the relationship as well as their larger social group as a whole. Thus, inferring guilt or innocence in the event of a transgression, and subsequently garnering appropriate evidence, is a critical part of the process.

Emotions

Under uncertainty, we use cues from the environment to help inform our judgments. In the realm of moral judgment, factors of the situation and the motives of those involved impact our judgments. However, what if no such evidence is available? What if, for instance, you came upon the same crowd of people surrounding two individuals and you see one of them strike the other in the face. How would you judge the actions of those involved? One salient cue that

people attend to is the reactions of others. Specifically, people attend to their emotional reactions.

Morality and emotions. Emotions are reliably linked to moral reactivity (Avramova & Inbar, 2013). The relationship is so consistent that some have posited an emotional reaction as necessary *by definition* for a situation to be subjectively experienced as 'moral' (e.g., Prinz, 2006). While this is a strong position, and primarily a philosophical one, it underscores the strength of the association between morality and emotions. Indeed, witnessing moral transgressions evokes feelings of disgust, anger, and contempt (i.e., "other-condemning emotions"; Haidt, 2003), and these emotions can serve in both the judgment of the transgression (Haidt, 2001; 2012) and motivation of behavior on the part of the perceiver (Tangney, Stuewig, & Mashek, 2007; but see Blasi, 1999).

Emotions are a complex, multi-faceted phenomenon involving the coordination of a suite of physiological and psychological phenomenon. According to Schwarz & Clore (2007), emotion involves three primary components: a cognitive/experiential component (e.g., the thoughts and conscious experiences that accompany an emotional reaction) a bodily component (e.g., the somatic reactions accompanying an emotion; postural changes, behavior), and an affective component (e.g., autonomic reactivity, physiological arousal; the 'feeling' of an emotion with a broad positive or negative valence). These three systems coordinate to produce what is colloquially lumped into an 'emotional reaction'. However, these three components, while associated in occurrence, may serve different adaptive functions.

Of specific interest to the present investigation are the somatic components of emotional reactions. Emotions are strongly and reliably linked to distinct non-verbal behaviors, such as facial expressions, posture, and body language. Here I define "non-verbal" as any behavior that

does not utilize language, but may include non-verbal auditory behaviors (e.g., a scream, whimper, cheer, gasp, sob, etc.) As a somatic phenomenon, they are outwardly expressive displays. They can be seen by others and interpreted as reliable indicators of what a person is experiencing emotionally.

Facial expressions, specifically, are theorized to have evolved to serve a communicative purpose: to convey meaningful information to nearby conspecifics about stimuli in the environment (Shariff & Tracy, 2011). Facial expressions are outward-facing, exaggerated, and are accurately displayed/interpreted across most, if not all, human cultures (Ekman et al., 1987). These are design features indicative of an evolved communicative signal (Eibl-Eibesfeldt, 1989). Facial expressions and some non-verbal behaviors are also present in our close evolutionary ape ancestors (Parr & Waller, 2006), indicating that they may have evolved early in our ancestral timeline. Thus, considering the association between moral judgment and emotional reactivity, and the link between emotional reactivity and facial expressions, I expect the facial expressions of others may serve as evidence to inform moral judgments.

The Present Studies

Moral judgment, here defined as an evaluation of the severity of a moral transgression, is an inferential process requiring the acquisition of evidence to make an appropriate assessment. However, little is known about how judgments occur under uncertainty, and how people use available evidence during the judgment process. Uncertainty, necessarily present in the judgment of others' internal states (e.g., motives, intent) and typically present to some degree in the environment, necessitates the use of relevant information during the judgment process. Thus, I seek to examine how moral judgments are informed by available evidence, particularly when the situation is uncertain. This examination also assesses whether facial expressions and auditory

cues serve to communicate information during evaluative judgments, and explores the link between morality and emotion.

I designed a series of 4 studies (denoted henceforth as Study 1 - Study 4) as a comprehensive test of whether people utilize the emotional content of others' reactions as evidential input during their moral judgments in uncertain situations. As emotion is reliably linked to moral transgressions (Avramova & Inbar, 2013) and produces distinct somatic reactivity thought to have evolved as a communicative device (Shariff & Tracy, 2011), here I utilize facial expressions and vocalizations in an experimental context to assess whether such signals serve as inputs to the judgment process. Study 1 was a pilot study to identify appropriate moral stimuli to be used in the package of studies that comprise this dissertation. Studies 2 and 3 examine the role that facial expressions play in forming moral judgments under uncertainty. Study 4 will examine whether auditory cues also inform moral judgment under uncertainty in a similar way to communicative facial expressions.

Specifically, I hypothesize that (1) emotional signaling (opeationalized via facial expressions and auditory vocalizations) will influence moral severity judgments. When viewing images of moral transgressions, emotional signaling will serve as evidence that inform the target of relevant contextual evidence. Subsequently, differences in severity judgments of the same image will vary systematically with different expressions. I also hypothesize that (2) this relationship will be stronger when the image being viewed is more uncertain. Perceivers in uncertain judgment contexts are motivated to glean evidence from the environment (e.g., Reimanet al., 1989). Communicative expressions, being reliably linked to emotional reactivity (Ekman et al., 1987; Anderson, Monroy, & Keltner, 2017; Oveis, Spectre, Smith, Liu, & Keltner, 2016) and by extension moral reactivity (Avramova & Inbar, 2013), will serve as a salient source

of evidence to inform their judgment. Finally, I will explore how a number of individual difference characteristics might moderate the degree to which people assess the uncertainty of moral situations, how uncertainty feels to perceivers, and how this may inform moral judgments under situational uncertainty.

Overview of Dissertation

A pilot study (Study 1; Chapter 2) was first implemented to identify stimuli appropriate for use in moral judgment research. Study 2 (Chapter 3) and Study 3 (Chapter 4) examined the influence of facial expressions on moral severity judgments in both clear and uncertain contexts. Study 2 utilized a within-subjects design where participants viewed morally salient images and rated the severity of each violation. Uncertainty was assessed via a self-report scale item for each image. To overcome some of the limitations identified in Study 2, Study 3 utilized a between-subjects design and experimentally manipulated uncertainty. Study 4 (Chapter 5) builds off of Study 2's design, but utilized auditory emotion cues instead of facial expressions. Should emotional expression serve as evidence in the moral judgment process, I predicted a similar pattern of results as in Studies 2 and 3.

Finally, I integrate the findings across the four studies into a discussion (Chapter 6) about the extent to which people use communicative information to inform moral judgment under uncertainty, individual differences in these processes, and directions for future research.

CHAPTER 2

Study 1: Pilot Study

Study 1 was designed to (1) identify appropriate, morally relevant stimuli, and (2) to assess whether the stimuli would allow for sufficient variance in both severity and uncertainty for use in subsequent studies on moral judgment.

In the broader literature, moral stimuli range from descriptive paragraphs of interpersonal situations to static images, videos, and even virtual reality situations in which participants act in real-time to moral dilemmas. However, there are several methodological problems with moral stimuli in particular that reduce variance in responding and may inhibit experimental manipulations from influencing judgment.

First, judgments about situations that have been experienced previously (e.g., a trolley dilemma scenario) may elicit 'canned' responses that are invariant to experimental assessment or manipulation. That is, participants have previously reasoned about the situation (or similar iterations), have prior knowledge about the types of responses being elicited (or the types of responses frequently given by others), and produce a previously formed decision instead of one that encapsulates an active in-the-moment decision (e.g., Chandler, Mueller, & Paolacci, 2014). This could be due, in part, to the popularity of trolley and other similar dilemmas and stimuli in morality research over the past decade, exposure to moral stimuli in broad societal contexts (e.g., Brain Games season 5 episode 3, "Morality", Kolber, Margol, & Nigro, 2015; The Good Place season 2 episode 5, "The Trolley Problem", Schur, M. & D. Holland, 2017) and/or social demand characteristics dictating socially acceptable responses regardless of the context of the moral act being assessed. Regardless of where people's problematic responses come from, these

are methodological hurdles to morality research that should be addressed prior to the implementation of stimuli in morality research.

Second, stimuli showing extreme moral transgressions are prone to ceiling (or floor) effects. For example, participants will always judge an act like murder or egregious interpersonal harm as severe, regardless of the situation or context. This creates a skewed response distribution that is difficult to work with statistically, and often violates the normality assumptions inherent in many standard statistical tests.

Finally, moral judgments of descriptive paragraphs or vignettes may not activate the psychological constructs of interest, as they do not contain the visual stimuli upon which these mechanisms operate. For example, describing a 'disgusting' act in words may not elicit the same feelings as sensing or witnessing the said act firsthand, as disgust is primarily triggered by sensory inputs such as vision, taste, or smell. This is particularly notable in the context of moral reactivity, as emotions are posited to serve as input into the judgment process. Should a vignette or written stimulus not properly engage related affective systems, the ecological validity of the findings may be undermined.

To overcome these potential issues with experimental moral stimuli, I examined judgment ratings of a number of interpersonal situations depicted in the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008). The IAPS has a number of key features that make it attractive for morality research. First, the images are well-guarded in the research community, such that the stewards protect and isolate the images from the general public domain[1]. Thus, it is unlikely that participants will have seen the stimuli previously. Second, they are detailed visual images that capture both interpersonal and impersonal moral violations in salient and ecologically valid contexts. As such, the images encapsulate sensory

inputs that are relevant for activating morally relevant psychological constructs of interest. Third, the images span a broad range of situations, such that judgments can be elicited from more or less severe moral transgressions to examine response tendencies across a variety of stimuli.

Participants

Five hundred and three undergraduate students volunteered to participate in a study titled "Interpersonal Perceptions" through the online subject pool recruitment website at Michigan State University. After excluding those who failed the self-reported attention check (n=7), declined to participate after reading the consent (n=22), or failed to complete a sufficient number of the image ratings (n=91)[2], the final sample size was N=383. Students received course credit for their participation. Mean age of participants was M = 19.69 (SD = 2.33), with approximately 71% identifying as female. The racial composition of the sample included 75% White, 13% Asian/Pacific Islander, 7% Black, 4% Hispanic/Latino, and 2% other or not specified.

Materials & Procedure

In consultation with the Center for the Study of Emotion and Attention, I was permitted access to the IAPS image set for use in these studies. After a visual assessment of the database, a subset of 42 images was selected for pre-rating. These images depicted a wide range of acts, from neutral or non-moral acts (e.g., a man riding a bike in the countryside) to theft (e.g., a woman shoplifting), drinking and driving (e.g., a person behind the wheel of a car drinking a beer), threat of interpersonal harm (e.g., a man pointing a gun at another man), and bodily mutilation or torture (e.g., a bloodied man being hanged; a deceased, bloodied body of a woman who appeared to have been tortured). A full list of IAPS image numbers that were rated in the study, as well as a brief description, can be found in the appendix. Of note, because these images

were to be used in conjunction with a facial expression manipulation, the faces of the actors in the images were blurred so as not to influence the manipulation in subsequent studies.

Participants arrived at the laboratory for their scheduled session, where a research assistant (RA) greeted them. Upon entering the lab, they were escorted to a computer where they completed all study procedures. Informed consent was completed by all participants prior to participation, followed by a brief instruction screen discussing the process of rating the images.

First, participants viewed the entire set of images in random order without providing any ratings. This was intended to prevent anchoring effects for participants, should they view a particularly gruesome or startling image first during the rating sequence. Participants were not timed and were instructed to simply look at the images and move to the next at their own pace. Then they viewed each image a second time, one at a time and in random order. Using a 7-point scale (1=*not at all*; 7=*extremely*), they provided ratings of the moral severity of the actions depicted ("How immoral do you think the picture is?"), how uncertain they found the images ("How morally ambiguous do you think the picture is?"), and how morally relevant they thought the images were ("How morally relevant do you think the image is?"). Definitions of each criterion (e.g., what 'morally ambiguous' means) were provided to the participants during the instructions to ensure participants understood specifically what types of judgments were being solicited.

They also provided ratings of their emotions toward each image ("How does the image make you feel?") using a 7-point scale (1=*not at all*; 7=*extremely*), including happy, sad, angry, fearful, physical disgust, and moral disgust. Finally, participants provided basic demographic information about their age, race/ethnicity, gender, and education.

Results

Table 1 shows the mean ratings for severity, relevance, uncertainty, and each emotion rating (happy, sad, angry, fearful, physical disgust, moral disgust). Across all images, mean severity ratings ranged from 1.33 to 6.47, showing robust variation in severity across the images. Mean relevance ratings ranged from 1.96 to 5.15, which also implied much variation, albeit tighter around the center of the scale. Mean ambiguity ratings varied the least, ranging from 2.45 to 4.11.

Table 2 shows the inter-correlations between the rating variables. Due to the large sample size of ratings, nearly every bivariate relationship is significant at the $p<.01$ level, so direction and relative size were used to guide interpretation. Severity and relevance were strongly associated at $r=.54$, indicating that as moral relevance increased, severity also increased. Severity was also positively associated with ambiguity ($r=.08$). However, this relationship was weaker.

Severity was positively associated with negative emotional reactivity (sad, angry, fearful, physical disgust and moral disgust; $rs > .64$), and negatively associated with happiness ($r = -.33$). Severity correlated most strongly with moral disgust ($r = .84$). Moral relevance displayed the same pattern of results (negative emotions: $rs > .34$, happiness: $r = -.09$), and also correlated most strongly with moral disgust ($r = .48$). The negative emotions showed strong inter-correlations ($rs > .74$), and a negative correlation with happiness ($rs > -.23$).

Distributions. As anticipated, a number of images displayed highly skewed distributions for severity judgments. Images of severe moral transgressions displayed strong negatively skewed distributions (i.e., ceiling effects), while neutral images displayed patterns of strong positive skew (i.e., floor effects). However, a number of the intermediately severe images

showed a relatively normal and varied distribution, ideal for use in statistical models. An example of each of the image rating distributions can be found in Figure 1.

Discussion

Having assessed the variance in severity and ambiguity of selected IAPS images, a subset of 16 images was identified as appropriate for use in Studies 2-4 based on (1) a visual inspection of the response distributions and q-q plots, such that images with the highest level of normality and least amount of skew were selected, (2) inspection of ambiguity ratings, such that a range of ambiguity levels were selected, and (3) the actions depicted in the images, such that the selection included a range of interpersonal and impersonal actions.

CHAPTER 3

Study 2

Study 2 was designed to directly test the hypotheses of interest: First, that emotional expressions conveyed by others will influence moral severity judgments by informing the target of relevant contextual evidence; Second, that this relationship will be strongest when the situation is uncertain, as perceivers with imperfect knowledge should glean and subsequently utilize the information contained in the emotion to inform their judgment.

Emotional evidence was operationalized using images of facial expressions; communicative signals (e.g., Shariff & Tracy, 2011) are theorized to have evolved to communicate information about the environment to others. Moral judgments were made via a subset of IAPS images selected based on the results of Study 1. Sixteen images were chosen based on (1) the relative normality of the response distributions, so as to avoid ceiling and floor effects and normality violations; (2) their ambiguity ratings, such that variance in uncertainty was present across the images; and (3) the explicit content of the images, to ensure that a range of moral transgressions was present across the images. Further, one neutral image (2026) and one extreme image (3530) were selected in order to examine responding in the most extreme and least extreme cases and serve as a valid responding check. Specifically, we applied the following criteria as indicators of invalid responding: anyone who rated the neutral image as extremely morally severe (i.e., 90-100 on a 0-100 scale) or indicated the extreme image as not at all morally severe (i.e., 0-10 on a 0-100 scale) was excluded from the study.

Participants

Two-hundred sixty-four undergraduate students volunteered to participate in a study titled “Interpersonal Perceptions 2” through the online subject pool recruitment website at

Michigan State University. Individuals who participated in Study 1 were barred from participating in the present study, due to previous exposure to the judgment stimuli. After excluding those who failed the self-reported attention check (n=5), failed to complete a sufficient number of the image ratings (n=5), or failed the aforementioned valid responding check (n=6), the final sample size was N=248. Students received course credit for their participation.

The mean age of participants was M = 19.77 (SD = 1.77), with approximately 57% identifying as female. The racial composition of the sample included 63% White, 21% Asian, 11% Black, and 5% other or not specified.

Materials & Procedure

Facial expression stimuli. The facial expressions implemented in the study were happy, neutral, anger, and disgust. These are basic, universal facial expressions (i.e., Ekman et al., 1987) that are reliably recognized in all human societies and cultures. As such, they should be reliably recognized by all participants in the study regardless of race, ethnicity, or country of origin. Anger and disgust, in particular, are defined as 'other-condemning' emotions (Haidt, 2003). They're linked to the condemnation of moral transgressions performed by others. As opposed to subtler other-condemning expressions such as contempt or self-condemning emotions such as shame, anger and disgust should serve as strong emotional signals to perceivers of a transgression. A neutral expression was included as a control expression against which the others could be compared. Finally, a happy expression was included to explore the possible influence of a positive emotion on the judgment of the images.

Facial expression images were taken from the Extended Cohn-Kanade AU-Coded Expression Database (CK+; Kanade, Cohn, & Tian, 2000; Lucey et al., 2010). Upon agreement with the terms of use, these images are free to download and use for research purposes. The

image set contains black-and-white and color photos of men and women performing a range of externally validated emotional expressions (e.g., happy, sad, fear, surprise, disgust, neutral).

Upon visual inspection of the database, not all subjects in the database had expressions for all of the target emotions, and some of the expressions were not visually distinct and clear across emotions. For the present study, I chose a single white male target (S037) in black-and-white. S037 was chosen because he had distinct and exaggerated expressions for the target emotions (anger, disgust, neutral, happy). I chose a white male target in order to control for race and variance in the target face across judgment scenarios.

Judgment stimuli. Sixteen images from the IAPS were chosen from the 42 piloted images. As discussed in Study 1, the images were selected for their severity rating distributions, variance in ambiguity ratings, and variety of actions depicted. The list of images used in Study 2, with IAPS number and short informal descriptions from Study 1, can be found in Table 3 of the appendix.

Judgment task. Similar to Study 1, participants arrived at the laboratory for their scheduled session where an RA greeted them. They were escorted to a computer where they completed all study procedures. Informed consent was completed by all participants followed by a brief instruction screen discussing the process of rating the images.

Participants first viewed each of the 16 IAPS images in random order with no time limit. The images were not accompanied by the facial expression stimuli. As in Study 1, this first viewing was intended to show participants all of the images *before* rating them, to prevent anchoring effects should they be randomly assigned to rate the most severe image first. Then participants viewed each image a second time, where they were asked to provide their judgment and emotion ratings. In this second viewing, each image was accompanied by a facial expression

with two images positioned immediately side-by-side, facial expression on the left and the target image on the right. Participants were instructed to "…answer the questions in regards to the image on the right only". This study utilized a within-subjects design, in which images and expressions were fully crossed and randomized, such that each participant provided a total of 64 image ratings (16 target images X 4 expressions).

For each image-face pairing, participants were asked to make three judgments. First, participants made a severity judgment about the actions in the picture ("How immoral are the actions in the picture?"), which served as the primary criterion of interest. Second, they were asked to judge how they thought others would judge the image ("Regardless of your own moral assessment of the image, how immoral do you think other people would find the actions in the picture?"). This was asked as a separate corroborative assessment of their judgments, to avoid potential social desirability bias in their assessments. Finally, they were asked how ambiguous they thought the actions in the image were ("How morally ambiguous do you find the actions in the picture?"). This was designed to assess uncertainty in the situation. Each of the three judgments were rated on a horizontal sliding scale ranging from (0=*Not at all*) to (100=*Extremely*), with scale marks at 0, 25, 50, 75, and 100.

Participants also provided ratings of their emotions toward each image ("How does the image make you feel?") using a 7-point scale (1=*not at all*; 7=*extremely*), including happy, sad, angry, fearful, physical disgust, and moral disgust.

After every 16-image rating set, participants were given a brief filler task that involved one of the following: solving three basic arithmetic math problems, writing a brief paragraph description of their morning routine, or writing a brief paragraph description of their evening routine. These were implemented to try and distract them from the specifics of the faces/images

they just saw, as well as break up the rating task into more manageable chunks. After the image ratings were completed, participants completed basic demographic questions about their age, gender, and race/ethnicity. Once the survey was completed, they were escorted out of the laboratory and granted compensation for their participation.

Results

Table 3 displays the severity and uncertainty means for each of the images, as well as the grand means across all images. Images were dummy-coded prior to analysis, and IAPS #2026 (neutral image) was used as the reference category. Across all images, the mean severity rating was $M = 54.53$ ($SD = 33.62$); about middle of the scale. Uncertainty fell just below the middle of the scale, $M = 42.38$ ($SD = 31.71$). Examining the mean severity ratings of each image, there appeared to be a wide range of severity ratings as well, ranging from the neutral image IAPS #2026 ($M = 14.59$, $SD = 19.74$) to the most severe image #3530 ($M = 86.17$, $SD = 16.55$). Figure 2 visually sorts the image severity ratings from most to least severe, highlighting the significant variance in ratings across images.

Table 4 displays the mean severity ratings for each facial expression condition. Across all images, the means are strikingly similar: neutral ($M = 54.32$, $SD = 33.68$), happy ($M = 54.82$, $SD = 33.53$), angry ($M = 54.63$, $SD = 33.38$), and disgust ($M = 54.77$, $SD = 33.53$). Indeed, it appears there was little difference between expressions when examined across the images. Looking again at Figure 2, it is clear that there is variability *between* images, but within each image there is little variability between the facial expression conditions.

Table 5 displays the bivariate relationships between severity ratings, other ratings, and uncertainty ratings. To control for the repeated measurement of variables within participants, each variable was mean-centered by subtracting the participant's overall mean on the variable

before estimating the correlation. Due to the high-powered nature of the sample and study design, nearly all correlations were significant at the $p<.05$ level. Therefore, I examined the relative size and direction of each pairing to look for patterns of relationships.

As expected, severity and others' ratings correlated strongly ($r = .82$). Severity and others' ratings both correlated with uncertainty, albeit weakly (r's = .05 and .08, respectively). The negative emotion ratings of sad, angry, fearful, physical disgust and moral disgust appear to correlate together tightly ($rs > .60$), indicating a tendency to rate similarly across these emotions. Happy was negatively associated with all variables ($rs > -.12$).

Severity was associated with the negative emotions such that as severity increased, these emotions also increased ($rs > .57$). Severity was also negatively associated with happiness ($r = -.20$). This same pattern was found for others' ratings as well, although associations were weaker overall. Finally, uncertainty correlated positively with the negative emotions, such that as uncertainty increased, the negative emotions tended to as well (and happiness tended to decrease). However, the relationships with uncertainty were the weakest across the entire set of bivariate relationships.

Main hypotheses. To test the hypothesis that emotional expressions conveyed by others will influence moral severity judgments by informing the target of relevant contextual evidence, I regressed severity ratings over facial expression condition (e.g., anger, disgust, etc.), image, and uncertainty. Two-way interactions between face condition x image and face x uncertainty, and image x uncertainty were also included. The face x uncertainty interaction was included as the key predictor, such that facial expressions would inform moral judgments under particularly uncertain situations.

Because participants saw multiple face-image sets, there is a source of non-independence in the data that needed to be modeled. Thus, we specified a random coeffient model by using the SPSS MIXED function, appropriate for repeated measure designs with nested data. I utilized maximum likelihood (ML) estimation with a covariance structure of homogeneous compound symmetry (CSR)[3], Type III sums of squares. As every participant rated every image x face combination, these were modeled as repeated effects nested within participants. The face x image interactions were also modeled to examine whether the effect of facial emotions varied across the IAPS stimuli. I also included predictors of gender and race to examine potential demographic influences on severity ratings. Because the racial groups had drastically uneven group sizes and the overall sample was 63% white, race was re-coded as a binary indicator (white/non-white). All significance tests were performed using two-tailed tests with $\alpha = 0.05$.

Table 6 shows the criterion ratings across each predictor, including the omnibus tests for each factor and specific parameter estimates. As seen in the omnibus test outcomes, there was a significant effect of image ($F(15, 14613) = 724.48$, $p < .001$). That is, there were significant differences in severity ratings across the images being rated. This makes intuitive sense, as they were selected specifically to provide a variety of stimuli to be rated by participants.

There was also a significant overall effect of uncertainty ($F(1, 14816) = 10.30$, $p = .001$). As seen in Table 6, higher levels of uncertainty are associated with higher levels of severity. That is, more uncertainty leads to slightly more severity, $\beta = .14$, $t(14678) = 5.58$, $p = .001$. This relationship is consistent with the bivariate relationship between uncertainty and severity ($r = .05$, $p < .05$) presented in Table 5.

Finally, there was a significant image x uncertainty interaction ($F(15, 14662) = 23.07$, $p < .001$), indicating that the relationship between severity and uncertainty varied across the

different images. Again, this makes intuitive sense as the images were also selected to vary on their level of ambiguity as well as their severity.

There were no significant effects of face ($F(3, 14606) = 0.33$, $p = n.s.$), face x uncertainty ($F(3, 14614) = 0.33$, $p = n.s.$), or face x image ($F(45, 14606) = 0.37$, $p = n.s.$) on moral severity of the images. That is, the facial expression conditions did not influence participants' overall severity judgments and did not interact with uncertainty or specific images.

Regarding demographic differences, there were no effects of gender ($F(1, 243) = 3.25$, $p = .073$) or race ($F(1, 243) = 3.48$, $p = 064$). While the data show that men ($M = 52.66$, $SD = 32.75$) rated the images as less severe than women ($M = 56.09$, $SD = 34.01$), this difference was not statistically significant.

Exploring alternatives. Considering the nature of morality research, and the vague colloquial definitions for terms like moral, ambiguous, and uncertain, it's possible these may be interpreted differently by native English speakers and non-native English speakers. Thus, I asked participants at the end of the survey whether English was their first language. The direction and pattern of significance for each predictor was the same for both groups and consistent with what is reported above.

I also collected participants' ratings of how they thought *other* people would rate the severity of the images. Other individuals' ratings may serve as a better measure for how emotions might be communicated to broader groups rather than individuals. I then ran the same model specified above but using others' ratings instead of participants' own severity ratings. Again, the pattern of results was the same as that specified above.

Discussion

The primary hypothesis tests indicated there were no observed effects of facial expressions on severity judgments and no effects of facial expressions on severity under uncertainty. While these results may indeed reflect the state of the relationship between facial expressions' communicative function and moral judgment severity, a number of methodological issues were identified that may have contributed to the results. These are discussed and addressed in Study 3, which sought to improve upon the present study to better test the primary hypotheses of interest.

CHAPTER 4

Study 3

After examination of the results of Study 2, I identified some methodological flaws that Study 3 was designed to overcome.

First, uncertainty was measured via self-reported ambiguity of each image. While the immorality of some actions and images were inherently more or less ambiguous, this was confounded with individual variance in how participants perceived the images and how they personally felt about the actions in each image. Subsequently, Study 3 experimentally manipulated uncertainty using a blurred version of each photo in order to impose situational uncertainty. Further, uncertainty was changed to a between-subjects factor so that participants would not be able to recognize the blurred versions of the images based on their viewing of the unaltered images.

Second, Study 2 presented the facial expressions side-by-side with the target images. While this was a straightforward implementation of the facial expression manipulation, the frequently changing faces being repeated across all 64 rating instances may have led participants to ignore the changing expressions. Further, due to the obvious presentation of the facial expressions as an experimental feature, and the exposure to all the facial expressions, participants may have disregarded the evidence in an effort not to be manipulated by their influence. Indeed, the presence of the face next to the target image allowed participants to fixate on it and consciously disregard its effects as non-credible (e.g., Schwarz & Clore, 2007). To overcome these issues, Study 3 presented each face briefly and prior to the presentation of each judgment stimuli.

Finally, the number of images implemented in the study was reduced from 16 to 8, and a distractor task was provided after every judgment task. This was designed to prevent carryover effects from the previous faces and images, and provide cleaner judgments about each image individually.

Aside from the aforementioned changes, the study procedure, questions, and materials were identical to that of Study 2. Therefore, I only present the parts of Study 3 below that diverge from Study 2's procedure.

Participants

Three-hundred fifty-eight undergraduate students volunteered to participate in a study titled "Interpersonal Perceptions 3" through the online subject pool recruitment website at Michigan State University. Individuals who participated in Study 1 and Study 2 were barred from participating in the present study, due to previous exposure to the judgment stimuli. After excluding those who failed to complete a sufficient number of the image ratings (n=5) or failed the judgment validation checks (n=2), the final sample size was N=351. No participants failed the self-reported attention check. Participants received course credit for their participation.

The mean age of participants was M = 19.33 (SD = 2.39), with approximately 81% identifying as female. The racial composition of the sample included 73% White, 14% Asian, 8% Black, and 5% other or not specified.

Methods & Procedure

Facial expression stimuli. The facial expressions were presented immediately prior to each target image for approximately 0.5 seconds. This allowed sufficient time for participants to view and register the facial expression before transitioning immediately into the judgment task. The same expressions (from subject #S037) were implemented, with the exception of the happy

expression. In Study 2, the happy expression did not influence judgments differently from the neutral expression, and it was removed for brevity in Study 3 (although ratings for all the emotions were retained). The presentation of the expressions was randomized across target images and participants.

Judgment stimuli. Eight of the 16 images were selected from those utilized in Study 2. The image number was reduced to decrease the number of judgment instances and help attenuate participant fatigue in making repetitive judgments. The particular images were chosen for their diversity of moral transgressions[4], as well as their distinctiveness in the blurred condition (see below).

Judgment under uncertainty. Uncertainty was experimentally manipulated by blurring the images using the Picture Tools function in Microsoft Word. The goal was to alter the images so that it was still apparent a moral transgression was occurring, but the details of which were unclear, creating a more ambiguous target image for judgment. After examining a number of blur options, it was determined that a 50% blur filter would be appropriate. This allowed the actions and context in the image to be discernible, but fuzzy enough so that the details could not be distinguished. For example, in the image of workplace harassment, it was clear that a female was sitting at a desk in a professional setting, with a male standing behind her with his hand on her shoulder. However, the details of their interaction and body language were difficult to discern, creating ambiguity in the nature of the interaction. The result is that it's apparent that the situation is in some way 'uncomfortable', but it's less obvious what exactly is happening in the situation.

Judgment task. The procedure followed the same format as in Study 2 but with the following changes: First, participants were not presented with the images prior to the rating of

them. This was to prevent participants from forming preconceptions of the images prior to viewing them for judgment. Second, participants were randomly assigned to either the uncertainty (blurred images) condition or the clear (unmodified images) condition, where they would see either 8 blurred target images for judgment or 8 clean images. Images were presented in random order within each condition. Each rating was followed by a filler task to distract participants from the judgments and attenuate any carryover effects image-to-image. Participants were first presented with a 6-digit number and asked to memorize it. The next screen asked them to solve three arithmetic problems (e.g., (3 x 12) -7 = ?). After, they were asked to recall the number.

Finally, the ratings of others' severity were removed (as self- and other-ratings yielded similar findings in Study 2), and the ambiguity question was changed slightly: "How certain are you of what's happening in the picture?" (0=*Not at all,* 100=*Extremely*). Higher values indicate higher certainty.

Results

Table 7 shows the mean severity and uncertainty ratings for both the clean and blurred image condition. The general mean trends reveal that clean images are rated as more severe than the blurred images, except for two images (IAPS # 2026 and 4233). A similar pattern emerged for certainty ratings—clean images are rated as more certain than the blurred images, except for three images (2026, 4233, and 9102).

The facial expression conditions show similar severity ratings across all the images, neutral M = 57.49 (SD = 33.90), disgust M = 58.73 (SD = 33.97), and anger M = 58.56 (SD = 33.82). Within the different images, the expressions don't show any clear patterns. Some images show the expected pattern of emotions on severity judgments (e.g., neutral < anger & disgust;

IAPS # 9423). However, some images show an ascending pattern where neutral > disgust > anger (3181), or a more 'flat' distribution where there are little or no differences across the conditions (9102).

Table 8 displays the aggregated bivariate correlations between severity, certainty, and emotion ratings across all images. The results show a very similar pattern to Study 2: the negative emotions correlate together strongly (*rs* > .58); happiness was negatively related to the negative emotions, severity, and certainty (*rs* > -.04). Severity and certainty were strongly correlated (*r* = .57), such that higher certainty was associated with higher severity ratings across conditions and images.

Main hypotheses. To assess whether the facial expressions influenced severity judgments, I implemented a similar procedure as outlined in Study 2 with severity regressed across facial expression condition (face), image, and uncertainty (blur) condition using the SPSS MIXED model procedure (restricted maximum likelihood, covariance type: compound symmetry correlation metric, type III SS). The facial expression and image factors were modeled as repeated effects nested within participants. I again included predictors of gender and race to examine potential demographic influences on severity ratings, and race was re-coded as a binary indicator (white/non-white). All significance tests were performed using two-tailed tests with α = 0.05.

To assess whether facial expression was more influential under conditions of uncertainty, I estimated two-way interactions between face and certainty (clean v. blur), image and certainty, and face and image. Finally, I modeled a three-way interaction of face x image x certainty to examine whether the relationship between facial expressions and level of certainty varied from image-to-image. Certainty was modeled as a between-subjects factor.

Table 9 shows the results of the model, including the omnibus tests for each factor and specific parameter estimates. Overall, the facial expression factor was not significant ($F(2, 2642) = 1.21$, $p = n.s.$), and the interaction terms were also not significant: face x certainty ($F(2, 2642) = 0.19$, $p = n.s.$), face x image ($F(14, 2639) = 1.40$, $p = n.s.$), face x certainty x image ($F(14, 2639) = 0.89$, $p = n.s.$). Thus, it appears the facial expressions do not have a significant influence on severity judgments for the images.

Image was found to be a significant factor, $F(7, 2402) = 223.60$, $p < .001$. This is expected and consistent with Study 2, as the images were selected for their variation in severity and content. The certainty (blur) manipulation was also significant, $F(1, 343) = 45.27$, $p < .001$. The blurred images were rated as significantly less severe ($M = 53.39$) than the clean images ($M = 63.22$). Further, the interaction between image and certainty condition was also significant, $F(7, 2403) = 44.90$, $p < .001$). Decomposing this interaction, the clean images are consistently rated as more severe than the blurred images, except for the two cases noted earlier (2026 and 4233) where the blurred images are rated as more severe. All mean differences are significant except for images 4233 and 9102. These effects are indicated on Table 7 and significant differences between means are denoted in bold (also see Figure 3).

Demographics. Race was not a significant factor in Study 3, $F(1, 347) = 0.73$, $p = n.s.$ Further, gender was not a significant predictor, $F(1, 345) = 3.13$, $p = n.s.$ To further test possible demographic differences, I also ran the model including interaction terms of gender with image, face, and blur condition. None of these interactions were significant and were not included in the final model.

Discussion

A similar pattern emerged here as in Study 2. Certainty was a significant predictor, such that those who saw the clean images rated them as significantly more severe than the blurred images except for the neutral image 2026, where the reverse was found. In that instance, it may be the case that obscuring the details of an innocent image created a sense of potential suspicion, as the image depicts a woman standing at a store checkout counter. However, the ratings in both instances were still quite low (M_{blur} = 17.67, M_{clean} = 7.82). In the other instances, it appears the blur reduced the severity of the perceived actions. This might be due to a reduction of culpability for the targets in the image or an apprehension to make a judgment without sufficient evidence. This is also congruent with the bivariate relationship between severity and self-reported certainty (r = .57), such that as certainty increased severity also increased.

Facial expression, gender, and race were not significant factors in Study 3, replicating the results from Study 2. This may be due to methodological limitations of the study, the stimuli used, or the undergraduate convenience sample. However, to ensure that the lack of evidence for emotional signals is not the result of the use of the specific visual expression stimuli used, Study 4 assessed moral severity judgments using auditory expressions to examine a different sensory input channel and experimental stimulus.

CHAPTER 5

Study 4

Both Studies 2 and 3 did not support the primary hypotheses that signals from the environment influence moral severity judgments, particularly under situations of uncertainty. While facial expressions are hypothesized to signal information about the environment to perceivers, it could be the case that the specific expressions used in Studies 2 and 3 were ineffective in eliciting the effect. Participants may have found the facial expression and target image pairings contrived, unbelievable, or may have perceived the facial expressions as unrelated to the scenes depicted in the images (i.e., spatially and/or temporally disconnected from their laboratory-based judgments).

Study 4 attempts to circumvent potential methodological issues with the facial expressions by implementing auditory expressions instead of visual facial expressions to communicate emotion. Emotional expressions are associated not only with physiological and somatic signals, but auditory sounds as well. For example, this could include cheering, jeering, yelling, screaming, or laughing, to name a few. Research in developmental psychology also shows evidence of an auditory expression effect in infants. Mumme, Fernald & Herrera (1966) found that pre-verbal infants were able to pick up on auditory expression (happy, neutral, or fear) and displayed avoidant behavior indicative of the fear responses given by the mother. Study 4 attempts to test the same hypotheses but implements an angry and neutral crowd sound manipulation as the expressions, as outlined below.

Participants

Seven-hundred and one undergraduate students volunteered to participate in a study titled "Interpersonal Perceptions 4" through the online subject pool recruitment website at Michigan

State University. Individuals who participated in Studies 1, 2, or 3 were barred from participating in the present study, due to previous exposure to the judgment stimuli. After excluding those who failed to complete a sufficient number of the image ratings or experienced technical difficulties (n=33), failed the judgment validation checks (n=50), failed the self-report attention check (n=5), the final sample size was N=613. Participants received course credit for their participation.

The mean age of participants was M = 19.51 (SD = 1.98), with 72.3% identifying as female. The racial composition of the sample included 68% White, 14% Asian, 11% Black, and 8% other or not specified.

Methods & Procedure

Auditory expressions. The auditory expressions were comprised of ambient 'crowd sounds' extracted from the internet (which were free for non-profit use). There were two conditions: angry and neutral. The neutral crowd sounded like a busy café or common crowded public area. It sounded as if there were a lot of people passively chatting and holding conversation. Particular words or snippets could be heard here or there, but no continuous conversations were noted. The angry condition, however, sounded like a crowd of jeering protesters. It sounded similar to the neutral condition but with bouts of shouting and yelling. Participants were randomly assigned to crowd condition, which was subsequently treated as a between-subjects variable in all analyses.

Judgment stimuli. Seven of the 8 images used in Study 3 were used in Study 4.[5] This included the same clean/blur variants for the uncertainty manipulation.

Judgment task. The procedure followed the same format as in Study 3 but with the following changes: First, participants were recruited in groups of 5 and seated at computer

workstations with barriers between them. Second, participants wore sound-isolating headphones for the judgment rating portion of the study. After participants reviewed the informed consent but before viewing the judgment stimuli, the research assistant started the randomly assigned crowd condition audio, which played for the duration of the judgment task. Prior to the judgment rating items, participants were instructed with the following: “Audio will be playing for the entirety of the session. While evaluating the scenes in the images try your best to imagine the audio is part of the scene, coming from a crowd of people out of view, as if these people are viewing the same actions depicted in the image as you are.” This was to ensure participants were engaged with the audio and connected it to the rating images being shown. Once the judgments were complete, participants were instructed to remove the headphones to complete the latter portion of the experiment.

In order to assess comfort and concern with uncertain judgments, participants were also asked two additional items (0=*Not at all*, 100=*Extremely*): 1. “Imagine you are on a jury and this scene is submitted as evidence for your jury decision. How comfortable would you be with making a judgment based only on this evidence?” 2. “Imagine your friend asked you to judge the actions in the scene, and then shared your response on social media to a local news outlet under your name. How worried would you be if the people involved in the scene found out about your judgment?”

Finally, to ensure the angry crowd manipulation was perceived as distinct from the neutral crowd, a manipulation check was performed at the end of the survey which asked participants “How did the crowd sound in the background? In a few words, describe what you think they sounded like”.

Results

For the manipulation check, responses were coded as 'angry' if they mentioned anything about the crowd sounding affectively negative or performing a negative act. For example, any mention of the crowd sounding angry, upset, perturbed, shouting, jeering, or enacting a riot, protest, or revolt was coded as 'angry'. Examining the results, 77% of the responses in the angry crowd condition were coded as angry, while only 7% were coded as such in the neutral crowd condition. This is highly suggestive that the angry crowd sounds were distinctly negative when compared to the neutral crowd sounds.

Table 10 shows the mean severity ratings for the angry and neutral crowd condition as well as the clean and blurred image condition. Across all images, the general mean trend revealed that clean images were rated as more severe than the blurred images (M_{clean} = 60.31, M_{blur} = 52.66), and participants in the angry crowd condition rated the images as more severe than those in the neutral crowd condition (M_{angry} = 57.71, $M_{neutral}$ = 55.69).

Table 10 also displays the bivariate correlations between severity, certainty, comfort, worry, and emotion ratings across all images, aggregated within participants. The results show a very similar pattern to Studies 2 and 3: the negative emotions correlate together strongly (*rs* > .58), and severity and certainty were strongly correlated (*r* = .57), such that higher certainty was associated with higher severity ratings across conditions and images. Happiness was negatively correlated with the negative emotions (*rs* > -.08), and negatively correlated with severity (*r* = -.25) and certainty (*r* = -.09).

Main hypotheses. To assess whether auditory expressions influenced severity judgments, I implemented a similar procedure as outlined in Study 3 with severity regressed across crowd condition, image, and certainty (blur) condition using the SPSS MIXED model

procedure (maximum likelihood, covariance type: unstructured, type III SS). The image factor was modeled as a repeated effect nested within participants, while crowd condition and uncertainty condition were modeled as between-subject effects. I again included predictors of gender and race to examine potential demographic influences on severity ratings, and race was re-coded as a binary indicator (white/non-white). All significance tests were performed using two-tailed tests with $\alpha = 0.05$.

To assess whether the crowd expression was more influential under conditions of uncertainty, or more influential for certain images, I estimated two-way interactions between crowd and certainty, image and certainty, and crowd and image. Finally, I modeled a three-way interaction of crowd x image x certainty to examine whether the relationship between facial expressions and level of uncertainty varied from image-to-image.

Table 11 shows the results of the model, including the omnibus tests for each factor and specific parameter estimates. Overall, the crowd factor was not a significant predictor of moral severity ($F(2, 610) = 1.18$, $p = n.s.$). Further, the crowd x blur interaction ($F(2, 609) = 0.03$, $p = n.s.$) and the three-way interaction ($F(16, 608) = 1.61$, $p = n.s.$) were not significant.

Consistent with studies 2 and 3, the image factor was found to be significant, $F(6, 608) = 564.17$, $p < .001$, demonstrating the variation in moral severity the scenes portray. The certainty (blur) manipulation was also significant, $F(1, 610) = 41.31$, $p < .001$, with participants consistently rating the blurred images less severely than the clean images (see Table 10). Further, the interaction between image and certainty was significant, $F(6, 608) = 69.55$, $p < .001$). Decomposing this interaction, the clean images are consistently rated as more severe than the blurred images ($p<.05$) except for two cases (IAPS image #3181 and #4233), for which there were no differences.

Finally, the crowd x image interaction was significant, ($F(6, 608) = 2.47$, $p = .023$). Decomposing this interaction, images #2026 (neutral) and #2691 (protester) showed heightened severity ratings in the angry crowd condition relative to the neutral crowd. The other images did not show this effect. See Figure 4 for a visual examination of mean differences.[6]

Exploring alternatives. To further examine the influence of expressions on moral judgments, I used the same modeling procedure to predict participants' responses to the two additional outcome measures (comfort with making a jury decision based on the evidence, and their concern if their judgment would be available on social media). See Table 10 for descriptive statistics.

For the comfort criterion, certainty ($F(1, 608) = 53.42$, $p < .001$), image ($F(6, 609) = 185.36$, $p < .001$), image x certainty ($F(1, 609) = 66.71$, $p < .001$), and race ($F(1, 610) = 16.56$, $p < .001$) were statistically significant. Unlike with severity judgments, the crowd x image effect was not significant ($F(1, 609) = 0.85$, $p = n.s.$). Aside from that latter point, the pattern of results is largely consistent with the other criterion outcomes across this study as well as studies 2 and 3.

For the concern criterion, only image ($F(1, 604) = 31.25$, $p < .001$) and image x certainty ($F(1, 604) = 3.17$, $p = .005$) were significant. Taken together, these indirect outcome measures may tap into slightly different aspects of the judgment process, but they are no more influenced by emotional expressions than the severity judgment outcome.

Finally, I examined the primary model using a subset of the data including only those who rated the angry crowd as 'angry'. This was intended to isolate the analysis only to those who actually perceived the angry crowd as angry. However, the pattern of significance of the results was the same as the overall model and did not show any deviations from the original model.

Demographics. Gender was not a significant factor in Study 3, $F(1, 610) = 0.01$, $p = n.s.$ However, race was found to be a significant predictor, $F(1, 620) = 5.48$, $p = .020$. Across images and conditions, non-white participants ($M = 58.23$, $SD = 33.06$) rated the images as more severe than white participants ($M = 55.96$, $SD = 33.87$).

Discussion

Overall the pattern of results was fairly consistent with the previous studies, with the exception of race and an image x crowd interaction predicting severity. As race was not found to be a significant predictor across Studies 2 and 3, this may be a random sampling effect such that non-white participants in this particular sample happen to find the images more severe than previous samples. Indeed, non-white participants found the images less severe than white participants in Study 2, and more severe in Study 3, further ambiguating the effect found in Study 4. As a result, this particular race finding should be interpreted with caution.

As for the image x crowd interaction, it appears the crowd expressions did have an effect on severity judgments for certain images. IAPS image #2026 is a neutral image depicting a woman standing at a store checkout counter. In the absence of other evidence in the image depicting an immoral act, the crowd may have supplied the needed inputs to evoke suspicion in the perceivers and inflate severity ratings. IAPS image #2691 depicts a protester in the streets throwing a rock toward an unknown target. It is possible that because there is a crowd in the image, and a protest is a typical place for a crowd, the match between the target stimulus and the crowd expression supplied relevant evidence for the judgment being made. These effects did not vary by level of certainty (e.g., the crowd x certainty x image interaction). The other images did not depict scenes that would necessarily have a crowd present, or not to the level of congruence that this particular image of a protest encapsulates.

CHAPTER 6

Discussion and Conclusion

Summary of Findings

The present studies assessed whether emotional expressions (both visual and auditory) influence moral severity judgments, including when the judgment is being made under uncertain conditions (defined here as a lack of clarity in available information about the situation at hand).

Study 1 was designed to assess morally relevant stimuli, and participants rated a subset of 42 images from the IAPS image database. Of those, a subset of images was selected for use in the subsequent studies that demonstrated normal rating distributions and presented adequate variability across both severity and ambiguity. We also ran a validation study to confirm that the images elicited a sufficient amount of moral reactance.[4] Indeed, these properties were also seen across Studies 2, 3, and 4; the images demonstrated variance in their severity ratings, as shown by the statistically significant differences in severity ratings across images.

Uncertainty was also found to have a significant impact on severity ratings. In Study 2, images with higher self-reported uncertainty were rated as significantly less severe than those with lower uncertainty. In Studies 3 and 4, the images with higher experimentally manipulated uncertainty (i.e., the 'blurred' images) were rated significantly less severe than the clean images.

In testing the primary hypothesis, across Studies 2 and 3 facial expressions (anger, disgust, vs. neutral expressions) were not found to influence judgments, either in a clear or uncertain context.

In Study 4 the expression stimulus was changed from a visual expression to an auditory expression (angry vs. neutral crowd sounds). The contextual uncertainty still did not have any influence on severity ratings. However, the angry crowd did increase severity ratings for two of

the images (IAPS #2026 and 2691). Thus, it may be the case that under certain circumstances, external expressions from others may provide relevant evidence for moral severity judgments.

Discussion

I set out to test whether third-party emotional expressions influenced the moral severity judgments of first-party perceivers of immoral actions, specifically when the act being assessed is uncertain or lacks clarity of information. Facial expressions, theorized to be evolved communicative devices (Shariff & Tracy, 2011), did not appear to influence perceivers' severity judgments in either clear or uncertain contexts. This could have been a methodological artifact; perhaps there was a disconnect between the target rating stimulus and the facial expression, or a single facial expression of one individual was not strong enough evidence to influence perceivers' judgments. Considering that a static facial expression is much different than an expression that unfolds actively during a situation, and that people are highly sensitive to change in the environment relative to a static stimulus, the effect of the faces may not have been strong enough to elicit an effect. Another possibility is that the evidence contained in the image for making an accurate assessment was already strong, and any additional effect of the expressions was superfluous to the initial perceiver assessment.

Additionally, this may be due to the bi-directionality of emotions. Emotions may influence the interpretation of a situation, as hypothesized here. However, the situation may also serve to disambiguate the emotion. In this way, the judgment stimuli may have been influencing the interpretation of the emotional expression and not the other way around. Either way, based on the narrow context of these studies and specificity of expression stimuli used, more work needs to be done before conclusions are made about the effectiveness of facial expressions as communicative devices in moral contexts.

The auditory expressions did appear to influence judgments in some cases. IAPS image #2026, the neutral image, may not have initially presented a case of an immoral act, but with an angry crowd virtually present on the scene, perhaps notably innocuous behavior can become suspicious to perceivers. Less innocuous behavior may be less susceptible to the influence of auditory communicative signals; as there is already an abundance of evidence for making a judgment, additional information becomes superfluous or corroborative in its impact. The influence of crowd sounds on severity judgments of the neutral image may also be due to floor effects. Specifically, ratings for the neutral image were low to begin with so an increase in moral severity is the most likely outcome. However, I find this less likely as the same effect was found in IAPS image #2691, where the neutral crowd was just above the mid-point of the rating scale. In this instance, the image depicted a protester throwing an object at an unknown target. The congruency between the protester and crowd sound (that a crowd is usually present at a protest) may have bolstered the relevance of the angry sounds and perceivers found it to be useful evidence in their judgment.

This latter point of congruency might also point to a more domain-specific connection between external expressions and judgments. That is, certain emotional expressions pair more strongly with a perceiver response when the judgment stimuli and emotional stimuli are from the same relevant domain (e.g., Garcia & Koelling, 1966). That might explain why the expressions from a crowd of people influence perceptions of a protester (crowds are known to be present at protests) or a woman waiting at a store (crowds may also be present in stores or public areas) but not the perception of other, less congruent situations (e.g., a sexual assault or one-on-one interpersonally violent situation). A promising future direction would be to assess the domain

specificity of emotional expressions, and the boundaries against which they may or may not impact others' judgments or behavior.

Aside from those two images, and barring methodological idiosyncrasies that may have influenced the effects of the studies as a whole, why wasn't external evidence influential for perceptions of moral transgressions? Jonathan Haidt's 'affect-first' perspective may help shed light on the situation. Perceivers of moral acts have an immediate, unconscious, affective reaction to the situation. Primarily autonomic in nature, it serves as the baseline response for the perceiver, characterized as a broadly valenced 'good' or 'bad' feeling (akin to an initial approach/avoidance behavior; Smillie, Pickering, & Jackson, 2006). Following this immediate reaction is a more conscious and deliberative period where the perceiver consciously processes the information at hand. Critical to his perspective is that the conscious rationalizing of the perceiver is *in the service of* the initial affective response. If the initial response is negatively valenced, then the conscious rationalizing will be in service of a negative viewpoint (and vice versa; Haidt, 2012).

This may partially explain why the severity ratings of perceivers in the present study were more or less impervious to external information sources. The initial reaction of the perceiver may have negated any subsequent influence from external sources, particularly when coupled with the contrived nature of experimental laboratory settings. Returning to the introductory example, imagine you come across a crowd of people while walking down the street. The crowd is surrounding two individuals, and you hear what sounds like a conflict. Are the two fighting? As you approach, you actively assess and make judgments about the situation. According to Haidt's view, whatever your initial reaction is to the situation will serve as a very strong valenced base that influences subsequent conscious assessments of the situation.

Certainly, concrete evidence contrary to your assessment would likely change your mind (e.g., if you initially thought they were fighting but they were acting). After changing your mind, you wouldn't judge the severity of their actions the same way, but the point illustrated is that the initial affective response is the starting point for subsequent cognitive processing about the situation. A future direction for research to explicitly test this possibility would be to examine whether initial affective responses affect the susceptibility of perceivers to use external communicative signals when constructing moral judgments.

Schwarz & Clore (2007) note that emotions and feelings contain information insofar as the information is relevant to the source at hand and not dismissible as either extraneous to the situation or attributable to an irrelevant source. In this way, they might argue that the external expressions of others were disconnected to the situation at hand and failed to evoke the appropriate feelings in the perceivers to influence judgments. Continuing with the same example, imagine that you approach the crowd of people it appears that two individuals are indeed having a fight. You wouldn't need to examine the rest of the crowd's facial expressions to make an evaluative judgment; their expressions would be redundant with the evidence you've gleaned with your own senses.

Further, this perspective might also suggest that any felt or perceived emotions might be attributable to either the facial expression seen or discounting the information of the facial expression as resulting from the immoral act in the image. Either explanation would suggest a disruption of the link between participants' internal states and their moral judgments.

Limitations

The studies presented are not without their limitations. Along with the aforementioned ambiguity in how participants' initial affective responses affect the susceptibility to external

information, there are a number of methodological ways that the current studies can be improved.

I went through several efforts to validate constituent components of each study. The pilot study and validation study allowed me to pre-select several stimuli that were appropriate for the studies. Likewise, the facial expressions (Kanade, Cohn, & Tian, 2000; Lucey et al., 2010) and auditory stimuli (see Study 4) both communicate the emotions they intended to successfully. However, the extent to which participants used these communicative devices to make judgments about the moral stimuli was the test of interest.

The current studies constituted a proximal test of how people use emotional expressions as communicative signals when constructing moral judgments. However, the designs of the current studies are far from an externally valid test of these processes. Laboratory studies are, by definition, contrived to some extent. Static facial expressions presented briefly before or during a moral judgment is very different context than an emergent judgment event encountered in everyday life. While the studies presented here are designed to assess whether expressions *can* influence severity judgments, the socioecological congruency to a real-world, active judgment event is lacking. Further, in a real judgment event, autonomic reactivity is likely stronger and more active. As autonomic nervous system reactivity (connected to emotional reactivity) is posited to play a role in these events. Reactivity is likely attenuated in a laboratory setting—the judgment effects found here may be less generalizable to the fast-paced reactivity present in ecologically valid situations. A more appropriate test may involve having confederates actively communicate their emotions (through facial or auditory cues) while an (ostensibly) ambiguously moral action is happening.

Further, the same judgment stimuli were used across all studies to control for varying targets across the experimental manipulations (e.g., facial expressions, crowds). These stimuli were pre-tested to ensure methodological and statistical viability during the analyses. However, if the images were not found to be morally compelling or motivating for judgment by participants in a laboratory, the judgments might not be as strong as the would be for more compelling stimuli. Without sufficient motivation to make a proper and accurate judgment, participants may have exercised less scrutiny in their assessments and felt less compelled to glean relevant evidence from the environment. This would have attenuated the effect of any external evidence.

Future research can examine the extent to which participants use emotional expressions in contexts such as those featured in the current studies, if at all. In the current studies, I did not explicitly ask the participants if they used the emotional stimuli when trying to discern the events of an uncertain situation. It could be that moral judgments were indeed affected, but only among those who reported using the emotional expressions as information when forming judgments. However, asking such a question might also nullify the effects of communicative emotional expressions on judgments (e.g., Schwarz & Clore, 2007).

Conclusion

There may be certain situations where the emotional expressions of others can influence someone's judgments. In the current studies, I tested whether exposure to emotional expressions affected subsequent moral judgments under uncertainty. The results from Studies 2-4 did not support this possibility. However, the specific domains and expressions that might evoke this phenomenon need additional empirical attention before a strong conclusion can be made. I identified a number of plausible explanations and hypotheses that can be tested in future

research. The hope for this research program is that future researchers can contextualize moral judgments within embedded social structures that allow us to form judgments and evaluations based on salient information in our environments.

FOOTNOTES

[1] This also prevents the disclosure of the images in their visual form within publications and manuscripts. However, a description of each image included in the study is provided in the appendix.

[2] Participants were informed at the beginning of the survey and in the consent form that they would be asked to view images that might make them uncomfortable, but that were not more graphic than what might be seen in a Rated-R film. However, due to the sensitive and graphic nature of some IAPS images assessed in this study, a number of participants declined participation or opted not to respond to multiple items in the survey.

[3] I attempted to implement competing covariance structures that may better fit the model, including heterogeneous compound symmetry and unstructured. However, the model would not converge under these specifications.

[4] An additional way of evaluating the moral nature of the stimuli is to have people generate stories about each image and code for the presence of moral features. To this end, a sample of N=105 participants was recruited via Amazon Mechanical Turk to examine if people naturally interpret the images in a moral framework outside of the experimental context. Each participant viewed a random 4 of the 8 images and provided an open-ended 3-5 sentence 'story' of what they thought the people in the images were doing (similar to instructions for the Thematic Apperception Test). They were coded by research assistants as either moral (containing any morally relevant themes from a provided list; e.g., fairness, interpersonal harm, disgust, etc.) or non-moral. The neutral image (IAPS #2026) was rated as containing moral features in only 15% of stories, suggesting that it did little to elicit moral concerns. All other images had a minimum proportion of 75% moral features (#9102), up to a high of 98% moral features (#3530). The pattern suggests the target images are being interpreted in a moral framework as expected, while the neutral image is not.

[5] Due to a methodological oversight, one of the images had to be dropped from the analyses for Study 3. However, as the pattern of results was largely consistent with the previous studies. I do not believe its inclusion would have influenced the conclusions made as, for the most part, the effects of certainty and crowd did not differ across the image set.

[6] As an exploratory investigation into predictors of uncertainty and severity judgments, I included a handful individual difference measures in addition to the standard demographic items for participants to complete. This included the intolerance for uncertainty scale (Freeston, Rhéaume, Letarte, Dugas, & Ladouceur, 1994), need for closure scale (Kruglanski, Atash, De Grada, Mannetti, & Pierro, 2013), the moral foundations questionnaire (Graham, 2011), and the right-wing authoritarianism scale (Altemeyer, 1981). I regressed both severity and uncertainty judgments separately on all of these measures. For severity ratings, only the purity and harm factors from the moral foundations questionnaire were significantly related to severity ($F(1, 492) = 3.93$, $p = .048$, $\beta = .34$; $F(1, 490) = 13.36$, $p < .001$, $\beta = .52$). Higher scores on both of these factors were associated with higher severity ratings. The other individual difference characteristics were unrelated to severity judgments. For uncertainty, the same pattern of results was found, with both the purity and harm factors being significantly related to self-reported certainty ratings ($F(1, 492) = 3.91$, $p = .049$, $\beta = .33$; $F(1, 492) = 7.54$, $p = .006$, $\beta = .55$). Higher values in both of these factors were associated with higher certainty ratings. The other individual difference characteristics were unrelated to certainty judgments.

APPENDIX

Table 1:

Study 1 IAPS image mean ratings (SD) and descriptive statistics, including severity, relevance, ambiguity, and emotions.

IAPS Image #	Relevance	Severity	Uncertainty	Happy	Sad	Angry	Fearful	P. Disgust	M. Disgust
Total	4.06	4.42	3.08	1.48	3.99	3.79	3.63	3.75	4.01
	(2.26)	(2.34)	(2.04)	(1.22)	(2.42)	(2.46)	(3.63)	(2.44)	(2.46)
1112	Snake								
	2.13	1.72	2.88	1.36	1.76	1.56	3.86	2.59	1.50
	(1.62)	(1.28)	(2.08)	(1.00)	(1.49)	(1.27)	(2.20)	(2.09)	(1.16)
1201	Spider								
	1.96	1.72	2.86	1.31	1.61	1.62	4.20	3.66	1.59
	(1.53)	(1.32)	(2.12)	(0.89)	(1.39)	(1.41)	(2.30)	(2.31)	(1.33)
2026	Woman store								
	2.35	1.61	3.11	1.43	1.41	1.34	1.30	1.26	1.30
	(1.80)	(1.13)	(2.19)	(1.01)	(1.06)	(0.97)	(0.91)	(0.87)	(0.93)
2039	Woman table								
	2.39	1.53	3.14	1.54	2.90	1.41	1.38	1.28	1.32
	(1.81)	(1.12)	(2.16)	(1.08)	(1.97)	(1.02)	(1.01)	(0.86)	(0.91)
2278	Refugees								
	3.51	2.83	3.91	1.49	3.77	2.45	2.62	2.09	2.41
	(1.79)	(1.79)	(1.89)	(1.03)	(2.11)	(1.84)	(1.86)	(1.64)	(1.85)
2373	Mariachis								
	2.37	1.33	2.52	3.89	1.16	1.15	1.15	1.16	1.16
	(1.95)	(0.92)	(2.08)	(2.08)	(0.63)	(0.61)	(0.64)	(0.68)	(0.65)
2377	Studying								
	2.69	1.33	2.52	2.76	1.29	1.19	1.24	1.19	1.20
	(2.27)	(1.00)	(2.13)	(1.95)	(0.95)	(0.78)	(0.90)	(0.79)	(0.80)
2390	Talking at table								
	2.55	1.41	2.85	2.34	1.45	1.21	1.34	1.22	1.24
	(1.98)	(0.97)	(2.09)	(1.74)	(1.10)	(0.75)	(0.99)	(0.81)	(0.82)
2488	Accordion player								
	2.40	1.39	2.60	3.32	1.32	1.24	1.26	1.23	1.24
	(1.96)	(1.03)	(2.05)	(2.00)	(0.95)	(0.87)	(0.95)	(0.90)	(0.91)
2590	Coffee older woman								
	2.79	1.76	3.31	2.51	2.07	1.42	1.45	1.34	1.38
	(1.89)	(1.27)	(1.98)	(1.82)	(1.53)	(1.04)	(1.02)	(0.91)	(0.96)
2691	Protester								
	4.25	4.50	4.05	1.26	3.42	3.42	3.75	2.83	3.39
	(1.77)	(1.55)	(1.73)	(0.74)	(1.97)	(1.95)	(1.97)	(1.89)	(1.91)
2717	Intravenous drug use								
	4.41	5.02	3.21	1.16	4.54	3.62	3.69	4.39	4.40
	(1.97)	(1.59)	(1.88)	(0.69)	(2.04)	(2.21)	(2.22)	(2.07)	(2.10)

Table 1 (cont'd)

2745	Store theft								
	4.47	4.93	2.87	1.17	3.08	3.30	1.98	2.11	3.80
	(1.90)	(1.51)	(1.71)	(0.63)	(2.01)	(1.96)	(1.66)	(1.68)	(1.87)
2751	Drinking and driving								
	5.06	5.80	2.61	1.18	4.62	5.36	5.01	4.06	5.32
	(2.01)	(1.42)	(1.92)	(0.72)	(2.14)	(1.89)	(2.05)	(2.21)	(1.85)
2981	Animal harm/hunting								
	4.08	4.63	3.29	1.15	4.81	4.18	3.08	4.89	4.27
	(1.92)	(1.81)	(1.74)	(0.55)	(2.12)	(2.26)	(2.17)	(2.11)	(2.16)
3181	Domestic abuse								
	3.76	4.05	4.11	1.23	3.77	3.11	3.55	3.80	3.56
	(1.76)	(1.79)	(1.80)	(0.78)	(2.20)	(2.16)	(2.11)	(2.15)	(2.17)
3500	Gun threaten								
	5.04	6.12	2.78	1.14	5.03	5.02	5.20	4.51	5.43
	(2.15)	(1.22)	(1.96)	(0.66)	(1.96)	(1.99)	(1.89)	(2.17)	(1.70)
3530	Gun threaten 2								
	4.79	6.00	3.01	1.15	4.92	4.90	5.08	4.72	5.28
	(2.12)	(1.31)	(2.04)	(0.66)	(2.04)	(2.00)	(2.00)	(2.11)	(1.85)
4232	Masturbation								
	3.25	3.10	3.30	1.79	1.90	1.87	1.69	2.61	2.55
	(1.92)	(2.03)	(1.86)	(1.49)	(1.70)	(1.64)	(1.47)	(2.07)	(2.06)
4233	Prostitution								
	4.01	4.20	3.48	1.27	3.28	2.51	2.31	2.92	3.63
	(1.85)	(1.75)	(1.91)	(0.81)	(2.18)	(1.91)	(1.90)	(1.99)	(2.11)
4621	Workplace harassment								
	4.31	4.70	3.91	1.27	3.41	4.01	3.72	3.73	4.18
	(1.88)	(1.72)	(1.89)	(0.90)	(2.10)	(2.12)	(2.12)	(2.13)	(2.07)
5875	Man on bike								
	2.41	1.86	3.28	2.41	1.70	1.50	1.46	1.44	1.53
	(1.71)	(1.30)	(2.13)	(1.81)	(1.30)	(1.13)	(1.07)	(1.08)	(1.19)
6312	Woman assault van								
	4.98	6.08	3.07	1.11	5.14	5.56	5.25	4.97	5.73
	(2.09)	(1.27)	(2.11)	(0.54)	(1.99)	(1.80)	(2.00)	(2.03)	(1.64)
6313	Knife threaten								
	5.11	6.41	2.48	1.19	5.32	5.67	5.52	5.24	5.97
	(2.19)	(1.06)	(1.94)	(0.79)	(1.87)	(1.67)	(1.82)	(1.96)	(1.45)
6315	Assault slap								
	5.13	6.18	2.73	1.11	5.44	5.91	5.04	5.17	5.92
	(2.07)	(1.22)	(2.08)	(0.58)	(1.79)	(1.54)	(2.16)	(2.07)	(1.53)
6415	Animal slain								
	4.83	6.06	2.87	1.14	5.86	5.68	4.36	5.75	5.78
	(2.21)	(1.42)	(2.05)	(0.73)	(1.69)	(1.83)	(2.36)	(1.79)	(1.74)
6520	Knife threaten 2								
	4.99	6.17	3.01	1.14	5.56	5.34	5.20	5.18	5.68
	(2.15)	(1.23)	(2.10)	(0.60)	(1.68)	(1.85)	(1.91)	(1.95)	(1.62)

Table 1 (cont'd)

7506	Gambling								
	3.12	2.48	3.17	1.69	1.61	1.53	1.53	1.39	1.73
	(1.75)	(1.43)	(1.86)	(1.26)	(1.20)	(1.20)	(1.18)	(1.05)	(1.30)
9007	Drug paraphernalia								
	4.46	5.31	3.68	1.16	4.76	3.85	4.05	4.61	4.71
	(2.02)	(1.58)	(1.99)	(0.64)	(2.00)	(2.19)	(2.22)	(2.06)	(1.97)
9102	Drug injection								
	4.33	4.92	3.21	1.21	4.40	3.53	3.56	4.36	4.27
	(2.01)	(1.70)	(1.92)	(0.80)	(2.12)	(2.22)	(2.23)	(2.09)	(2.14)
9145	Cow branding								
	3.97	4.45	3.29	1.28	4.64	4.03	3.10	4.18	4.02
	(1.85)	(1.86)	(1.77)	(0.93)	(2.06)	(2.22)	(2.18)	(2.16)	(2.19)
9163	Soldier abuse								
	4.99	5.81	3.27	1.17	5.29	5.23	4.79	4.79	5.45
	(2.04)	(1.40)	(2.04)	(0.69)	(1.86)	(1.87)	(2.07)	(2.13)	(1.73)
9183	Dog abuse								
	5.03	6.30	2.57	1.16	6.33	6.09	4.69	6.01	6.23
	(2.24)	(1.28)	(2.10)	(0.77)	(1.35)	(1.59)	(2.27)	(1.63)	(1.46)
9252	Corpse dragged								
	5.06	6.47	2.61	1.15	5.83	5.77	5.33	6.10	6.24
	(2.28)	(1.15)	(2.08)	(0.69)	(1.69)	(1.71)	(2.01)	(1.54)	(1.35)
9253	Dead woman								
	5.15	6.40	2.87	1.16	6.06	5.89	5.44	6.16	6.21
	(2.24)	(1.19)	(2.10)	(0.73)	(1.50)	(1.70)	(2.02)	(1.47)	(1.39)
9295	Ocean pollution								
	4.98	5.39	2.72	1.17	5.29	5.11	4.10	4.46	5.04
	(1.96)	(1.61)	(1.78)	(0.78)	(1.82)	(1.91)	(2.26)	(2.16)	(1.85)
9410	Boy's body								
	4.98	5.26	3.70	1.19	6.18	5.32	4.95	5.64	5.40
	(2.10)	(1.94)	(2.22)	(0.75)	(1.42)	(2.00)	(2.14)	(1.81)	(2.00)
9413	Men hanged								
	5.08	6.28	2.93	1.13	5.63	5.31	5.20	5.26	5.82
	(2.22)	(1.13)	(2.13)	(0.65)	(1.69)	(1.92)	(1.98)	(1.99)	(1.59)
9414	Gun threaten assault								
	4.94	6.15	3.16	1.16	5.49	5.49	5.13	5.06	5.73
	(2.19)	(1.29)	(2.08)	(0.70)	(1.84)	(1.80)	(2.05)	(2.06)	(1.60)
9423	Hostages in pit								
	4.89	5.87	3.20	1.14	5.42	5.15	5.16	4.64	5.40
	(2.13)	(1.56)	(2.05)	(0.64)	(1.82)	(1.96)	(2.00)	(2.18)	(1.87)
9800	Neo-nazi teen								
	4.62	5.67	2.91	1.17	4.42	5.07	4.44	4.12	5.26
	(2.13)	(1.54)	(1.95)	(0.63)	(2.23)	(2.02)	(2.17)	(2.28)	(1.95)
9810	KKK burning cross								
	5.03	6.22	2.45	1.23	5.42	5.80	5.31	5.17	6.03
	(2.28)	(1.45)	(2.06)	(0.90)	(2.04)	(1.85)	(2.10)	(2.23)	(1.68)

Table 2:

Study 1 judgments, bivariate correlations (Pearson's r)

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. Relevance	-								
2. Severity	**.54**	-							
3. Uncertainty	**.08**	**-.03**	-						
4. Happy	**-.09**	**-.33**	**.04**	-					
5. Sad	**.43**	**.72**	-.01	**-.26**	-				
6. Angry	**.44**	**.78**	**-.06**	**-.25**	**.85**	-			
7. Fearful	**.34**	**.64**	.00	**-.23**	**.74**	**.77**	-		
8. P. Disgust	**.37**	**.69**	**-.02**	**-.25**	**.77**	**.81**	**.75**	-	
9. M. Disgust	**.48**	**.84**	**-.06**	**-.29**	**.83**	**.90**	**.74**	**.82**	-

Note: P. Disgust = physical disgust; M. Disgust = Moral disgust. Bold values significant at $p < .01$.

Table 3:

Study 2 IAPS image mean ratings (SD) and descriptive statistics, including severity, others' perceived ratings, ambiguity, and emotions.

IAPS Image #	Severity	Others	Uncertainty	Happy	Sad	Angry	Fearful	P. Disgust	M. Disgust
Total	54.53	56.55	42.38	1.25	3.33	3.07	2.80	2.91	3.35
	(33.62)	(29.54)	(31.71)	(.86)	(2.23)	(2.21)	(2.14)	(2.17)	(2.22)
2026	Woman store								
	14.59	17.07	41.88	1.27	1.39	1.36	1.24	1.27	1.34
	(19.74)	(19.63)	(37.50)	(.93)	(1.05)	(.98)	(.77)	(.84)	(.94)
2278	Refugees								
	28.11	30.37	42.55	1.4	3.58	2.27	2.37	1.91	2.20
	(30.17)	(28.84)	(34.66)	(1.10)	(2.21)	(1.86)	(1.9)	(1.59)	(1.82)
2691	Protester								
	54.07	57.99	49.68	1.24	2.95	3.02	3.21	2.30	2.91
	(25.12)	(21.97)	(26.84)	(.74)	(1.96)	(1.93)	(2.01)	(1.77)	(1.88)
2745	Store theft								
	59.68	62.81	37.65	1.14	2.76	2.76	1.75	1.94	3.18
	(25.83)	(23.28)	(28.10)	(.54)	(1.85)	(1.81)	(1.43)	(1.55)	(1.88)
2981	Animal harm/hunting								
	58.98	56.72	40.04	1.15	4.25	3.63	2.74	4.02	3.73
	(31.29)	(24.53)	(28.55)	(.56)	(2.16)	(2.26)	(2.06)	(2.25)	(2.25)
3181	Domestic abuse								
	46.43	50.89	51.19	1.12	2.90	2.53	3.09	3.20	2.86
	(29.69)	(26.57)	(30.56)	(.54)	(2.00)	(1.95)	(2.08)	(2.05)	(2.03)
3530	Gun threaten 2								
	86.17	85.49	42.24	1.11	4.47	4.37	4.55	3.96	4.83
	(16.55)	(15.60)	(35.00)	(.62)	(2.09)	(2.13)	(2.12)	(2.29)	(2.00)
4232	Masturbation								
	30.80	44.78	36.40	1.89	1.59	1.45	1.42	2.00	2.12
	(30.22)	(25.71)	(28.87)	(1.67)	(1.34)	(1.13)	(1.14)	(1.66)	(1.66)
4233	Prostitution								
	47.23	57.13	45.15	1.25	2.73	1.97	1.80	2.24	2.77
	(28.16)	(25.73)	(28.28)	(.82)	(1.96)	(1.56)	(1.45)	(1.68)	(1.83)
4621	Workplace harassment								
	63.36	60.82	50.19	1.16	3.44	3.83	3.27	3.61	4.16
	(28.35)	(25.53)	(31.77)	(.66)	(2.14)	(2.16)	(2.17)	(2.19)	(2.12)
7506	Gambling								
	22.06	32.69	34.19	1.58	1.47	1.41	1.42	1.30	1.58
	(22.65)	(22.73)	(28.08)	(1.23)	(1.05)	(.99)	(1.03)	(.82)	(1.11)
9102	Drug injection								
	65.04	69.13	43.15	1.17	3.84	2.95	3.03	3.70	3.75
	(28.85)	(23.86)	(38.74)	(.67)	(2.20)	(2.13)	(2.13)	(2.26)	(2.16)
9145	Cow branding								

Table 3 (cont'd)

	54.54	51.69	38.74	1.16	4.01	3.36	2.50	3.41	3.47
	(31.04)	(24.04)	(27.76)	(.61)	(2.09)	(2.19)	(1.94)	(2.14)	(2.15)
9295	Ocean pollution								
	72.34	63.95	37.34	1.11	4.84	4.68	3.75	3.93	4.71
	(24.59)	(24.37)	(31.34)	(.51)	(1.97)	(2.03)	(2.18)	(2.17)	(1.98)
9423	Hostages in pit								
	83.69	81.61	45.90	1.11	5.02	4.82	4.72	4.10	5.10
	(21.97)	(21.16)	(35.48)	(.57)	(1.96)	(2.11)	(2.11)	(2.35)	(2.04)
9800	Neo-nazi teen								
	76.14	75.16	41.27	1.12	4.11	4.67	3.89	3.75	4.86
	(26.51)	(23.57)	(34.45)	(.63)	(2.25)	(2.18)	(2.25)	(2.38)	(2.12)

Table 4:

Study 2 severity judgments by facial expression condition.

Facial expression	*M*	*SD*
Neutral	54.32	33.68
Happy	54.82	33.53
Angry	54.63	33.38
Disgust	54.77	33.53

Table 5:

Study 2 judgments, bivariate correlations (Pearson's r) and demographic means.

	1.	2.	3.	4.	5.	6.	7.	8.
1. Severity	-							
2. Projecting Others	.82	-						
3. Uncertainty	.05	.08	-					
4. Happy	-.20	-.15	-.03	-				
5. Sad	.61	.50	.02	-.16	-			
6. Angry	.70	.56	.01	-.15	.75	-		
7. Fearful	.57	.50	.09	-.12	.64	.68	-	
8. P. Disgust	.59	.49	.04	-.14	.64	.69	.60	-
9. M. Disgust	.77	.62	.00	-.16	.70	.81	.65	.74

	Severity		Ambiguity	
	M	*SD*	*M*	*SD*
Gender				
Male	52.66	32.75	42.29	30.52
Female	56.09	34.01	42.85	32.50
Race				
Black	60.54	35.17	46.74	31.91
White	53.24	30.99	42.01	32.07
Asian	55.81	35.52	40.21	29.75
Other/not specified	53.24	33.95	47.58	33.49

Note: P. Disgust = physical disgust; M. Disgust = Moral disgust. Variables were mean-center aggregated within each participant to account for repeated measures within each subject. All correlations significant at $p < .05$.

Table 6:

Study 2 severity ratings, omnibus effects and parameter estimates.

Predictor		*df*	*F*	*b*	*t*	*p*	*95% CI* *Upper*	*Lower*
Gender		1	3.25			.073		
	Male			2.96	1.80	.073	-0.27	6.19
Race		1	3.48			.064		
	Non-white			-3.14	1.86	.064	-6.45	0.18
Face		3	0.47			.703		
	Anger				0.29	.771	-3.70	4.99
	Disgust				0.71	.478	-2.79	5.95
	Happy				0.52	.605	-3.22	5.52
Uncertainty		1	10.30	0.14	5.58	.001	0.09	0.18
Image		15	724.48			< .001		
	2278			12.19	6.47	< .001	10.14	16.54
	2691			38.28	17.45	< .001	33.98	42.58
	2745			44.54	20.40	< .001	40.26	48.82
	2981			43.43	19.86	< .001	39.14	47.72
	3181			30.99	14.12	< .001	26.69	35.29
	3530			71.28	32.60	< .001	67.00	75.57
	4232			18.27	8.27	< .001	13.94	22.60
	4233			33.70	15.43	< .001	29.42	37.98
	4621			49.63	22.61	< .001	45.33	53.94

Table 6 (cont'd)

	7506			8.35	3.77	< .001	4.00	12.69
	9102			50.91	23.32	< .001	46.63	55.19
	9145			39.33	18.01	< .001	35.05	43.61
	9295			56.89	25.86	< .001	52.57	61.20
	9423			67.96	30.99	< .001	63.66	72.26
	9800			61.15	27.92	< .001	56.85	65.44
Face x Uncertainty		3	0.33			.806		
Face x Image		45	0.37			1.00		
Image x Uncertainty		15	23.07			< .001		

Note: Regression coefficients are unstandardized. Categorical variables dummy coded prior to analysis. Reference categories for Gender = Female, Race = White, Face = Neutral, Image = 2026 (neutral).

Table 7:

Study 3 IAPS mean severity and uncertainty ratings (SD) by image, facial expression.

IAPS Image #	Severity (clean)	Severity (blurred)	Certainty (clean)	Certainty (blurred)	Face (neutral)	Face (disgust)	Face (anger)
Total*	63.22	53.39	58.82	51.30	57.49	58.73	58.56
	(34.64)	(32.41)	(31.57)	(32.27)	(33.90)	(33.97)	(33.82)
2026*	Woman store						
	8.88	18.29	33.98	36.75	10.26	15.86	14.76
	(13.76)	(22.96)	(32.53)	(29.70)	(15.50)	(22.50)	(19.63)
2691*	Protester						
	71.33	60.05	62.17	53.90	66.10	64.25	66.61
	(24.41)	(24.90)	(25.18)	(25.93)	(26.83)	(24.17)	(24.84)
3181*	Domestic abuse						
	54.09	44.48	38.39	27.27	51.22	50.48	46.06
	(29.79)	(28.62)	(28.49)	(25.67)	(29.24)	(29.21)	(30.20)
3530*	Gun threaten 2						
	93.92	55.72	79.66	43.54	74.00	73.40	76.36
	(11.24)	(30.35)	(22.61)	(27.47)	(29.61)	(31.24)	(28.96)
4233	Prostitution						
	56.75	64.08	61.74	71.10	57.72	65.24	58.44
	(30.28)	(27.18)	(29.29)	(25.84)	(28.57)	(28.05)	(29.85)
4621*	Workplace harassment						
	65.70	55.16	56.92	56.77	60.44	58.72	61.99
	(29.50)	(30.86)	(29.76)	(27.87)	(30.77)	(31.29)	(29.93)
9102	Drug injection						
	68.11	74.33	70.33	82.39	72.88	69.59	71.24
	(30.77)	(27.97)	(29.12)	(21.71)	(27.62)	(32.26)	(28.66)
9423*	Hostages in pit						
	86.83	54.99	67.36	38.83	66.79	72.90	72.69
	(20.64)	(33.96)	(26.47)	(32.83)	(32.94)	(32.35)	(31.57)

Note: * Severity ratings significantly different at $p<.05$.

Table 8:

Study 3 judgments, bivariate correlations (Pearson's r) and demographic means across both the clean and blur conditions.

	M (SD)	1.	2.	3.	4.	5.	6.	7.
1. Severity	58.26 (33.89)	-						
2. Certainty	55.03 (32.14)	.57	-					
3. Happy	1.26 (0.81)	-.21	-.04	-				
4. Sad	3.39 (2.21)	.66	.50	-.14	-			
5. Angry	3.17 (2.17)	.67	.46	-.15	.70	-		
6. Fearful	3.42 (2.18)	.58	.29	-.13	.64	.64	-	
7. P. Disgust	3.36 (2.20)	.66	.47	-.16	.64	.67	.58	-
8. M. Disgust	3.66 (2.22)	.79	.54	-.17	.69	.76	.60	.78

	Severity		Certainty	
	M	*SD*	*M*	*SD*
Gender				
Male	54.93	34.59	55.23	32.81
Female	59.01	33.70	54.96	31.96
Race				
Black	60.85	37.38	63.61	33.85
White	58.13	33.75	53.83	32.02
Asian	56.90	32.40	55.80	30.06
Other/not specified	59.72	34.34	57.17	34.67

Note: P. Disgust = physical disgust; M. Disgust = Moral disgust. Variables were mean-center aggregated within each participant to account for repeated measures within each subject. All correlations significant at $p < .05$.

Table 9:

Study 3 severity ratings, omnibus effects and parameter estimates.

Predictor		***df***	***F***	***b***	***t***	***p***	***95% CI***	
							Upper	***Lower***
Gender		1	3.13			.078		
	Female			3.27	1.64	.078	-0.36	6.90
Race		1	0.73			.395		
	Non-white			1.39	0.85	.395	-1.81	4.58
Face		2	1.21			.297		
	Disgust			4.40	0.91	.362	-5.06	13.87
	Anger			5.58	1.16	.246	-3.84	14.99
Certainty		1	45.27			< .001		
	Blur			10.11	2.09	.037	0.61	19.61
Image		7	223.60			< .001		
	2691			67.68	14.37	.001	58.44	76.91
	3181			48.45	10.18	< .001	39.12	57.78
	3530			87.93	18.39	< .001	78.55	97.30
	4233			45.93	9.67	< .001	36.62	55.24
	4621			57.44	11.97	< .001	48.03	66.85
	9102			67.41	14.30	< .001	58.17	76.66
	9423			78.90	16.64	< .001	69.61	88.20
Face x Certainty		2	0.19			.825		
Image x Certainty		7	44.90			< .001		

Table 9 (cont'd)

Face x Image	14	1.40	.143
Face x Image x Certainty	14	0.89	.571

Note: Regression coefficients are unstandardized. Categorical variables dummy coded prior to analysis. Reference categories for Gender = Male, Race = White, Face = Neutral, Image = 2026 (neutral), Certainty = clean image.

Table 10:

Study 4 judgments, bivariate correlations (Pearson's r) and demographic means across experimental conditions.

	M (SD)	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. Severity	56.69 (33.62)	-								
2. Certainty	58.69 (31.05)	.57	-							
3. Comfort	48.85 (32.85)	.65	.73	-						
4. Worried	41.89 (30.58)	.24	.06	.09	-					
5. Happy	1.25 (.73)	-.25	-.09	-.11	-.08	-				
6. Sad	3.55 (2.21)	.65	.43	.51	.21	-.19	-			
7. Angry	3.12 (2.12)	.64	.38	.51	.19	-.19	.71	-		
8. Fearful	3.51 (2.18)	.55	.25	.35	.26	-.18	.62	.65	-	
9. P. Disgust	3.37 (2.13)	.65	.37	.47	.20	-.20	.67	.64	.58	-
10. M. Disgust	3.68 (2.17)	.78	.46	.58	.24	-.22	.71	.74	.60	.77

	Certainty condition				Crowd condition			
	Blur		Clean		Angry		Neutral	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Severity	52.66	31.84	60.31	34.75	57.71	33.65	55.69	33.57
Gender								
Male	53.57	31.49	60.63	34.40	57.60	32.61	57.13	34.05
Female	52.33	31.98	60.35	34.93	57.74	34.10	55.33	33.44
Race								
Black	50.89	32.10	59.14	36.92	57.20	35.66	53.34	34.29
White	52.23	32.12	59.64	35.13	56.71	33.97	55.24	33.76
Asian	55.76	30.16	60.06	32.32	59.54	31.31	56.95	31.60
Other/not specified	53.51	31.47	67.06	32.56	64.40	30.73	59.64	34.39

Table 10 (cont'd)

Note: P. Disgust = physical disgust; M. Disgust = Moral disgust. Variables were mean-center aggregated within each participant to account for repeated measures within each subject. All correlations significant at $p < .01$.

Table 11:

Study 4 severity ratings, omnibus effects and parameter estimates.

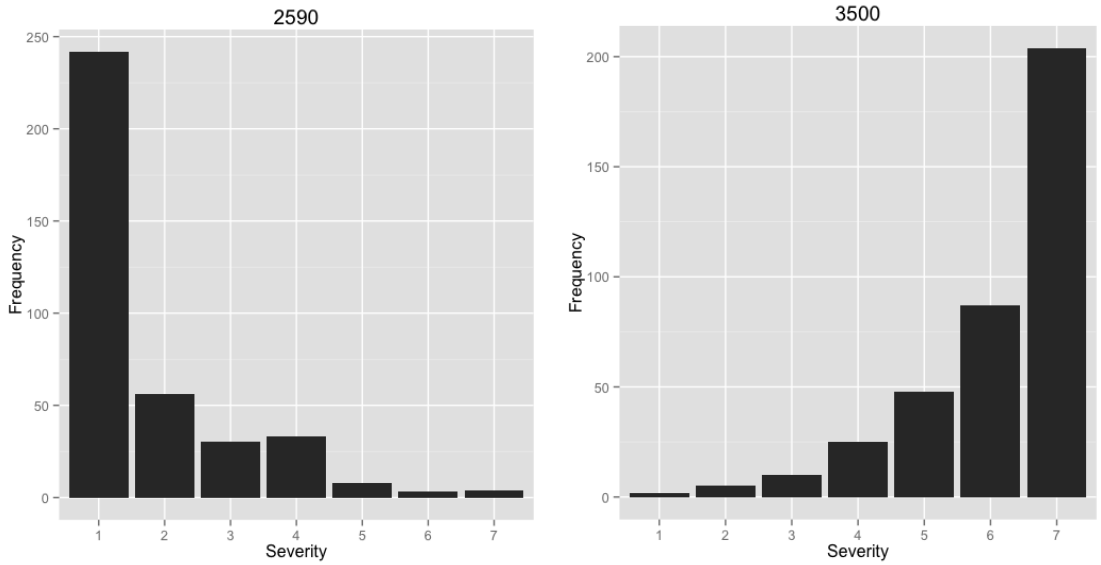
Predictor		*df*	*F*	*b*	*t*	*p*	*95% CI*	
							Upper	*Lower*
Crowd		2	1.18			.278		
	Anger			3.56	1.63	.105	-0.74	7.85
Certainty		1	41.31			< .001		
	Blur			6.02	2.70	.007	1.64	10.40
Gender		2	0.01			.957		
	Female			-0.06	-0.05	.957	-2.33	2.21
Race		1	5.48			.020		
	Non-white			2.59	2.34	.020	0.42	4.76
Image		6	564.17			< .001		
	2691			51.64	21.04	< .001	46.82	56.46
	3181			33.62	13.12	< .001	28.59	38.65
	3530			81.00	35.04	< .001	76.47	85.54
	4233			51.68	19.31	< .001	46.43	56.94
	9102			59.93	22.22	< .001	54.63	65.23
	9423			73.45	25.47	< .001	67.79	79.11
Crowd x Certainty		1	0.03			.873		
Crowd x Image		6	2.47			.023		
Image x Certainty		6	69.55			< .001		
Crowd x Certainty x Image		16	1.61			.142		

Table 11 (cont'd)

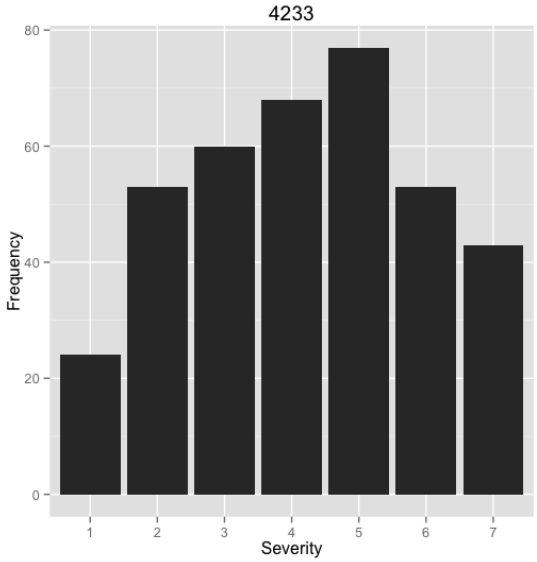
Note: Regression coefficients are unstandardized. Categorical variables dummy coded prior to analysis. Reference categories for Gender = Male, Race = White, Crowd = neutral, Image = 2026 (neutral), Certainty = clean image.

Figure 1:

Example severity rating distributions of a neutral (left), severe (right), and intermediate (bottom) image.



Neutral image example (left) and severe image example (right), demonstrating deviations from normality and significant floor and ceiling effects, respectively.



Intermediate image example demonstrating a relatively normal distribution with significant variance across the rating scale.

Figure 2:

Study 2 mean severity ratings, by image and facial expression condition.

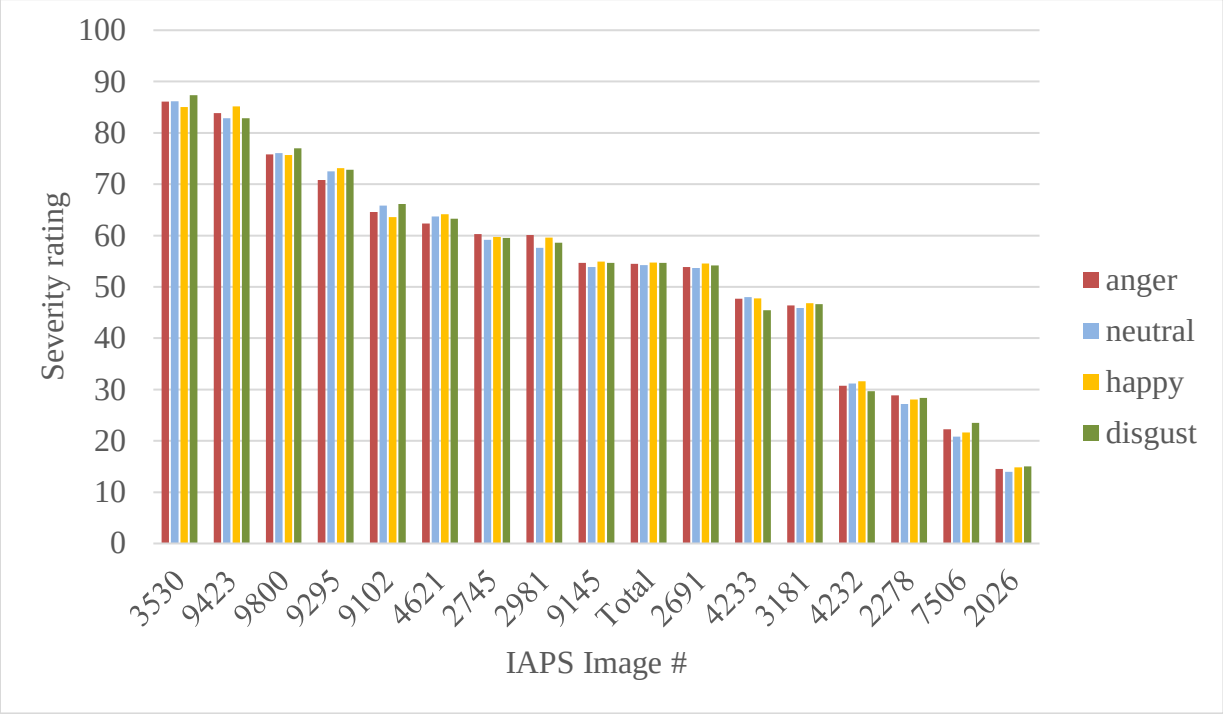
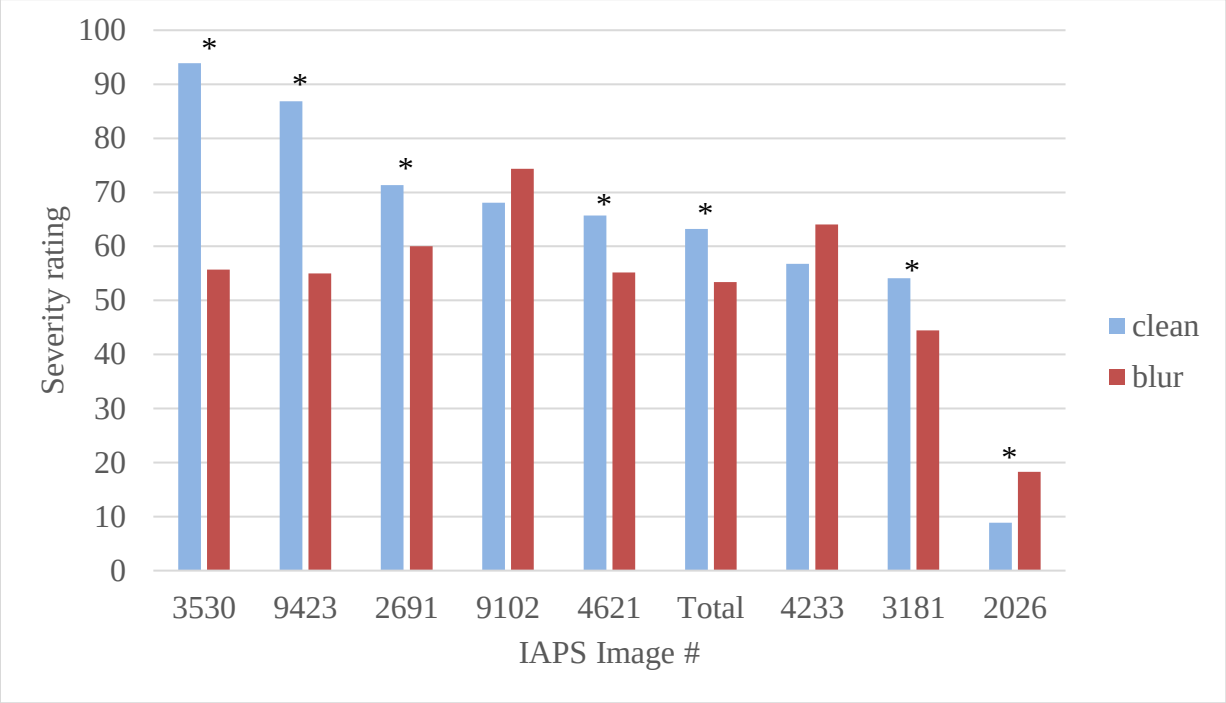


Figure 3:

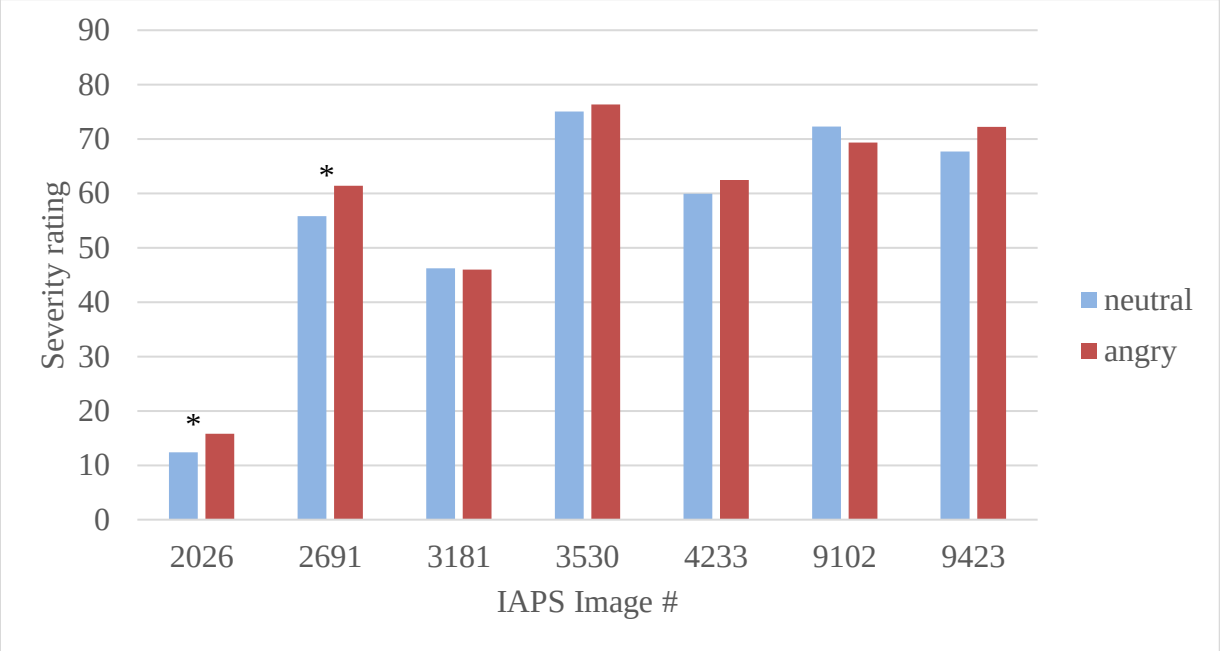
Study 3 mean severity ratings, by image and certainty (blur) condition.



Note: *$p < .05$

Figure 4:

Study 4 mean severity ratings, by image and crowd condition.



Note: *$p < .05$

REFERENCES

REFERENCES

Altemeyer, B. (1981). *Right-wing authoritarianism.* Manitoba, Canada: University Press.

Anderson, C., Monroy, M., & Keltner, D. (2017). Emotions in the wilds of nature: The coherence and contagion of deat during threatening group-based outdoors experiences. *Emotion, Advanced obnline publication*. doi:10.1037/emo0000378

Avramova, Y., & Inbar, Y. (2013). Emotion and moral judgment. *WIREs Cognitive Science, 4*(2), 169-178. doi:10.1002/wcs.1216

Bartels, D., Bauman, C., Cushman, F., Pizarro, D., & McGraw, A. (2014). Moral Judgment and Decision Making. In G. Keren & G. Wu (Eds.), *Blackwell Reader of Judgment and Decision Making*. Malden, MA: Blackwell.

Blasi, A. (1999). Emotions and Moral Behavior. *Journal for the Theory of Social Behavior, 29*(1), 1-19. doi:10.1111/1468-5914.00088

Brosnan, S., & de Waal, F. (2014). Evolution of responses to (un)fairness. *Science, 346*(6207), 1251776-1251776.

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods, 46*, 112-130.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition, 108*, 353-380. doi:10.1016/j.cognition.2008.03.006

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment. *Psychological Science, 17*(12), 1082-1089. doi:10.1111/j.1467-9280.2006.01834.x

DeScioli, P., & Kurzban, R. (2009). Mysteries of morality. *Cognition, 112*(2), 281-299. doi:10.1016/j.cognition.2009.05.008

Eibl-Eibesfeldt, I. (1989). *Human ethology*. New York, NY: Aldine de Gruyter.

Ekman, P., Friesen, W., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., . . . Tzavaras, A. (1987). Universals and cultural differences in judgments of facial expressions of emotion. *Journal of Personality and Social Psychology, 53*(4), 712-717.

Freeston, M.H., Rhéaume, J., Letarte, H., Dugas, M.J., & Ladouceur, R. (1994): Why do people worry? *Personality and Individual Differences, 17*(6), 791-802.

Garcia, J., & Koelling, R. (1966). Relation of cue to consequence in avoidance learning. *Psychological Science, 4*(1), 123-124. doi:10.3758/BF03342209

Graham, J., Nosek, B., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology, 101*(2), 366-386.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814-834. doi:10.1037//0033-295X. 108.4.814

Haidt, J. (2003). The moral emotions. In R. Davidson, K. Scherer, & H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 852-870). Oxford: Oxford University Press.

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York: Pantheon Books.

Haselton, M., & Buss, D. (2000). Error Management Theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology, 78*(1), 81-91. doi:10,1037110022-3514.78.1.81

Kanade, T., Cohn, J., & Tian, Y. (2000). Comprehensive database for facial expression analysis. *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00)*, Grenoble, France, 46-53.

Kolber, J., & Margol, B. (Writers) & M. Nigro (Director). (2015, Jan 26). [Television series episode] Morality, *Brain Games*. USA: National Geographic.

Krahé, B., Möller, I., Huesmann, L., Kirwil, L., Felber, J., & Berger, A. (2015). Desensitization to media violence: Links with habitual media violence exposure, aggressive cognitions, and aggressive behavior. *Journal of Personality and Social Psychology, 100*(4), 630-646. doi:10.1037/a0021711

Kruglanski, A. W., Atash, M. N., De Grada, E., Mannetti, L., & Pierro, A. (2013). Need for Closure Scale (NFC). *Measurement Instrument Database for the Social Science.* Retrieved from: www.midss.ie

Kurzban, R., & Leary, M. (2001). Evolutionary origins of stigmatization: The functions of social exclusion. *Psychological Bulletin, 127*(2), 187-208. doi:10.1037//0033-2909.127.2.187

Lang, P., Bradley, M., & Cuthbert, B. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8. University of Florida, Gainesville, FL.

Lucey, P., Cohn, J., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. *Proceedings of the Third International Workshop on CVPR for Human*

Communicative Behavior Analysis (CVPR4HB 2010), San Francisco, USA, 94-101. doi:10.1109/CVPRW.2010.5543262

McDonald, M.M., Defever, A.M., & Navarrete, C.D. (2017). Killing for the greater good: Action aversion and the emotional inhibition of harm in moral dilemmas. *Evolution and Human Behavior, 38*(7), 770-778.

Mikhail, J. (2007). Universal moral grammar: theory, evidence, and the future. *Trends in Cognitive Sciences, 11*(4), 143-152. doi:10.1016/j.tics.2006.12.007

Mrug, S., Maden, A., Cook, E., & Wright, R. (2015). Emotional and physiological desensitization to real-life and movie violence. *Journal of Youth and Adolescence, 44*(5), 1092-1108. doi:10.1007/s10964-014-0202-z

Oveis, C., Spectre, A., Smith, P., Liu, M., & Keltner, D. (2016). Laughter conveys status. *Journal of Experimental Social Psychology, 65*, 109-115. doi:10.1016/j.jesp.2016.04.005

Parr, L., & Waller, B. (2006). Understanding chimpanzee facial expression: insights into the evolution of communication. *Social Cognitive and Affective Neuroscience, 1*, 221-228. doi:10.1093/scan/nsl031

Prinz, J. (2006). The emotional basis of moral judgments. *Philosophical Explorations, 9*(1), 29-43. doi:10.1080/13869790500492466

Reiman, E., Fusselman, M., Fox, P., & Raichle, M. (1989). Neuroanatomical correlates of anticipatory anxiety. *Science, 243*(4894), 1071-1074.

Schur, M. (Writer) & D. Holland (Director). (2017, Oct 19). [Television series episode] The Trolley Problem, *The Good Place*. USA: NBC.

Schwarz, N., & Clore, G. (2007). Feelings and phenomenal experiences. In A. Kruglanski & E. Higgins (Eds.), *Handbook of basic principles* (2nd ed., pp. 385-407). New Yrok, NY: Guilford Press.

Shariff, A., & Tracy, J. (2011). What are emotion expressions for? *Current Directions in Psychological Science, 20*(6), 395-399. doi:10.1177/0963721411424739

Smillie, L., Pickering, A., & Jackson, C. (2006). The new reinforcement sensitivity theory: Implications for personality measurement. *Personality and Social Psychology Review, 10*(4), 320-335. doi:10.1207/s15327957pspr1004_3

Smithson, M. (2008). Psychology's ambivalent view of uncertainty. In G. Bammer & M. Smithson (Eds.), *Uncertainty and Risk: Multidisciplinary Perspectives* (pp. 205-217). London: Earthscan.

Tangney, J., Stuewig, J., & Mashek, D. (2007). Moral emotions and moral behavior. *Annual Review of Psychology, 58*(1), 345-372. doi:10.1146/annurev.psych.56.091103.070145

Thomson, J. (1985). The Trolley Problem. *The Yale Law Journal, 94*(6), 1395-1415.

Young, A., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition, 120*(2), 202-214. doi:10.1016/j.cognition.2011.04.005