SOCIAL PSYCHOLOGICAL PROCESS AND EFFECTS ON THE LAW

EDITED BY: Colleen M. Berryessa, Yael Granot, Melissa de Vel-Palumbo and Clare Sarah Allely

frontiers Research Topics

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

SOCIAL PSYCHOLOGICAL PROCESS AND EFFECTS ON THE LAW

Table of Contents

Editorial: Social psychological process and effects on the law

Colleen M. Berryessa[1]*, Clare S. Allely[2],
Melissa de Vel-Palumbo[3] and Yael Granot[4]

[1]School of Criminal Justice, Rutgers University, Newark, NJ, United States, [2]School of Health and Society, University of Salford, Manchester, United Kingdom, [3]College of Business, Government and Law, Flinders University, Adelaide, SA, Australia, [4]Department of Psychology, Smith College, Northampton, MA, United States

Editorial on the Research Topic
Social psychological process and effects on the law

Studying social psychological processes entails disentangling how people perceive, interact in, and react to our social world. This framework has been increasingly applied to studying law, with growing interest in the ways in which social contexts intersect with legal institutions and decision-making. Although the law may be viewed as insulated from social contexts, it is in fact ever-changing and shaped over time by society. Work at this intersection offers insights into how social psychology can impact the law, but also informs the law about the ways in which the public engages with and perceives legal principles, practices, and proceedings.

Although research at this intersection has begun to grow in recent years, many areas of empirical and theoretical work in this area continue to be understudied, particularly across different countries and legal systems, those using more interdisciplinary frameworks in the study of these relationships, and in considering broader understandings and applications of social psychological processes to studying the law.

The 13 papers included in this Research Topic approach varied aspects of social cognition and its relationship to legal processes, providing important guidance on how we might explore these questions in future international work across different jurisdictions and countries. We are thrilled that researchers represented in this international collection hail from Australia, Canada, Chile, China, Hong Kong, Italy, Japan, the Netherlands, Spain, Sweden, and the United States.

Grosfeld et al. survey members of the European Union (EU), finding that value alignment, particularly in relation to binding values, plays a significant role in affecting the public's views on the perceived legitimacy of the Court of Justice of the EU (CJEU) and the EU more broadly.

Younan and Martire present two experimental studies to U.S. participants on the effects of expert likeability. They find that likeability may influence judgments on experts' persuasiveness and testimony quality, but may not necessarily affect support for particular sentencing outcomes.

Kurinec and Weaver also use two online experiments to show that the speech stereotypicality of Black Americans may activate racial stereotypes and racial phenotype bias, which influence suspect descriptions and eyewitness identifications.

Albrecht and Nadler test how the composition of crime news articles contributes to reader perceptions of moral blameworthiness and corresponding punishment attributions of vehicular homicide offenders. This study suggests that lay support for more severe punishment is affected by participants' characteristics, particularly political affiliation, when the immigration status of a suspect is provided and uncovers how differential reporting on suspects' personal characteristics may affect public views on blameworthiness.

Shang et al. present a study that measures types of social behavior, such as empathy, perspective-taking, and self-control, in Chinese adolescents in order to further work on whether age should continue to be the primary attribute by which to judge a juvenile's criminal responsibility. Their findings support the notion that legal systems may want to consider juvenile responsibility in terms of the social and interpersonal maturity and decision-making, rather than solely in terms of age.

Pettersson et al. study Swedish police officers and their ability to conduct investigative interviews with intoxicated witnesses, primarily looking at how police decision-making and perceptions of witness credibility may be biased by pre-existing social norms. This study adds to the existing literature in the field by showing that breath alcohol concentration far lower than the legal maximum still significantly affected officers' views on witness credibility.

Watamura et al. from Japan, develop a ratio measure to see how people weigh and justify different punishment philosophies when considering sentences for child abuse cases. Results show that ratio justifications differ across cases involving either severe or moderate abuse, with both retribution and utilitarian justification considered in the sentencing decisions of such cases.

Ansems et al. study Dutch criminal court hearings involving defendants with non-Western backgrounds to examine how prior discrimination and outcome judgments might interfere with the effects of procedural justice. Their findings help to illuminate the importance of promoting procedural justice in Dutch courts as a way to decrease social costs associated with continued justice system involvement.

Guan and Lo present a systematic review on drug offending within a certainty–severity framework of punishment, covering a wide body of literature on the importance of exposing certain types of information on punishment as a way to deter drug offending. Main themes identified in this literature focus on restrictive deterrence strategies, particularly surrounding pre-arrest, deterrability, and perceptions of risk, which suggest expanding future work on after-arrest strategies and across different types of drug offenders.

Ewanation and Maeder use an online experiment of U.S. participants to study the effects of a defendant's race and the presence or absence of expert testimony on jurors' perceptions of recanted confessions. Results support a "watchdog hypothesis" as White mock-jurors were found to be more receptive to legally relevant evidence when a defendant was identified as Black.

Angioletti et al. using common moral dilemmas from psychology research, demonstrate how individual, situational, contextual, and internal factors may influence the moral decision-making of lawyers in Italy. Results show that lawyers' internal states (e.g., interoceptive ability) may influence their fairness in decision-making during trial.

Saad et al. in an experimental study of New Jersey parole officers, find that officers' implicit social cognition may influence their behaviors toward and empathy for those whom they supervise. Findings may help to improve therapeutic and supervision relationships between officers and their clients.

Camplá et al. assess informal reasoning and biases that may affect the decision-making of Chilean legal actors in rape cases. Results find that these actors commonly overestimate probabilities of false or unfounded allegations and myths about sexual offending, and show attributional biases toward victims.

Ultimately, this paper collection represents an expansive and comprehensive account of international research on widespread ways in which law and social psychology interact. These issues are not only important to common legal practices, such as eyewitness identification, interviewing, or trial proceedings, and how they may be influenced by discrimination, bias, and other social processes, but also when considering how social psychological processes could influence larger philosophical questions on why we punish, why we use and support various legal practices, and the design and evaluation of legal rules. Thus, interactive relationships between law and social psychology should be viewed as "two-way streets" that will continue to shape criminal-legal outcomes across the globe moving forward.

Author contributions

Acknowledgments

We would like to thank the authors for their contributions, as well as Frontiers for providing the opportunity to solicit and publish these papers.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Restrictive Deterrence in Drug Offenses: A Systematic Review and Meta-Synthesis of Mixed Studies

Xin Guan and T. Wing Lo*

Department of Social and Behavioral Sciences, City University of Hong Kong, Hong Kong, China

Deterrence by punishment aims to prevent a crime; however, it is not always successful. Restrictive deterrence explains the continuous criminal activities that occur despite deterrence; offenders enact various strategies to avoid detection, which is more typical among drug offenders given that they have a high frequency of offending and exposure to punishment. This systematic review provides an in-depth understanding of restrictive deterrence of drug offenders. Two prominent themes, "restrictive deterrence strategy" and "deterrability and restrictive deterrence," depict drug offenders' restrictive deterrence and effectively fit within the certainty–severity framework of punishment. Future studies should investigate restrictive deterrence strategies in the after-arrest context, the facilitative effect of perception of risk on strategy development, and facilitators or inhibitors affecting the diffusion of restrictive deterrence strategies.

INTRODUCTION

For decades, researchers, and theorists in criminology have investigated punishment and its deterrent effect. A large portion of deterrence research has focused on how punishment exerts an influence on people's determination to engage in or refrain from illegal behavior. Two elements of punishment, certainty, and severity, are the most commonly cited and explored, and are considered to be influential factors in motivating people to avoid committing crimes. Although punishment aims to deter people from crime altogether (absolute deterrence), it has a chance of encouraging people to commit crimes in insidious ways, which echoes the concept of restrictive deterrence (partial deterrence) (Gibbs, 1975, p. 33).

The distinction between absolute and restrictive deterrence is the extent to which people adjust their criminal behavior in reaction to risks. As Gibbs (1975, p. 32) defined, absolute deterrence denotes "an individual has refrained throughout life from a particular type of criminal act because in whole or in part he or she perceived some risk of someone suffering a punishment as a response to the crime." Restrictive deterrence denotes "the curtailment of a certain type of criminal activity by an individual during some period because in whole or in part the curtailment is perceived by the individual as reducing the risk that someone will be punished as a response to the activity" (Gibbs, 1975, p. 33). It can be derived from two definitions that some persons may stop committing crimes to lessen their likelihood of punishment, while others may only curtail the frequency of crime. Beyond the magnitude of behavioral change, the two kinds of deterrence apply to different types of offenders. Absolute deterrence pertains to the people who refrain from participating in crime from a time onwards, regardless of their previous crime involvement. However, restrictive deterrence is only applicable to those who have committed a particular crime at least once.

Jacobs (1996a) expanded Gibbs's definition of restrictive deterrence by classifying it into two distinct types: probabilistic and particularistic restrictive deterrence. The former corresponds to the definition proposed by Gibbs (1975, p. 33), which emphasizes the reduction of crime frequency. The latter refers to the "skills for evasion" (Jacobs, 1996a, p. 425), implying that offenders develop various situational measures, namely restrictive deterrence strategies, to carry out an offense more likely to go undetected. For example, an offender committing street crimes takes advantage of everyday social activity to disguise the act of committing a crime (e.g., shaking hands with another using complex street handshake etiquette while holding an illegal substance in his hands and exchanging it with his partner). An offender commits offenses of lesser severity than the one anticipated because he believes that there will be less penalty for a less serious crime (e.g., an offender only sells cannabis rather than heroin). Both of these are typical restrictive deterrence strategies among drug offenders.

Drug offenders, referring to those who use, deal/traffic, or cultivate/manufacture illegal drugs, are of particular relevance in the theoretical development of restrictive deterrence. Since Gibbs (1975, p. 33) introduced the concept of restrictive deterrence, it has been substantively explored on samples of drug offenders (Jacobs, 1993, 1996a,b). Originally based on research concerning drug offenders, restrictive deterrence was also gradually extended to a broader range of criminals, such as auto thieves, sex workers, sexual offenders, and computer hackers, among others (Jacobs and Miller, 1998; Cherbonneau and Copes, 2006; Beauregard and Bouchard, 2010; Gallupe et al., 2011; Jacobs and Cherbonneau, 2014; Maimon et al., 2014; Wilson et al., 2015).

Likewise, restrictive deterrence is of particular relevance in shaping the character of drug offenders. First, the high recidivism levels of drug offenders (Harrison, 2001) and their involvement in multiple crimes (Casey, 2015) suggest that they are among the most judicially entrenched offenders. The high recidivism rate of drug offenders may be partly because restrictive deterrence strategies facilitate them to avoid arrest and thus build a criminal career. Second, the restrictive deterrent effect is more potent for drug offenders than liquor drinkers, petty thieves, or vandalizers (Paternoster, 1989; Eck and Wartell, 1998). The high risk-responsiveness of drug offenders may be partly because restrictive deterrence contributes to converting drug offenders' risk perception into action against risk rather than just quitting from crime or ignoring the risk.

Drug offenders evolve strategies to counteract the threats of punishment, and punishment threats are developed in return to discourage offenders more efficiently; such progress repeatedly continues and becomes an inevitable cycle (Ryan, 1994). Drug offenders have shown their adaptiveness to cope and innovate ways to commit crimes. Consequently, practitioners of criminal justice and scholars need to thoroughly grasp restrictive deterrence to better respond to newly-developed patterns in drug offender behavior.

The bulk of this systematic review examined restrictive deterrence of drug offenders, including the concrete strategy and possible prerequisites for strategy use. The current systematic review synthesizes 34 quantitative, qualitative, and mixed-method studies that focused on restrictive deterrence of drug offenders, and analyses findings based on a certainty–severity framework of punishment, providing an explicit picture and revealing the understudied field of restrictive deterrence of drug offenders.

To complete the synthesis and interpretation, we introduce a certainty–severity framework of punishment (**Figure 2**). Punishment is a system of conditional probabilities (Nagin, 2013), including multiple probabilistic events between arrest and final sentencing. As the sequence of probabilistic events evolves, the certainty and severity of the punishment also undergoes an increase. The certainty and severity of punishment serve as deterrents (Piliavin et al., 1986; Williams and Hawkins, 1989), where the certainty of punishment is viewed as the most influential (DeJong, 1997; Pogarsky, 2002; Nagin and Pogarsky, 2003) and the severity of punishment only produces a modest effect (Pogarsky and Piquero, 2003; Apel, 2013). The framework based on these two elements of punishment assists in the better visualizing of restrictive deterrence of drug offenders.

In the certainty–severity framework of punishment, the x-axis represents the certainty of punishment and the y-axis represents the severity of punishment. Along the x-axis and y-axis, different restrictive deterrence strategies are presented, and in the middle of this coordinate system factors that influence the strategy implementation are listed. The clustered themes scatter along axes and the coordinate system, explaining how drug offenders implement strategies to "move" to the origin place (0,0) representing successfully avoiding detection, and exploring the potential prerequisite for strategy implementation.

MATERIALS AND METHODS

Meta-Synthesis of Mixed Studies

In the present study, meta-synthesis of mixed studies is adopted. Meta-synthesis is an analytical technique used to combine and compare the outcomes or metaphors of various qualitative studies to create interpretations, ground narratives, or theories (Sandelowski et al., 1997; Beck, 2002). Meta-synthesis expands the qualitative results by analyzing the distinctiveness of a study as a through and interpretive whole as opposed to a meta-analysis, which transforms quantitative research into averages (Clemmens, 2003). Though frequently focused on qualitative research, it can also be used to integrate qualitative, quantitative, and mixed-method studies to provide a more holistic view of the problem than could be obtained from a one study approach (Panda et al., 2018). A meta-synthesis is still not commonly used and is a relatively new method in criminology. Nevertheless, it is a worthy instrument to promote gap-finding. Therefore, the meta-synthesis of mixed studies might be an effective way to achieve a thorough analysis of restrictive deterrence of drug offenders.

Procedure of Meta-Synthesis of Mixed Studies

The review consists of four successive phases: data selection, data extraction, theme identification, and finding synthesis.

Data Selection

The review complies with the guideline of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Moher et al., 2009) (**Figure 1**).

The HeinOnline, Social Science Database, Sociological Abstracts, Scopus, SAGE, JSTOR, PsycINFO, and Web of Science research databases were searched in December 2020. Furthermore, the SPIDER (Sample, Phenomenon of interest, Design, Evaluation, Research type) approach (Cooke et al., 2012) was used to decompose the targets and reinforce the search strategy (see **Appendix 1**).

Data screening was proceeded in three steps. The first step was de-duplication. The second step was screening with titles. Researchers followed the idea of the SPIDER approach, leaving those studies with titles containing terms such as "drug" and its derivatives, "deterrence" and its cognates or "avoid" and its cognates. The third step was screening with abstracts. As the drug-related studies have a broad perspective, the researcher only selected studies with their abstract indicating how drug offenders commit crimes or the factors influencing the ways they commit crimes, and excluded those studies that focused on the subjects of law enforcement, victims of drug crime or other subjects involved in drug crime.

After the data screening, the Mixed Methods Appraisal Tool (MMAT) version 2018 was used to evaluate the methodological quality of the included studies. The MMAT contains 27 methodological quality criteria for appraising qualitative, quantitative and mixed-methods studies. Each criterion in the tool would be labeled with one asterisk if a requirement was met and would be labeled without asterisk if a requirement was not met or "cannot tell." There are two comprehensive screening questions for all types of studies, namely "Are there clear research questions?" and "Do the collected data allow to address the research questions?" A further appraisal may not be feasible when the answer is "no" or "cannot tell" to one or both screening questions. Of the remaining 25 questions, 5 of them are expressly set up for appraising the qualitative study, 15 of them for quantitative study (as the tool divides quantitative study into three types, including quantitative randomized controlled trials, quantitative non-randomized and quantitative descriptive), and 5 of them for mixed-methods study. Accordingly, for each study, it will be labeled with 5 asterisks, or scored 100%, if it meets all criteria for the type of study. By analogy, a study that meets 4 criteria will be labeled with 4 asterisks (or scored 80%). The MMAT does not have a specific standard cut-off value. However, two categories (low and high) or three categories (low, medium and high) can be adopted. The crucial aspect is to carefully utilize the results of the appraisal in the review. One author of the present study analyzed the methodological quality of each included study and verified its final score. Studies with at least 3 asterisks (or scored 60%) were kept.

Data Extraction

A data extraction table was developed to report a full description of each included study, including its purpose, sample size, design,

FIGURE 1 | PRISMA flow chart presenting search results.
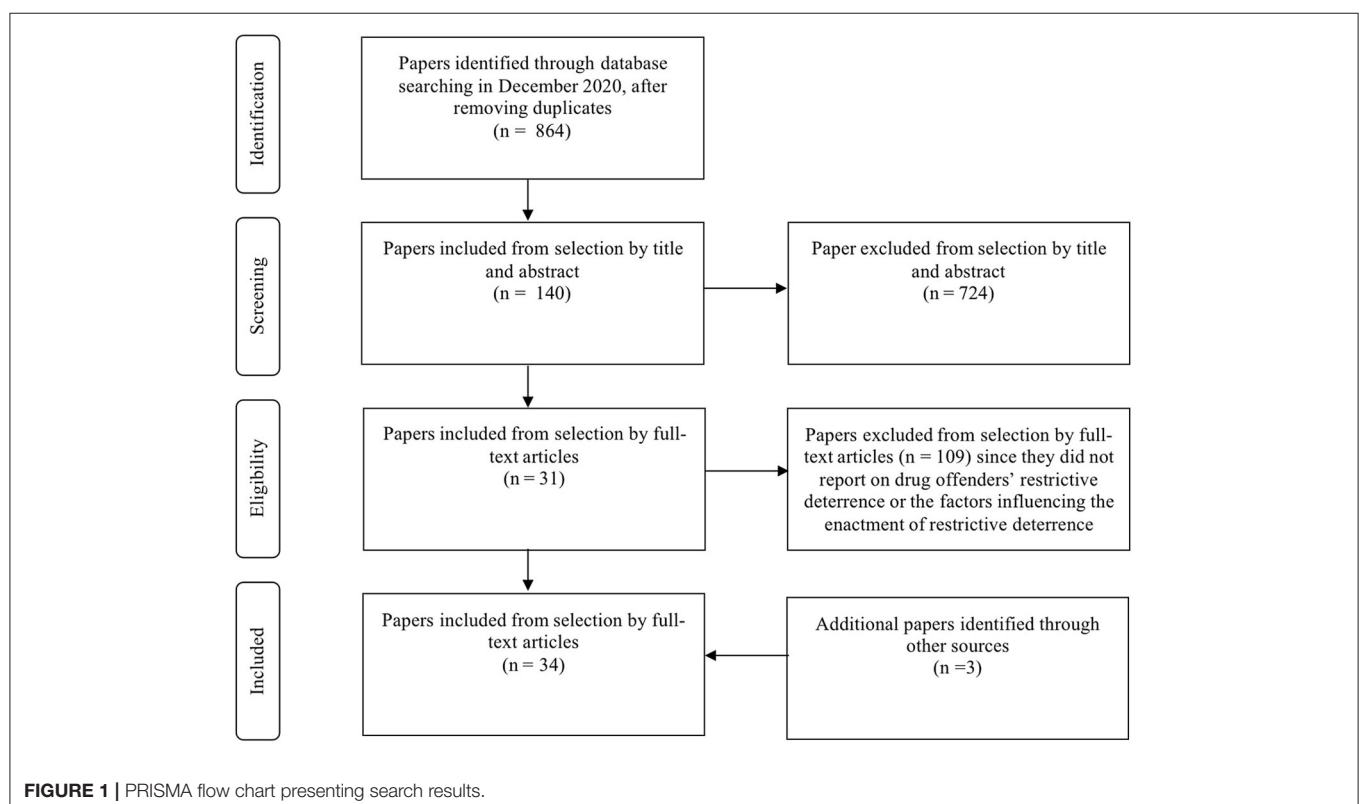
TABLE 1 | Results of search strategy for each database.

Database	Year	Results
Heinonline	1975–2020	618
Social science database	1975–2020	172
Sociological abstracts	1975–2020	160
Scopus	1975–2020	159
SAGE	1975–2020	78
JSTOR	1975–2020	24
PsycINFO	1975–2020	14
Web of science	1975–2020	12
Total		**1,237**

participants, research setting, and the drug crime type reported by the author.

Theme Identification

Themes were extracted and grouped from individual studies into wider themes and subthemes before being synthesized. A thematic analysis and meta-synthesis were performed rather than a meta-analysis since meta-analyses are not feasible when there is considerable heterogeneity among qualitative studies.

Synthesis Identification

Based on the search strategy, 1,237 individual titles were retrieved and 864 studies remained after removing any identified duplicates ($n = 373$) (**Table 1**). A total of 724 studies were excluded after reviewing their "titles" and "abstracts." Hence, 140 studies remained for a full-text review by the author, and 109 of these were excluded since they failed to report drug offenders' restrictive deterrence or the factors influencing restrictive deterrence. Meanwhile, an additional three papers identified through other sources were added (**Figure 1**). Finally, 34 included papers were reviewed for quality appraisal. All of these papers scored as moderate (score of 60 to 80%) ($n = 10$) or high quality (score of 100%) ($n = 24$).

The 34 studies were published from 1984 to 2019: 18 studies focused mainly on factors that affect the action of restrictive deterrence; 15 of them depicted restrictive deterrence strategies; and only one literature review discussed restrictive deterrence with respect to multiple crimes. Studies were carried out mostly in the United States and other Western countries. Nine studies used a quantitative design (surveys/questionnaires/systematic observations), whilst 22 used qualitative designs (individual or focus group interviews), and three were designed using mixed methods (interviews and surveys/systematic observation).

Based on the systematic review and meta-synthesis of mixed studies, the characteristics of the included studies were summarized in **Appendix 2** and the most noticeable themes and subthemes in **Table 2**. Specifically, three main areas of concern were identified: restrictive deterrence strategies, the contingency of restrictive deterrence, and the iteration of restrictive deterrence. These themes fit within the certainty–severity framework of punishment (**Figure 2**). Our examination

TABLE 2 | Restrictive deterrence domains and strategy used by drug offenders.

Theme	Subtheme	Detail
Restrictive deterrence strategy	1. Certainty reduction	a. Camouflage in front stage (Jacobs, 1996b; Jacobs and Miller, 1998)
		b. Pick safe time and position (Jacobs and Miller, 1998; VanNostrand and Tewksbury, 1999)
		c. Counter-reconnaissance (Jacobs, 1993, 1996a; Jacques and Reynald, 2012)
	2. Severity mitigation	a. Choose less severe activity (Knowles, 1999; Fleetwood, 2014)
		b. Pass risk (Knowles, 1999; Piza and Sytsma, 2016)
		c. Stash product (Jacobs, 1996b; Jacobs and Miller, 1998; Moloney et al., 2015)
		d. Cooperate with police (Cross, 2000; Dickinson and Wright, 2015)
Deterrability and restrictive deterrence	1. Perception of risk	a. Individual characteristic
		b. Crime milieu characteristic
	2. Crime skill	a. Self-reflection
		b. Collective wisdom
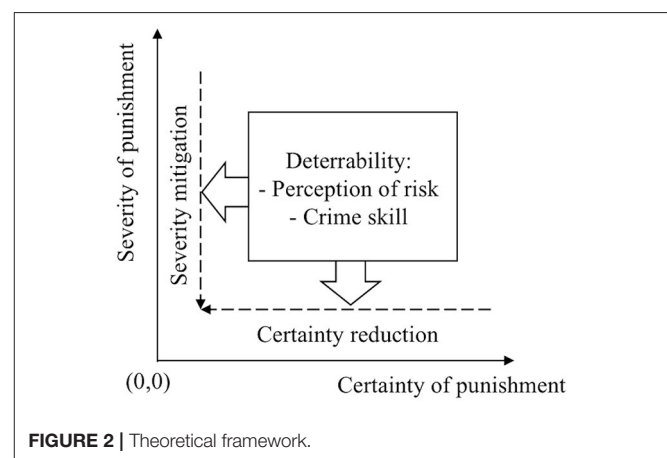
FIGURE 2 | Theoretical framework.

of the constitution of the differences within the themes is reflected along these axes. In the subsequent section, a series of inferences and generalizations about restrictive deterrence of drug offenders and any uncharted areas are explored.

RESULTS

Restrictive Deterrence Strategy

Offenders are risk-respondents rather than risk-takers (Jacobs and Cherbonneau, 2014). In other words, offenders adopt various strategies to alter the risk environment in which they are placed. Punishment is one of the most significant risks associated with

offenders. It can mainly be divided into two types. One is formal punishment (legal punishment), such as arrest/apprehension. Another is informal punishment (e.g., moral sanction) (Jacques and Allen, 2014), such as stigmatization/labeling. Informal punishment, at some level, can be seen as a subsequent punishment triggered by formal punishment (Nagin, 1998). Therefore, all allusions to punishment in the following indicate formal punishment.

Many terms such as detection, arrest, apprehension, conviction, prosecution, and sanction have been used to refer to punishment or a part of punishment. There is an obvious need for clear and coherent definitions of punishment and/or its associations in light of the expansion of literature. As Nagin (2013) noted, punishment is more accurately characterized as a system of conditional probabilities. There are multiple probabilistic events between detection and final sentencing, such as from arrest to detention, to prosecution, to conviction, and to sentencing. In this sequence of conditional probabilities, punishment is shown as a process of severity in legal or judicial responses.

We extended this concept of punishment to the restrictive deterrence strategies of drug offenders. Drug offenders adopt various strategies to minimize their odds of arrest. These strategies help avoid one kind of adverse event. They also impact the probability of the occurrence of the subsequent event and affect the final sentencing. Hence, in what follows, the restrictive deterrence strategy used by drug offenders, whether to avoid arrest or to reduce the length of a sentence, can be understood as an attempt to avoid the punishment.

The commonly used classification of restrictive deterrence strategy is contended by Jacobs (1996b), including probabilistic strategies and particularistic strategies. However, the current study adopts Moeller et al.'s (2016) classification of restrictive deterrence strategies because it is in line with the certainty–severity framework of punishment. Hence, restrictive deterrence strategies are divided into certainty reduction strategies and severity mitigation strategies. The former corresponds to criminal thinking about "what should I do to commit a crime while keeping myself from arrest," and the latter corresponds to "what should I prepare to do if I am arrested?"

Certainty Reduction: "I Need a Plan"

Strategies of certainty reduction are designed to allow drug offenders to remain "invisible" to police when a crime occurs. Drug offenders disguise themselves under the cloak of legal activities, keeping a low-profile, choosing a less-risky time and area, and discreetly uncovering their adversary's invasion. In such a way, they create an illusion of being a law-abiding person.

Camouflage

Drug offenders camouflage their drug offending in two ways: integrating crimes into existing legal daily routines or producing staged performances to disguise offenses and allow them to evade police attention continually. Jacobs and Miller (1998) noted that drug dealers make good use of gender advantages to blend into their environments. Female drug dealers take their children to transactions and project a self-image that does not

use dramatic clothes and accessories to reduce police suspicion (Jacobs and Miller, 1998; Carbone-Lopez, 2015; Moloney et al., 2015). Some drug dealers engage in legal occupations or a legitimate business because these can generate unpredictable street activity routines which can reduce law enforcement (VanNostrand and Tewksbury, 1999; Fader, 2016b, 2019). Some people who sell drugs need to cooperate with others. They have to set up a flawless "stage performance" by preparing a set of props, feasible locations, helpmates, and even specialized words, the so-called "transactional mediation" contended by Jacobs (1996b). This requires that buyers be regular customers who commonly understand "the same kind of action" (Schutz, 1972, p. 155). Jacobs (1996b) identified three ways of using transaction mediation: flash decoys, moving screens, and sleight of hand. Flash decoys refers to finishing the drug deal in automobiles, camouflaging the whole process as a kind of favor to hitchhikers. Moving screens refers to rehearsing the movements between participants to speed up the transaction and smoothness of dealing. Sleight of hand refers to using normal hand gestures, including slaps and hugs, to finish the final step of drug delivery.

Picking a Safe Time and Position

Deciding when and where to commit the crime is also a significant part of certainty reduction. Drug offenders only commit crimes in locations they consider safe (Carbone-Lopez, 2015; Olaghere and Lum, 2018). A transaction can be arranged at local entertainment facilities, such as restaurants or markets where participants can eat and shop, appearing no different from others to evade intense surveillance (Jacobs and Miller, 1998). For others, selling from home is an ideal option. This decreases the sense of insecurity from being on the street (Moloney et al., 2015). Female drug dealers usually invite friends for after-parties in which drug dealing has been established (Fleetwood, 2014). Along with trading in public areas, some drug trade occurs in secluded lots in which drug dealers can easily perceive risk. Drug dealers guide buyers to walk into a covert space in an apartment while partners watch every move of the buyers if things "go down" (Jacobs, 1996b). Drug cultivators who operate in less visible areas (e.g., stores, hotels) are also more likely to avoid detection (Gallupe et al., 2011). By manipulating their trading location, dealers can squeeze in additional moments to "launder" the illegal income (Jacobs, 1996b). In addition to picking a safe place, timing matters. It is easy to be exposed if selling occurs at an unusual time in the day. Female drug dealers set up an unbreakable principle about opening hours and do not respond to any demands outside the pre-set time (Jacobs and Miller, 1998). Sometimes dealers might suddenly change the location. Drug offenders select areas and times based on police patrol intensity, often diverting to other areas if police patrol is increased in a given region (VanNostrand and Tewksbury, 1999).

Counter-Reconnaissance

Drug dealers generally deal with regular customers; however, it is almost inevitable that many of them have connections with unfamiliar, new buyers. Undercover police usually utilize such trading opportunities and operate buy–bust approaches. Thus, offenders develop strategies to identify the presence of

police without getting arrested. First of all, they avoid dealing with immature people who could be under intensive police surveillance. Maturity can be judged by whether an individual engages in overly risky behavior, such as an overdose of drugs, or purely based on age stereotypes (Jacques and Allen, 2014). They can also tell police from the real drug customers by their image, in addition to verbal and physical clues, or perhaps test potential buyers in a number of different ways (Jacobs, 1993, 1996a; Johnson and Natarajan, 1995; VanNostrand and Tewksbury, 1999; Jacques and Reynald, 2012). They scan the counterparts repeatedly and collectively to ensure their real identity. To decide if something seems unusual or potentially dangerous, dealers must identify their turf, practicing what they call a "peep game" (Jacobs, 1996b), such as using a foreign language to separate real drug buyers from potential undercover police (Knowles, 1999). Once they identify an undercover police officer, drug dealers withhold exchanges (Jacques and Allen, 2014).

Severity Mitigation: "If I Were Arrested"

Compared to fruitful observations about certainty reduction strategies, severity mitigation is understudied (Moeller et al., 2016). One possibility is that severe punishments without certain odds of arrest may have little effect on individual behavior (Carbone-Lopez, 2015). However, many strategies of drug offenders are to prepare for future arrest. Frequently used strategies for severity mitigation include engaging in less severe activity, passing the risk on to associates, stashing stock in converted places, and cooperating with police. In doing so, they believe the probability of a guilty conviction or the severity of their punishment could be decreased if they were prosecuted.

Choosing a Less Severe Activity

Choosing a less severe activity is the easiest way to mitigate severity. Some drug runners do not become a dealer because of the fear of hard prison time (Knowles, 1999). Many drug dealers only sell less toxic drugs, such as marijuana rather than crack, because selling crack is regarded as a more severe crime (Fleetwood, 2014; Moloney et al., 2015). Drug cultivators who worry about manufacturing charges perceive cooking and purchasing as precursors that could bring about more severe penalties (Carbone-Lopez, 2015). The drug producers who face manufacturing charges also have several strategies for getting around legislative restrictions to mitigate charges with more severity, such as replacing purchasing precursors with production (Vidal and Décary-Hétu, 2018).

Passing on Risks

The passing on of risks is a method of diverting dangers to lower power gang members. The use of selling partners can be a defense against severe charges (Piza and Sytsma, 2016). There are multiple roles in a drug dealing group, with some members charged with the duty to check and receive money and others merely being responsible for drug delivery (Johnson and Natarajan, 1995; VanNostrand and Tewksbury, 1999). In contrast, the "big boss," who is the actual owner of both the drugs and money, never shows up in the police observation; thus, they could reduce criminal culpability. Lower-level distributors are often hired for the riskiest work (Johnson and Natarajan, 1995; Jacobs and Miller, 1998). Drug runners enable true dealers to be shielded from potential police surveillance or detention, which helps them fully escape the criminal justice system (Knowles, 1999).

Stashing Products

Unlike high-level drug dealers who have many helpmates, street drug dealers have to hide drugs in safe locations by themselves to minimize the potential accusation of drug trafficking, which is more severe than drug possession. A frequently used tactic is to hide the majority of their stock and only keep a small number of drugs to be sold quickly (Johnson and Natarajan, 1995). They usually hide the majority of the drugs in caps, under bottles, in newspaper stands, on the ground, or in paper bags that have been placed at a particular angle (Jacobs, 1996b). Women innovate the concealment in their homes, such as a stash inside the hollow shaft of a curtain rod or a box under the carpet over which the pet dog sleeps (Jacobs and Miller, 1998). When carrying the drug, they have to practice drug-handling techniques to avoid scrutiny when they encounter police. Due to the legal constraints stating that police cannot ask suspects to strip, this leaves room to hide drugs within clothes. Drugs are tightly packaged in plastic wrap which can be placed in the hand and mouth without notice or can be swallowed if risk is perceived (Jacobs, 1996b). In drug dealing, women's bodies are viewed as an advantage since they possess "more hiding spots" (Moloney et al., 2015). An on-person or off-person stash is also dependent on settings. In commercial areas with increased foot traffic, an on-person stash is deemed safer than an off-person one (Piza and Sytsma, 2016).

Cooperating With Police

To cooperate with the police is to admit drug use if approached by the police (Ribeiro et al., 2010). Fooling the police might lead to violent conflict and even a more severe sentence. Passive cooperation with police means that offenders are "turned" by the police to seek the possibility of a less severe punishment. Such a strategy could be inferred from certain studies. Some sellers have emphasized that they should be careful when dealing with dealers that have recently been charged with a large number of drugs and released soon after because they are more likely to be a decoy under the instruction of the police or an informant (Dickinson and Wright, 2015). Further, the informant may only be charged and convicted as a low-level drug employee (Cross, 2000). In Johnson and Natarajan (1995), a high-level drug dealer recalled that his first jail experience was due to being set up by a drug user.

Deterrability and Restrictive Deterrence

The aim of restrictive deterrence strategies is to reduce the risk of punishment, or reduce the offenders' perceived risk. This implicitly presupposes that offenders have the ability to perceive and calculate risk. Jacobs (2010) used deterrability to highlight such an ability, explicitly referring to the "offender's capacity and/or willingness to perform risk calculation." Deterrability is crucial in understanding restrictive deterrence strategies because it supports the idea that such strategies are not arbitrary and

thoughtless. Instead, it can be seen as the prerequisite of drug offenders' use of restrictive deterrence strategies.

Jacobs (2010) suggested that deterrability should be measured by risk sensitivity. The current systematic review follows this line of thought and divides drug offenders' deterrability into two parts related to risk sensitivity in criminology, including the perception of risk (Roche et al., 2020) and crime skill (Casey, 2015).

Perception of Risk

A growing body of research has highlighted the importance of risk perception in the decision making of offenders (Cherbonneau and Copes, 2006; Beauregard and Bouchard, 2010; Gallupe et al., 2011; Jacobs and Cherbonneau, 2014, 2016, 2017; Maimon et al., 2014; Wilson et al., 2015; Moeller et al., 2016). Perception of risk (subjective risk of punishment) is an extension of actual risk (objective risk of punishment). First, perception of risk is a vital gateway to connect risk and subsequent behavior (Decker et al., 1993; Pogarsky et al., 2004; Paternoster, 2010). Researchers often explain criminal behavior and the vast majority of human behavior by assuming a reality-perception correspondence. Restrictive deterrence does not require a perfect correspondence between the real and the perceived risk. Still, some correspondence (a net positive effect) is necessary if it is to serve as an explanation or a predictor. Second, perception of risk is the individualization of the actual risk for a group of offenders. An example might clarify this. A drug offender lives in a city where 5% of drug offenders are punished each year. This rate of punishment is the average actual risk for both him and the group. However, he is neither a drug lord nor a drug addict who feeds on trafficking but a regular company employee who traffics drugs for subsidizing the household. He is not a gang member and traffics only small amounts of cannabis rather than cocaine. These factors may have reduced his perceived risk of being punished from the average actual risk, although the variation is hard to measure. Studies indicated a moderate or weak relationship between perceived and actual risk of offenders (e.g., Kleck et al., 2005), suggesting that perceived risk is always influenced by other factors. In the current theme, individual characteristics and crime milieu characteristics contribute to the variation on drug offenders' perception of risk and use of strategy.

Individual Characteristics

People with specific characteristics become flexible in perceiving risks, acting out planned strategies, and avoiding punishment. Examples illustrate that gender (Jacobs and Miller, 1998), age (Paternoster, 1989) and social attachment (Ekland-Olson et al., 1984; DeJong, 1997) affect the action of punishment avoidance.

In most criminal subcultures, gender inequality exists (Jacobs and Miller, 1998). Women are more likely to perceive risks than men because they have lower fault tolerance in society (Carbone-Lopez, 2015). Women prefer not to implement the detection avoidance strategies that men frequently use, even if they share similar motivations. Jacobs and Miller (1998) identified that female drug dealers developed female-oriented restrictive deterrence strategies that exploited gender and normative beliefs about femininity to render an antagonistic audience neutral or perhaps even friendly.

In addition to gender, age also affects how drug offenders perceive risk and adopt strategies. Adolescents who are potential marijuana users perceive a higher perception of risk as they age; in other words, they become sensitive to a set of opportunities to commit delinquency (Paternoster, 1989).

As for the social attachment, drug offenders with solid bonds with conventional society (marriage and employment) are likely to reconsider risk before the crime. They fear losing the investment they have made in prosocial domains, leading them to commit crimes less frequently and a longer time before re-arrests (DeJong, 1997). Compared to conventional social bonds, ties with other offenders also variate the drug dealers' risk perception. Drug dealers in a dense and closed criminal network perceive less risk as they trust their co-actors; therefore, they discourage the formation and maintenance of weak ties and act as a restrictive deterrence strategy (Ekland-Olson et al., 1984).

Psychological status is another relevant individual characteristic that affects drug offenders' perception of risk and subsequent behavior. Drug offenders with experience in avoiding detection undermine risk sensitivity, as they reckon that they are more capable of escaping detection than anyone else and they become overly confident and reckless when carrying out crimes (Piliavin et al., 1986; Jacobs, 2010; Carbone-Lopez, 2015). Correspondingly, offenders who have been previously deterred are likely to produce a "reset" estimation (Pogarsky and Piquero, 2003) since they believe that arrest is rare and unlikely to occur again so soon afterward (Gallupe et al., 2011; Dickinson and Wright, 2015).

In addition to the experience in avoidance or being deterred, the existence of co-offenders also spurs drug offenders' self-serving bias and compromises their perception of risk. Accomplices decrease the fear of detection and generate social support for severe illegal acts. Co-offenders in drug offenses offer a greater feeling of control, scatter blame for the crime, and foster feelings of invulnerability; thus, the perception of risk is further compromised and spurs on individual participants (Johnson and Natarajan, 1995; Jacobs and Miller, 1998; Cross, 2000).

Another psychological status that impairs drug offenders' perception of risk is the perceived benefits of crime. The longer the drug offenders make a "career" in drug crime, the more immersed they become in the lucrative lifestyle, which reduces their perceptions of risk and boosts the perceptual rewards of crime (Ekland-Olson et al., 1984; Jacobs, 1993; Moloney et al., 2015). However, drug cultivators are an exception; they are involved in considerable planning and investment. Starting a cultivation site, large or small, can take several months with ongoing maintenance and care (Nguyen et al., 2015). Thus, drug cultivators have to remain sober and cautious of risk changes to readily adjust the drug plants, such as reducing the area cultivated.

Crime Milieu Characteristics

Recent research has focused on how crime milieu affects offenders' perceptions and responses to risk (Pratt et al., 2006; Piquero et al., 2011). The crime milieu is full of unexpected and twisted events which spur offenders' fast response. Offenders with a present-minded propensity are more responsive to unexpected

risks and have a greater capacity to adapt to them when compared to those with a future-minded propensity (Jacobs and Cherbonneau, 2018). Being present-minded assists offenders in committing crime successfully where rationality is strictly limited. It is consistent with the concept of "mindfulness" in psychology, which stresses a capacity that helps decision-makers to block out the "noise" that hinders effective choices in unpredictable settings (Jacobs and Cherbonneau, 2018).

Drug dealing is a socially situated phenomenon (Dickinson and Wright, 2015). Drug offenders have to pay attention to the crime milieu in which the crime is about to be committed. For example, when considering natural surveillance, dealers prefer to adopt immediate transactions in a commercial area with a high level of both vehicle and pedestrian traffic. When considering formal surveillance, they prefer to avoid places with CCTV (Piza and Sytsma, 2016). Sometimes, the crime milieu is full of complexity. Drug offenders have to deal with multiple risks at the same time. It has been revealed that drug dealers have to deal on busy street segments even if there are intensified police patrols or CCTV cameras since the buyers often show up there. They develop detection avoidance strategies, including walking around and not staying in a single spot for long periods, hiding drugs in off-person stash spots, and being careful to keep their faces or bodies out of reach of the view of CCTV cameras (Bernasco and Jacques, 2015).

Skills in Crime

Crime skills work as a guidance for offenders to implement restrictive deterrence strategies. It lets offenders know how effective their efforts are and helps them adjust strategies in real-time (Topalli et al., 2015). Two ways that offenders acquire their skills in crime to enhance their performance are self-reflection and collective wisdom.

Self-Reflection

Drug offenders' crime skills largely depend on the experiential learning effect; in short, offenders learn by doing (Gallupe et al., 2011). Regardless of an experience of failure or success, experience is always a chance to advance an evolving crime-specific learning curve. Even spending time in jail stimulates restrictive deterrence. It has been found that imprisonment is related to an increased likelihood of ongoing violation for certain criminals (DeJong, 1997). Drug offenders who have long intervals before re-arrests have learned from earlier failures and have personally enacted restrictive deterrence strategies (Gallupe et al., 2011). Learning from personal experience enables drug offenders to survive longer and hence expand the scale of their operations.

The acquisition of crime skills through self-reflection is subtle. Many offenders do not notice the improvements so they deem criminal skills a certain intuition or instinct instead of an intellectual process (Johnson and Natarajan, 1995; VanNostrand and Tewksbury, 1999). It is undeniable that the more proficient the offenders in committing crime, the more natural the crime skill becomes (Nee and Ward, 2015), but it does not obscure the fact that crime skill is built up through learning. Like other specialzed criminals, drug offenders have to devote time and energy to integrate in specific scenes to acknowledge the social

nuances within drug markets. Instead of intuition, repeated exposure in observing the streets enables drug offenders to identify undercover police officers by their movements, speech, and actions (Jacobs, 1996a; VanNostrand and Tewksbury, 1999; Jacques and Reynald, 2012). In some cases, nuanced changes in the accumulation of crime skills facilitate restrictive deterrence. For instance, Gallupe et al. (2011) suggested that punishment avoidance techniques can be more successful if the drug offenders conduct well-thought-out adjustments rather than impulsively implementing a complete revamp.

Collective Wisdom

Collective wisdom is more important for facilitating the learning process than self-reflection, especially for novices. Novices have limited experience in recognizing undercover police officers. They need vicarious experience to form punishment avoidance strategies. Re-offenders also rely on vicarious experience, as criminal experience is not only obtained based on how many times a crime is committed but also by how many types of crime are committed (Knowles, 1999). Under the screening of collective wisdom, useless punishment avoidance strategies are discarded and effective ones are pursued. Gossip plays an indispensable role in spreading the collective wisdom among active drug dealers (VanNostrand and Tewksbury, 1999; Dickinson and Wright, 2015). Drug dealers maintain informal information channels to keep track of police routines, such as shifts or patrol timetables (Jacobs, 1993; Johnson and Natarajan, 1995). They keep an eye on and gossip about clients, staff, associates, and suppliers who have had some contact with police or have recently behaved dubiously (Dickinson and Wright, 2015). Besides verbal communication, observing others' dealing activities is essential to understand the local drug markets, such as nuanced details while trading (Johnson and Natarajan, 1995; Jacobs, 1996a).

Learning from vicarious experience, drug offenders accelerate their learning curve (Bouchard and Nguyen, 2010; Fader, 2016a; Malm et al., 2017); however, not all drug offenders take advantage of collective wisdom. Some offenders proactively or passively obtain less access to drug organizations and information (Ekland-Olson et al., 1984; Jacobs and Miller, 1998; Erickson et al., 2013; Moloney et al., 2015). Besides, offenders do not blindly obey every instruction that the collective wisdom provides. Reactions to the gossip rely on how gossip subjects are caught, the social distance between listeners, and gossip subjects and sources (Dickinson and Wright, 2015). To illustrate, when hearing gossip about possible police informants, drug dealers commonly prevent connections with all or any of associates considered as police informants, at least for a short time. However, they would not alienate a recent associate who was arrested, if it was for a traffic matter. Additionally, they prefer to avoid a gossip subject when they are close to the gossip source and are less likely to avoid gossip when they are close to the gossip subject.

DISCUSSION

Conducting a meta-synthesis of the findings from 34 studies, this systematic review offers evidence relating to drug offenders' restrictive deterrence. Two prominent themes, namely

"restrictive deterrence strategy" and "deterrability and restrictive deterrence," emerge as a picture that depicts the whole process of drug offenders' restrictive deterrence and fit well in the certainty–severity framework of punishment (**Figure 2**).

Perhaps the most important conclusion of this review relates to the finding that the two types of restrictive deterrence strategies are explored equally in the reviewed papers. Restrictive deterrence strategies directly influence "whether" (certainty reduction) and "how" (severity mitigation) drug offenders will be punished. The parity of discussion between the severity and the certainty of punishment is uncommon in prior deterrence studies. The imbalanced topic distribution, specifically that most of the studies focused on the certainty of punishment, may be due to the different deterrent effects of certainty and severity of punishment. Firstly, it is generally accepted that the certainty of punishment exerts a significantly stronger and more stable deterrent effect on offenders than the severity of the punishment (DeJong, 1997; Pogarsky, 2002; Nagin and Pogarsky, 2003). Secondly, the deterrent effect of the severity of punishment relies on the certainty of punishment. As Beccaria (1963, p 58) wrote, "[t]he certainty of a punishment, even if it be moderate will always make a stronger impression than the fear of another which is more terrible but combined with the hope of impunity; even the least evils, when they are certain, always terrify men's minds".

Although the severity of punishment has been devalued compared to the certainty of punishment, it acts as a significant catalyst that stimulates the whole deterrence process. Roche et al. (2020) revealed that offenders' perception of punishment severity significantly affected certainty. Furthermore, evidence of the anchoring effect from behavioral economics indicated that an individual is influenced by a specific number (or "anchor") when making a statistical estimation (e.g., about a probability), and unintentionally keeps the statistical estimation close to the anchor (Tversky and Kahneman, 1974). Studies revealed that an individual's perceived certainty of risk is highly volatile and one must rank certainty by anchoring the reality, and the anchor here refers to the perceived severity of consequences for committing different offenses (Nagin, 1998; Pogarsky et al., 2018; Thomas et al., 2018).

Under the terrain of restrictive deterrence, drug offenders are no less apprehensive about the severity of punishment than its certainty. This could be due to the strategy of severity mitigation, which influences the extent of using the strategy of certainty reduction. To illustrate, a marijuana seller is less concerned about the timing and location of sales than a heroin dealer. A drug dealer with 0.01 grams of heroin on their person is less likely to care if they dress or behave in a way that will attract the attention of the police than a dealer with 100 grams of heroin. Theoretically, this echoes the aforementioned idea that the perception of the severity of punishment sets an anchor for that of certainty (Nagin, 1998; Pogarsky et al., 2018; Thomas et al., 2018). Another potential explanation is that, among deterrable offenders, the severity of punishment provides a more significant deterrent effect than the certainty of punishment (Pogarsky, 2002). In this way, the severity effect reasserts its vital power throughout the deterrence process.

It is worth noting that the severity of punishment is primarily examined in quantitative studies and certainty of punishment is usually discussed in qualitative studies. These two observations may reveal the difference in design between qualitative and quantitative studies. Qualitative studies, using mainly semi-structured interviews, usually design interviews with a relatively broad range of questions and do not strictly separate the severity of punishment from the certainty of punishment. In conjunction with what has been mentioned earlier that the certainty of punishment has a higher profile in deterrence research overall, it is easy to attribute the role, effect, or importance of severity of punishment to the certainty of punishment when interpreting drug offenders' responses to punishment. In contrast, quantitative research can clearly separate the two elements through questionnaire design and examine and demonstrate the role of severity of punishment while controlling for the role of certainty of punishment. The severity of punishment can also be explored interactively with other variables, which facilitates the identification of the role of severity of punishment in a given population or a given situation and tap into its once-overlooked position. Therefore, the qualitative study tends to examine the role of certainty of punishment, while the quantitative study is better equipped to uncover the role of severity of punishment.

The second key conclusion from this review is that the perception of risk is not only an inhibitor in using restrictive deterrence strategies, but also a facilitation of strategy differentiation. First, we found that due to the psychology of self-serving bias, drug offenders with less perception of risk implement restrictive deterrence strategies with confidence. This is consistent with the literature on the perceptual risk, which states that the lower the perception of risk, the more significant the crime (e.g., Pratt et al., 2006). However, we also discovered that the perception of risk motivates, rather than undermines, drug offenders to innovate strategies for committing crimes. For instance, female drug dealers struggled in a male-dominated field to innovate new strategies to avoid detection. Drug dealers selling drugs on CCTV-equipped streets develop a strategy to move without having their faces be captured by a camera. Such a contribution of risk perception to a cautious mindset is consistent with "flaw hunting" (Walsh, 1986), which refers to the notion that offenders sometimes utilize the high perception of risk of getting caught as an incentive for proper planning (Cherbonneau and Copes, 2006; Jacobs and Cherbonneau, 2014, 2016).

The final conclusion of this review is that the formation of crime skill relies on a combination of the slow internalization of self-reflection and the rapid input of collective wisdom. Self-reflection leads to nuanced adjustments of restrictive deterrence strategies. At the same time, collective wisdom accelerates the progress of skill learning because it reduces the individual's cost of trial and error. Perhaps because of the two different contributions to the speed and magnitude of crime skill learning, crime skill formed through self-reflection is not considered a learning process and is thus classified as an intuition. This echoes the findings of offenders' Bayesian learning based on personal experience. Bayesian learning is a way in which individuals incorporate newly learned information to update subjective prior

beliefs. Anwar and Loughran (2011) found that the weight of unobserved signals (including peers' experience) when offenders consider potential risk is nearly eight times greater than the weight of considering their own arrest rate. The slight weight that is put on personal experience implies that this component is not being taken seriously.

While the reviewed research revealed fruitful restrictive deterrence strategies and their potential prerequisites, it still leaves room to explore uncharted topics that can promote our understanding of the issues. First, restrictive deterrence strategies were mainly discussed in the pre-arrest context, and future research could further explore strategies used during and after arrest. Before the arrest, by the use of proactive situational control over the context, offenders can minimize their chances of being arrested. However, this much focuses on the context of pre-arrest, delivers an incorrect presumption that drug offenders do not respond to risk once they are arrested. An arrest does not equate with the final sentence, e.g., imprisonment. Punishment is a system of conditional probabilities: restrictive deterrence strategies before an arrest can influence the outcome of punishment, and restrictive deterrence strategies after an arrest can achieve this effect as well. Between arrest and the final sentence, a series of judicial proceedings can affect the outcome, such as prosecution, conviction, and the dismissal of charges. As the "Cooperating with police" section of the current review shows, drug offenders beg or cooperate with the police to mitigate the expected severity of punishment. It is an evidence of the restrictive deterrence strategy adopted during and after arrested. Similar restrictive deterrence strategy has also been found in the study of other offenders. Sex worker, for example, might be very polite and compliant with police during arrest in the hope that they would be charged with a less severe crime (Dewey and Germain, 2014). In addition, offenders make decision on guilty plea or withdraw it in the hope that the punishment would be changed (Cheng et al., 2018). It is evident that offenders have to negotiate and deal with the authorities during and after arrest. They may come up with a completely different restrictive deterrence strategy than that used pre-arrest. Therefore, exploring the after-arrest strategy provides insights into how offenders negotiate with authority.

Second, determining the facilitative effect of the perception of risk on the innovation of restrictive deterrence strategies is an area for future research. While much of the broader deterrence literature has quantified the role of the perception of risk in curbing crime or an individual's intention to commit a crime (e.g., Pratt et al., 2006), in ethnographic studies of restrictive deterrence it is implied that the perception of risk stimulates innovation in crime strategies (e.g., Jacobs and Miller, 1998). This is a relatively novel idea that emphasizes the vital power of the perception of risk from the opposite perspective. Nevertheless, it is unclear what type of perception of risk motivates offenders' innovation or planning ability and willingness rather than acting as a deterrent. For instance, increased offender risk perception and vigilance may be an early warning signal. Indeed, prior restrictive deterrence research based on a sample of hackers suggested that a warning banner significantly reduces the duration of trespassing incidents (Maimon et al., 2014). The

mediation that connects the perception of risk and strategy innovation is also unclear, both psychological and social. Prior research revealed that offenders with better emotional management are good at translating perceptions of risk into better risk coping strategies, and peer support reinforces their emotional management (Jacobs and Cherbonneau, 2017). We definitely do not wish to only dwell on how the perception of risk exerts its facilitative effect on strategy differentiation. Exploring the aforementioned issues would allow the literature to better understand how the two roles of perception of risk (curbing crime/facilitating strategy development) reconcile.

Another future research line for exploration is to look into the impact of collective wisdom on crime skill learning and strategy use. The influence of collective wisdom on offenders' crime skills is more significant than individual reflection regarding speed and quantity. However, the effectiveness of collective wisdom highly depends on the offender's closeness to other offenders or crime organizations. It is noted that the connection to criminal groups is beneficial for drug offenders to obtain advanced and effective strategies (Ekland-Olson et al., 1984; Jacobs and Miller, 1998; Moloney et al., 2015). Others have attempted to quantify the role of collective wisdom (peer experience) in the learning process (e.g., Pogarsky et al., 2004), but as of yet it remains unclear how ties play into this. Hence, research can be extended to explore what kind of tie is efficient to diffuse collective wisdom on crime skill and through what kind of communication paths offenders are more likely to accept and adopt the crime skill. In other words, how relationships affect the rate of transmission and acceptance of collective wisdom. Studying different channels that spread crime skills and making comparisons among them can generate insights about the iteration of restrictive deterrence. Furthermore, the literature needs more details on restrictive deterrence advances in drug offender groups.

Interpretation of our results should be tempered by several limitations. First, although the literature search was comprehensive, only a small number of studies could be included compared to previous literature reviews focusing on deterrence or drug criminality (e.g., Pratt et al., 2006). Simply put, only studies that considered and discussed restrictive deterrence of drug offenders seriously were included. Studies that merely referred to crime strategies of drug offenders in a broader research question, e.g., drug economy (e.g., Dickinson, 2020), were excluded but may provide additional insights. Second, we specifically selected only English studies when conducting data search, which means that studies published in other languages may have reported different conclusions and impacts. Therefore, this review is limited in its evaluation of cross-cultural aspects in restrictive deterrence of drug offenders. Finally, several important questions remain unanswered in the current review. For instance, which type of deterrent strategies is most effective against which kind of drug offenses (e.g., using, selling, cultivating drugs, etc.), and which type of restrictive deterrence strategy works best in which kind of situations (e.g., gender, drug type, time, places). We were not able to perform further analysis due to a general dearth of quantitative study in the literature related to restrictive deterrence of drug offenders.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

XG organized the database and wrote the first draft of the manuscript. TWL contributed to reviewing and editing the manuscript. Both authors contributed to manuscript revision, read, and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.727142/full#supplementary-material

REFERENCES

Anwar, S., and Loughran, T. A. (2011). Testing a Bayesian learning theory of deterrence among serious juvenile offenders. *Criminology* 49, 667–698. doi: 10.1111/j.1745-9125.2011.00233.x

Apel, R. (2013). Sanctions, perceptions, and crime: Implications for criminal deterrence. *J. Quant. Criminol.* 29, 67–101. doi: 10.1007/s10940-012-9170-1

Beauregard, E., and Bouchard, M. (2010). Cleaning up your act: Forensic awareness as a detection avoidance strategy. *J. Crim. Justice* 38, 1160–1166. doi: 10.1016/j.jcrimjus.2010.09.004

Beccaria, C. (1963). *On Crimes and Punishments*. Indianapolis: Bobbs—Merrill.

Beck, C. T. (2002). Mothering multiples: a meta-synthesis of qualitative research. *MCN: The Am. J. Mater. Child Nurs.* 27, 214–221. doi: 10.1097/00005721-200207000-00004

Bernasco, W., and Jacques, S. (2015). Where do dealers solicit customers and sell them drugs? *A micro-level multiple method study. Journal of Contemporary Criminal Justice* 31, 376–408. doi: 10.1177/1043986215608535

Bouchard, M., and Nguyen, H. (2010). Is it who you know, or how many that counts? criminal networks and cost avoidance in a sample of young offenders. *Justice Q.* 27, 130–158. doi: 10.1080/07418820802593386

Carbone-Lopez, K. (2015). Above the law: Changes in methamphetamine laws and the deterrent impact on market-involved women. *Am. J. Crim. Justice* 40, 682–701. doi: 10.1007/s12103-014-9281-3

Casey, S. (2015). Offending: drug-related expertise and decision making. *Aggr. Viol. Behav.* 20, 82–91. doi: 10.1016/j.avb.2014.12.010

Cheng, K. K.-,y., Chui, W. H., Yong, S. N., and Rebecca, O. (2018). Why Do Criminal Trials 'Crack'? An empirical investigation of late guilty pleas in Hong Kong. *Asian J. Comparat. Law* 13, 1–25. doi: 10.1017/asjcl.2017.27

Cherbonneau, M., and Copes, H. (2006). "Drive it like you stole it:" auto theft and the illusion of normalcy. *Br. J. Criminol.* 46, 193–211. doi: 10.1093/bjc/azi059

Clemmens, D. (2003). Adolescent motherhood: a meta-synthesis of qualitative studies. *MCN Am. J. Mater. Child Nurs.* 28, 93–99. doi: 10.1097/00005721-200303000-00010

Cooke, A., Smith, D., and Booth, A. (2012). Beyond PICO: the SPIDER tool for qualitative evidence synthesis. *Qual. Health Res.* 22, 1435–1443. doi: 10.1177/1049732312452938

Cross, J. C. (2000). Passing the buck: Risk avoidance and risk management in the illegal/informal drug trade. *Int. J. Sociol. Soc. Policy* 20, 68–94. doi: 10.1108/01443330010789232

Decker, S., Wright, R., and Logie, R. (1993). Perceptual deterrence among active residential burglars: A research note. *Criminology* 31, 135–147. doi: 10.1111/j.1745-9125.1993.tb01125.x

DeJong, C. (1997). Survival analysis and specific deterrence: Integrating theoretical and empirical models of recidivism. *Criminology* 35, 561–576. doi: 10.1111/j.1745-9125.1997.tb01230.x

Dewey, S., and Germain, T. S. (2014). "It depends on the cop:" Street-based sex workers' perspectives on police patrol officers. *Sex. Res. Soc. Policy* 11, 256–270. doi: 10.1007/s13178-014-0163-8

Dickinson, T. (2020). Identity, situational elements, and responses to drug debt. *J. Res. Crime Delinquency* 57, 66–104. doi: 10.1177/0022427819866495

Dickinson, T., and Wright, R. (2015). Gossip, decision-making and deterrence in drug markets. *Br. J. Criminol.* 55, 1263–1281. doi: 10.1093/bjc/azv010

Eck, J. E., and Wartell, J. (1998). "Improving the management of rental properties with drug problems: A randomized experiment," in *Civil Remedies and Crime Prevention*, eds L. G. Mazerolle and J. Roehl (Monsey, NY: Criminal Justice Press), 161–185.

Ekland-Olson, S., Lieb, J., and Zurcher, L. (1984). The paradoxical impact of criminal sanctions: some microstructural findings. *Law Soc. Rev.* 18, 159–178. doi: 10.2307/3053401

Erickson, P. G., Van Der Maas, M., and Hathaway, A. D. (2013). Revisiting deterrence: legal knowledge, use context and arrest perception for cannabis. *Czech Sociol. Rev.* 49, 427–449. doi: 10.13060/00380288.2013.49.3.04

Fader, J. J. (2016a). Criminal family networks: criminal capital and cost avoidance among urban drug sellers. *Deviant Behav.* 37, 1325–1340. doi: 10.1080/01639625.2016.1177388

Fader, J. J. (2016b). "Selling smarter, not harder:" Life course effects on drug sellers' risk perceptions and management. *Int. J. Drug Policy* 36, 120–129. doi: 10.1016/j.drugpo.2016.04.011

Fader, J. J. (2019). "The game ain't what it used to be:" Drug sellers' perceptions of the modern day underground and legal markets. *J. Drug Issues* 49, 57–73. doi: 10.1177/0022042618803057

Fleetwood, J. (2014). Keeping out of trouble: female crack cocaine dealers in England. *Euro. J. Criminol.* 11, 91–109. doi: 10.1177/1477370813491177

Gallupe, O., Bouchard, M., and Caulkins, J. P. (2011). No change is a good change? restrictive deterrence in illegal drug markets. *J. Crim. Just.* 39, 81–89. doi: 10.1016/j.jcrimjus.2010.12.002

Gibbs, J. P. (1975). *Crime, Punishment, and Deterrence*. New York, NY: Elsevier.

Harrison, L. D. (2001). The revolving prison door for drug-involved offenders: challenges and opportunities. *Crime Delinq.* 47, 462–485. doi: 10.1177/0011128701047003010

Jacobs, B. A. (1993). Undercover deception clues: a case of restrictive deterrence. *Criminology* 31, 281–299. doi: 10.1111/j.1745-9125.1993.tb01131.x

Jacobs, B. A. (1996a). Crack dealers and restrictive deterrence: identifying narcs. *Criminology* 34, 409–431. doi: 10.1111/j.1745-9125.1996.tb01213.x

Jacobs, B. A. (1996b). Crack dealers' apprehension avoidance techniques: a case of restrictive deterrence. *Justice Q.* 13, 359–381. doi: 10.1080/07418829600093011

Jacobs, B. A. (2010). Deterrence and deterrability. *Criminology* 48, 417–441. doi: 10.1111/j.1745-9125.2010.00191.x

Jacobs, B. A., and Cherbonneau, M. (2014). Auto theft and restrictive deterrence. *Justice Q.* 31, 344–367. doi: 10.1080/07418825.2012.660977

Jacobs, B. A., and Cherbonneau, M. (2016). Managing victim confrontation: auto theft and informal sanction threats. *Justice Q.* 33, 21–44. doi: 10.1080/07418825.2014.891638

Jacobs, B. A., and Cherbonneau, M. (2017). Nerve management and crime accomplishment. *J. Res. Crime Delinq.* 54, 617–638. doi: 10.1177/0022427817693037

Jacobs, B. A., and Cherbonneau, M. (2018). Perceived sanction threats and projective risk sensitivity: auto theft, carjacking, and the channeling effect. *Justice Q.* 35, 191–222. doi: 10.1080/07418825.2017.1301536

Jacobs, B. A., and Miller, J. (1998). Crack dealing, gender, and arrest avoidance. *Soc. Probl.* 45, 550–569. doi: 10.2307/3097212

Jacques, S., and Allen, A. (2014). Bentham's sanction typology and restrictive deterrence: a study of young, suburban, middle-class drug dealers. *J. Drug Issues* 44, 212–230. doi: 10.1177/0022042613497936

Jacques, S., and Reynald, D. M. (2012). The offenders' perspective on prevention: guarding against victimization and law enforcement. *J. Res. Crime Delinq.* 49, 269–294. doi: 10.1177/0022427811408433

Johnson, B. D., and Natarajan, M. (1995). Strategies to avoid arrest: crack sellers' response to intensified policing. *Am. J. Police* 14, 49–69. doi: 10.1108/07358549510111947

Kleck, G., Sever, B., Li, S., and Gertz, M. (2005). The missing link in general deterrence research. *Criminology* 43, 623–660. doi: 10.1111/j.0011-1348.2005.00019.x

Knowles, G. J. (1999). Deception, detection, and evasion: a trade craft analysis of Honolulu, Hawaii's street crack-cocaine traffickers. *J. Crim. Justice* 27, 443–455. doi: 10.1016/S0047-2352(99)00015-X

Maimon, D., Alper, M., Sobesto, B., and Cukier, M. (2014). Restrictive deterrent effects of a warning banner in an attacked computer system. *Criminology* 52, 33–59. doi: 10.1111/1745-9125.12028

Malm, A., Bouchard, M., Decorte, T., Vlaemynck, M., and Wouters, M. (2017). More structural holes, more risk? network structure and risk perception among marijuana growers. *Soc. Netw.* 51, 127–134. doi: 10.1016/j.socnet.2017.01.006

Moeller, K., Copes, H., and Hochstetler, A. (2016). Advancing restrictive deterrence: a qualitative meta-synthesis. *J. Crim. Justice* 46, 82–93. doi: 10.1016/j.jcrimjus.2016.03.004

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 6, 1–9. doi: 10.1371/journal.pmed.1000097

Moloney, M., Hunt, G., and Joe-Laidler, K. (2015). Drug sales, gender, and risk: notions of risk from the perspective of gang-involved young adults. *Substance Use Misuse* 50, 721–732. doi: 10.3109/10826084.2015.978642

Nagin, D. S. (1998). Criminal deterrence research at the outset of the twenty-first century. *Crime Justice* 23, 1–42. doi: 10.1086/449268

Nagin, D. S. (2013). Deterrence in the Twenty-First Century. *Crime Justice* 42, 199–263. doi: 10.1086/670398

Nagin, D. S., and Pogarsky, G. (2003). An experimental investigation of deterrence: cheating, self-serving bias, and impulsivity. *Criminology* 41, 167–194. doi: 10.1111/j.1745-9125.2003.tb00985.x

Nee, C., and Ward, T. (2015). Review of expertise and its general implications for correctional psychology and criminology. *Aggress. Violent Behav.* 20, 1–9. doi: 10.1016/j.avb.2014.12.002

Nguyen, H., Malm, A., and Bouchard, M. (2015). Production, perceptions, and punishment: restrictive deterrence in the context of cannabis cultivation. *Int. J. Drug Policy* 26, 267–276. doi: 10.1016/j.drugpo.2014.08.012

Olaghere, A., and Lum, C. (2018). Classifying "micro" routine activities of street-level drug transactions. *J. Res. Crime Delinq.* 55, 466–492. doi: 10.1177/0022427818760103

Panda, S., Begley, C., and Daly, D. (2018). Clinicians' views of factors influencing decision-making for caesarean section: a systematic review and metasynthesis of qualitative, quantitative and mixed methods studies. *PLoS ONE* 13, 1–27. doi: 10.1371/journal.pone.0200941

Paternoster, R. (1989). Decisions to participate in and desist from four types of common delinquency: deterrence and the rational choice perspective. *Law Soc. Rev.* 23, 7–40. doi: 10.2307/3053879

Paternoster, R. (2010). How much do we really know about criminal deterrence? *The Journal of Criminal Law and Criminology* 100, 765–824.

Piliavin, I., Gartner, R., Thornton, C., and Matsueda, R. L. (1986). Crime, deterrence, and rational choice. *Am. Sociol. Rev.* 51, 101–119. doi: 10.2307/2095480

Piquero, A. R., Paternoster, R., Pogarsky, G., and Loughran, T. (2011). Elaborating the individual difference component in deterrence theory. *Ann. Rev. Law Soc. Sci.* 7, 335–360. doi: 10.1146/annurev-lawsocsci-102510-105404

Piza, E. L., and Sytsma, V. A. (2016). Exploring the defensive actions of drug sellers in open-air markets: a systematic social observation. *J. Res. Crime Delinq.* 53, 36–65. doi: 10.1177/0022427815592451

Pogarsky, G. (2002). Identifying "deterrable" offenders: Implications for research on deterrence. *Justice Q.* 19, 431–452. doi: 10.1080/07418820200095301

Pogarsky, G., and Piquero, A. R. (2003). Can punishment encourage offending? Investigating the "resetting" effect. *J. Res. Crime Delinq.* 40, 95–120. doi: 10.1177/0022427802239255

Pogarsky, G., Piquero, A. R., and Paternoster, R. (2004). Modeling change in perceptions about sanction threats: the neglected linkage in deterrence theory. *J. Quant. Criminol.* 20, 343–369. doi: 10.1007/s10940-004-5868-z

Pogarsky, G., Roche, S. P., and Pickett, J. T. (2018). Offender decision-making in criminology: Contributions from behavioral economics. *Ann. Rev. Criminol.* 1, 379–400. doi: 10.1146/annurev-criminol-032317-092036

Pratt, T. C., Cullen, F. T., Blevins, K. R., Daigle, L. E., and Madensen, T. D. (2006). "The empirical status of deterrence theory: a meta-analysis," in *Taking Stock: The Status of Criminological Theory.*, eds. F.T. Cullen, J.P. Wright and K.R. Blevins (Piscataway, NJ, US: Routledge), 367–395.

Ribeiro, L. A., Sanchez, Z. M., and Nappo, S. A. (2010). Surviving crack: a qualitative study of the strategies and tactics developed by Brazilian users to deal with the risks associated with the drug. *BMC Public Health* 10, 1–10. doi: 10.1186/1471-2458-10-671

Roche, S. P., Wilson, T., and Pickett, J. T. (2020). Perceived control, severity, certainty, and emotional fear: testing an expanded model of deterrence. *J. Res. Crime Delinquency* 57, 493–531. doi: 10.1177/0022427819888249

Ryan, K. (1994). Technicians and interpreters in moral crusades: the case of the drug courier profile. *Deviant Behav.* 15, 217–240. doi: 10.1080/01639625.1994.9967971

Sandelowski, M., Docherty, S., and Emden, C. (1997). Qualitative metasynthesis: issues and techniques. *Res. Nurs. Health* 20, 365–371. doi: 10.1002/(SICI)1098-240X(199708)20:4andlt;365::AID-NUR9andgt;3.0.CO;2-E

Schutz, A. (1972). *The Phenomenology of the Social World.* Evanston, IL: Northwestern University Press.

Thomas, K. J., Hamilton, B. C., and Loughran, T. A. (2018). Testing the transitivity of reported risk perceptions: evidence of coherent arbitrariness. *Criminology* 56, 59–86. doi: 10.1111/1745-9125.12154

Topalli, V., Jacques, S., and Wright, R. (2015). "It takes skills to take a car:" perceptual and procedural expertise in carjacking. *Aggress. Violent Behav.* 20, 19–25. doi: 10.1016/j.avb.2014.12.001

Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science* 185, 1124–1131. doi: 10.1126/science.185.4157.1124

VanNostrand, L.-M., and Tewksbury, R. (1999). The motives and mechanics of operating an illegal drug enterprise. *Deviant Behav.* 20, 57–83. doi: 10.1080/016396299266597

Vidal, S., and Décary-Hétu, D. (2018). Shake and bake: Exploring drug producers' adaptability to legal restrictions through online methamphetamine recipes. *J. Drug Issues* 48, 269–284. doi: 10.1177/0022042617751685

Walsh, D. (1986). "Victim Selection Procedures Among Economic Criminals," in *The Reasoning Criminal: Rational Choice Perspectives on Offending*, eds. D.B. Cornish and R.V. Clarke. (New York, NY: Springer-Verlag) 39–52.

Williams, K. R., and Hawkins, R. (1989). The meaning of arrest for wife assault. *Criminology* 27, 163–182. doi: 10.1111/j.1745-9125.1989.tb00867.x

Wilson, T., Maimon, D., Sobesto, B., and Cukier, M. (2015). The effect of a surveillance banner in an attacked computer system: Additional evidence for the relevance of restrictive deterrence in cyberspace. *J. Res. Crime Delinq.* 52, 829–855. doi: 10.1177/0022427815587761

The Interactive Effects of Race and Expert Testimony on Jurors' Perceptions of Recanted Confessions

Logan Ewanation[1]* and Evelyn M. Maeder[2]

[1] Department of Psychology, Carleton University, Ottawa, ON, Canada, [2] Institute of Criminology and Criminal Justice, Carleton University, Ottawa, ON, Canada

We examined the effect of defendant race and expert testimony on jurors' perceptions of recanted confessions. Participants (591 jury-eligible community members) read a first-degree murder trial transcript in which defendant race (Black/White) and expert testimony (present/absent) were manipulated. They provided verdicts and answered questions regarding the confession and expert testimony. When examining the full sample, we observed no significant main effects or interactions of defendant race or expert testimony. When exclusively examining White participants, we observed a significant interaction between expert testimony and defendant race on verdicts. When the defendant was White, there was no significant effect of expert testimony, but when the defendant was Black, jurors were significantly more likely to acquit when given expert testimony. These findings support the watchdog hypothesis, such that White jurors are more receptive to legally relevant evidence when the defendant is Black.

Keywords: juror decision-making, recanted confessions, watchdog hypothesis, expert testimony, juries, confession evidence, defendant race

INTRODUCTION

Empirical research indicates that suspects falsely confess to crimes for a variety of reasons (Kassin and Kiechel, 1996; King and Snook, 2009). According to the Innocence Project (2021), false confessions[1] were involved in approximately a quarter of the cases that have been exonerated through DNA evidence. However, confessions remain one of the most influential forms of evidence in the courtroom (Kassin and Neumann, 1997; Lieberman et al., 2008). Although expert witnesses are sometimes used to safeguard against issues with confession evidence, the effect of expert testimony on jurors' perceptions of recanted confessions is unclear (Moffa and Platania, 2007; Blandon-Gitlin et al., 2011). Further, jurors' perceptions of recanted confessions may depend upon the suspect's race, as jurors perceive confessions as more voluntary when the defendant belongs to a racial minority (Pickel et al., 2013).

Most research focusing on the interaction between juror and defendant race has found that jurors are more lenient toward same-race defendants (see Devine and Caughlin, 2014). However, Sargent and Bradfield (2004) found that White mock jurors were more sensitive to legally relevant

---

[1]The Innocence Project includes false admissions and other forms of self-incrimination in their definition of false confessions.

evidence in a trial transcript when the defendant was Black as compared to White. These authors argued that White jurors may attend to evidence more closely when the defendant is Black in an effort to serve as "watchdogs" against racism (termed the watchdog hypothesis). In a case in which a defendant has recanted their confession, the watchdog hypothesis would suggest that jurors may be more receptive to expert testimony (regarding factors that increase the likelihood of false confessions) when the defendant is Black, resulting in fewer guilty verdicts. The current study examines the interactive effects of defendant race and expert testimony on jurors' perceptions of recanted confessions.

Confession Evidence

Empirical research has demonstrated that individuals may falsely confess to crimes that they did not commit (e.g., Kassin and Kiechel, 1996; Redlich et al., 2010) for a variety of reasons, including coercive interrogation tactics (Kassin et al., 2003; King and Snook, 2009). However, in a criminal trial, a defendant's confession is one of the most influential forms of evidence that the prosecution can present (Kassin and Neumann, 1997; Lieberman et al., 2008; Schweitzer and Nuñez, 2018). For instance, Lieberman et al. (2008) demonstrated that among several types of evidence, the only type that participants perceived as more persuasive than a suspect's confession was DNA analysis. Even then, there are plenty of anecdotal instances in which law enforcement officials have ignored exculpatory DNA evidence in investigations when the suspect has confessed (e.g., the Central Park jogger case, Juan Rivera, the Norfolk Four; Duru, 2003; Leo and Davis, 2010). Furthermore, Appleby and Kassin (2016) conducted a series of jury studies involving conflicting DNA and confession evidence. Although participants were overall more likely to render a verdict in line with the DNA evidence, the authors also observed that perceptions of culpability and the proportion of guilty verdicts rose significantly when the prosecution presented a theory to explain the contradicting exculpatory DNA evidence (e.g., the DNA evidence only indicated that the defendant had not ejaculated).

Unfortunately, the phenomenon of false confessions is by no means a rare occurrence. Depending on particular definitions and methodology[2], scholars estimate that false confessions contribute to 12–26% of wrongful convictions (Gudjonsson, 2003; Innocence Project, 2021; National Registry of Exonerations, 2020). Further research suggests that 73–81% of individuals who falsely confess are eventually convicted of the crime (Leo and Ofshe, 1998; Drizin and Leo, 2004). This type of evidence may be so problematic partly because jurors are unable to distinguish between true and false confessions (Kassin et al., 2005; Levine et al., 2010) and are unreceptive to the idea that

___

[2]The National Registry of Exonerations defines a false confession as "a statement made to law enforcement at any point during the proceedings which was interpreted or presented by law enforcement as an admission of participation in or presence at the crime, even if the statement was not presented at trial. A statement is not a confession if it was made to someone other than law enforcement. A statement that is not at odds with the defense is not a confession. A guilty plea is not a confession." In comparison, the Innocence Project's definition includes false admissions and other forms of self-incrimination in their definition of false confessions.

an innocent person would ever falsely admit to a crime (Leo and Davis, 2010; Blandon-Gitlin et al., 2011). In an attempt to safeguard against the serious implications of a false confession, some states have allowed expert witnesses to testify about the science concerning this type of evidence (Kassin, 2008; Fulero, 2010).

Expert Testimony

A small number of studies have investigated the effect that expert testimony has on jurors' perceptions of recanted confessions. Jurors themselves report that such testimony would assist in their evaluation of this form of evidence (e.g., Chojnacki et al., 2008; Costanzo et al., 2010). However, findings concerning expert testimony's actual influence on verdicts in confession trials are mixed (Blandon-Gitlin et al., 2011; Gomes et al., 2016; Henderson and Levett, 2016). For instance, a number of studies have observed no differences in verdicts between jurors who have and have not been presented with expert testimony in mock homicide trials (Moffa and Platania, 2007; Henderson and Levett, 2016, Study 2; Jones and Penrod, 2016). In comparison, Blandon-Gitlin et al. (2011) found that participants who read a trial transcript involving a recanted confession were less likely to render a guilty verdict after being exposed to expert testimony.

If working as intended, expert testimony should sensitize jurors to the quality of the confession (Cutler et al., 1989; Levett and Kovera, 2008), thus leading to fewer convictions when the confession is low quality (e.g., when several interrogation tactics known to elicit false confession have been used). However, rather than sensitizing jurors to variations in the quality of confession evidence, expert testimony may instead induce a general skepticism concerning confessions. For example, Woody and Forrest (2009) found that jurors presented with expert testimony concerning false confessions were significantly less likely to convict the defendant, regardless of whether a false evidence ploy was used in the interrogation. In a similar study involving mock jurors, Woestehoff and Meissner (2016, Study 3) manipulated the pressure and number of coercive interrogation tactics used in a defendant's confession, as well as the presence of expert testimony. Overall, jurors were less likely to convict the defendant when exposed to expert testimony. This effect held regardless of how much pressure was involved in the defendant's interrogation. Interestingly, Woestehoff and Meissner (2016) observed an effect of interrogation pressure independent of the expert witness; jurors given the high and medium-pressure conditions were significantly less likely to convict as compared to jurors who read about a low-pressure confession. These participants therefore appeared to be sensitive to variations in the confession evidence's strength without the help of expert testimony.

Overall, there are conflicting findings regarding the effectiveness of expert testimony in cases involving a recanted confession. One factor that may explain these contradictory results is the race of the defendant. As described below, previous research has suggested that defendant race may influence jurors' perceptions of the confession itself (Ratcliff et al., 2010; Pickel et al., 2013).

Race in the Criminal Courtroom

An abundance of research examining the influence of defendant race has found that jurors often discriminate against defendants belonging to a racial minority (e.g., ForsterLee et al., 2006; Struckman-Johnson et al., 2008; Pickel et al., 2013). Specifically, jurors' perceptions of confession evidence appear to depend upon the defendant's race (Ratcliff et al., 2010; Pickel et al., 2013). Pickel et al. (2013) presented mock jurors with a confession video wherein the defendant's race was ambiguous. Jurors who were told the defendant was Arab American rated the confession as more voluntary and were more likely to convict than those who believed the defendant was White. In similar research, Ratcliff et al. (2010) found that participants shown a video of a confession believed that the confession was more voluntary, and that the suspect was more likely to be guilty, if the suspect was Asian or Black, as opposed to White.

Several studies have further observed the existence of an overall similarity-leniency bias within the courtroom, such that jurors perceive defendants of the same race more favorably than other-race defendants (see Mitchell et al., 2005; Devine and Caughlin, 2014 for meta-analyses). This similarity-leniency bias may be explained by social identity theory (SIT; Tajfel and Turner, 1986), which argues that people have a motivation to favor and prefer individuals belonging to their groups (rather than those outside of their groups) as a method of promoting a positive self-concept. In a criminal trial, social identity theory would therefore predict that jurors are more likely to acquit a defendant of the same race (and more likely to convict a defendant of a different race).

However, the watchdog hypothesis (Sargent and Bradfield, 2004) suggests that White jurors are motivated to protect against discrimination and thus pay more attention to legally relevant information when the defendant is Black. Using simulated vignettes describing a robbery trial, Sargent and Bradfield (2004) manipulated alibi evidence strength (Study 1), cross-examination effectiveness (Study 2), and defendant race, in two samples of White mock jurors. They found that White jurors were more sensitive to manipulations of alibi strength and cross-examination effectiveness when the defendant was Black as opposed to White. The authors argued that White jurors may have attended to this information more closely in an effort to be "watchdogs" against racism. In trials involving recanted confessions, jurors may therefore be more receptive to expert testimony (concerning the phenomenon of false confessions) when the defendant is Black as compared to White.

Current Study

Previous research has observed conflicting findings concerning the effect of expert testimony in trials involving recanted confessions. Furthermore, although research examining juror and defendant race has demonstrated a similarity-leniency bias (see Devine and Caughlin, 2014), Sargent and Bradfield (2004) suggest that White jurors may pay more attention to legally relevant evidence when the defendant is Black. The current study

therefore aimed to examine the interactive effects of defendant race and expert testimony on jurors' perceptions of recanted confessions. Drawing upon previous research, we developed two hypotheses.

> **Hypothesis 1:** Based on previous literature suggesting that jurors perceive confessions to be less voluntary for White defendants (e.g., Ratcliff et al., 2010), and other research demonstrating an outgroup bias in verdict decisions (e.g., Devine and Caughlin, 2014), we predicted a main effect for defendant race such that [predominantly White (United States Census Bureau, 2019)] participants would be more likely to convict the Black defendant than his White counterpart.

> **Hypothesis 2:** However, in accordance with the watchdog hypothesis (Sargent and Bradfield, 2004), we predicted an interaction between defendant race and expert testimony for White participants. Specifically, White participants would render fewer convictions in conditions with expert testimony as compared to conditions with no expert testimony for the Black defendant (with no such effect for the White defendant), as this testimony is legally relevant information to which they could attend to be "watchdogs" against racism.

MATERIALS AND METHODS

Participants

Research has demonstrated that crowd sourced samples can be more heterogenous as compared to traditional undergraduate college samples (Paolacci et al., 2010; Paolacci and Chandler, 2014; Baker et al., 2016). We therefore recruited participants using Amazon's Mechanical Turk (MTurk). We compensated participants with $3 for successfully completing the study. Although we had 1133 responses to our task, one participant did not give informed consent, 248 participants failed manipulation/attention checks[3], 235 participants were ineligible for jury duty in the United States, and 58 participants quit the survey prior to completion. Our final sample therefore consisted of 591 jury-eligible community members (i.e., citizens of the United States who were at least 18 years old with no unpardoned felony conviction). Three-hundred and thirty-one (55.4%) of the participants were women, 263 (44.1%) were men, and three (0.5%) identified as another gender. Participants' ages ranged from 19 to 69 years old ($M = 36$). Four-hundred and eighty-seven (81.6%) of the participants were White, 51 (8.5%) were Black, 27 (4.5%) were Hispanic, 21 (3.5%) were Asian, four (0.7%) were Native American, and seven (1.2%) identified as another race. Our participants' racial demographics are similar to what other researchers have observed using MTurk

---

[3]Sixty-two participants failed the manipulation check concerning defendant race, while 184 participants failed the manipulation check concerning the content of the expert's testimony. For attention checks, participants were instructed to select a particular response. Twenty-six participants failed the first attention check, 29 participants failed the second attention check, and 33 failed the third attention check.

(e.g., Burnham et al., 2018), and are comparable to the general United States population (United States Census Bureau, 2019), although our sample contained a slightly lower percentage of individuals identifying as Black (8.5 vs. 13.4%).

Materials

Screening/Demographic Questionnaire

We used a demographic questionnaire in order to screen participants to ensure that they were jury-eligible. Participants were also asked to provide information regarding their race and gender.

Trial Transcript

We used a trial transcript adapted from previous research (Kassin and Sommers, 1997; Sommers and Kassin, 2001; Henkel, 2008). The transcript involved a defendant charged with murdering his wife and neighbor. The prosecution argued that the defendant had arrived home to find his wife and neighbor together, and believing they were having an affair, he killed them in an act of jealousy. However, the defendant claimed that his wife and neighbor were already dead when he came home. The defendant had initially confessed to the crime, but later recanted the confession. Apart from this confession, the remaining evidence was circumstantial (e.g., a witness saw someone fleeing the crime scene who matched the general physical description of the defendant). The defendant testified that during his interrogation he was handcuffed to a desk in a small room for more than 5 h and claimed the interrogating officer had physically threatened him with his service weapon. The defendant also stated that he was experiencing an immense amount of stress and in a state of shock during the interrogation because he had learned of his wife's death only hours before. Finally, the defendant testified that the interrogating officer had repeatedly told him that his actions (killing his cheating spouse and her lover) were understandable, and that no one would blame him for what he did (i.e., minimization, Kassin and McNall, 1991). In each transcript, we manipulated the defendant's race (Black, White) and presence of expert testimony (present, not present). Defendant race was manipulated by including a color photograph [matched in a pilot study ($N = 30$) on perceived age, likeability, and attractiveness] of the defendant, along with varying his name (Charles Smith for the White defendant and Jamaal Washington for the Black defendant) to strengthen our race manipulation. Previous research has used names to manipulate race (e.g., Bertrand and Mullainathan, 2004; Widner and Chicoine, 2011; Alhabash et al., 2014), as names can reinforce racial stereotypes and elicit biased judgments (Bodenhausen and Wyer, 1985; Watson et al., 2011; Garcia and Abascal, 2016). In half of the transcripts, an expert witness specializing in confession research testified for the Defense. The expert primarily testified about two situational factors – minimization techniques and extended periods of time – that increase the likelihood of a false confession, both of which he noted were present in the defendant's confession and interrogation. The expert also discussed independent knowledge of the crime (underscoring the fact that the defendant's confession did not include details that only the true perpetrator of the crime would know), as

well as the prevalence of wrongful convictions that involve a false confession.

Jury Instructions

Before and after the transcript, we provided participants with juror instructions adapted from the California Criminal Jury Instructions (Judicial Council of California Civil Jury Instructions, 2020). The instructions discussed the criteria for first-degree murder, as well as the lesser-included second-degree murder and voluntary manslaughter charges, and also informed participants about the burden of proof and reasonable doubt.

Juror Questionnaire

In accordance with the legal instructions, we first asked participants to render a dichotomous verdict concerning the first-degree murder charge (guilty/not guilty). Participants who selected not guilty were then asked to render a dichotomous verdict concerning a second-degree murder charge (guilty/not guilty). Participants who still selected not guilty were then finally asked to render a dichotomous verdict concerning a voluntary manslaughter charge (guilty/not guilty). Logistically, we felt this method most appropriately reflected how jurors decide verdicts in California, as a juror who renders a guilty verdict for first-degree murder would not need to vote on the lesser-included offenses. We also asked participants to indicate on a scale from 1 (*not at all*) to 9 (*very much*) the degree to which they felt the defendant's confession was voluntary ("How voluntary was the defendant's confession?").

The questionnaire also included a manipulation check, which asked participants to identify the race of the defendant from a list of options. In conditions with expert testimony, we asked participants to identify what the expert witness testified about ("What was a factor that the false confession expert, Dr. Turner, discussed?") from a list of options to demonstrate that they had attended to this material. We embedded three other attention checks that required participants to select a specific response (e.g., "This is an attention check. Select Strongly Agree.").

Procedure

Participants were recruited from MTurk and completed the study online using Qualtrics survey software. Once participants had given informed consent, they were screened to ensure they met jury-eligibility requirements. We randomly assigned eligible participants to one of the four trial transcripts. Before and after reading the transcript, participants were provided with relevant legal instructions. After reading the transcript, participants responded to the juror questionnaire. Upon completion, participants were thanked, debriefed, and compensated. Participation in the study lasted approximately 30–45 min.

RESULTS

Verdicts

Tables 1–3 display a breakdown of verdicts by condition for the first-degree murder, second-degree murder, and voluntary

TABLE 1 | First-degree murder verdicts by defendant race and expert testimony.

Defendant race	Expert testimony	Verdict	Frequency	Percent
White	No Expert Testimony	Not guilty	113	72.9%
		Guilty	42	27.1%
		Total	155	100.0%
	Expert Testimony	Not guilty	105	72.9%
		Guilty	39	27.1%
		Total	144	100.0%
Black	No Expert Testimony	Not guilty	96	68.1%
		Guilty	45	31.9%
		Total	141	100.0%
	Expert Testimony	Not guilty	120	79.5%
		Guilty	31	20.5%
		Total	151	100.0%

TABLE 2 | Second-degree murder verdicts by defendant race and expert testimony.

Defendant race	Expert testimony	Verdict	Frequency	Percent
White	No Expert Testimony	Not guilty	75	66.4%
		Guilty	38	33.6%
		Total	113	100.0%
	Expert Testimony	Not guilty	77	73.3%
		Guilty	28	26.7%
		Total	105	100.0%
Black	No Expert Testimony	Not guilty	68	71.6%
		Guilty	27	28.4%
		Total	95	100.0%
	Expert Testimony	Not guilty	89	74.2%
		Guilty	31	25.8%
		Total	120	100.0%

TABLE 3 | Voluntary manslaughter verdicts by defendant race and expert testimony.

Defendant race	Expert testimony	Verdict	Frequency	Percent
White	No Expert Testimony	Not guilty	60	80.0%
		Guilty	15	20.0%
		Total	75	100.0%
	Expert Testimony	Not guilty	51	66.2%
		Guilty	26	33.8%
		Total	77	100.0%
Black	No Expert Testimony	Not guilty	52	76.5%
		Guilty	16	23.5%
		Total	68	100.0%
	Expert Testimony	Not guilty	70	78.7%
		Guilty	19	21.3%
		Total	89	100.0%

manslaughter charges, respectively. **Table 4** summarizes the percentage of not guilty verdicts by verdict option for all participants as well as for White participants only. Prior to running the regression, we indicator coded our predictor variables (0 = no expert, 1 = expert present, and 0 = White defendant, 1 = Black defendant, respectively). We coded our ordinal outcome variable as 0 = not guilty, 1 = guilty of

TABLE 4 | Percentage of not guilty verdicts by charge.

Charge	% of Not guilty verdicts	
	Full sample	White participants
First-degree murder	72.6%	74.5%
Second-degree murder	52.0%	53.3%
Voluntary manslaughter	39.3%	40.5%

manslaughter, 2 = guilty of second-degree murder, and 3 = guilty of first-degree murder. To test our hypotheses, we conducted an ordinal regression with verdict being regressed on expert testimony, defendant race, and the interaction between expert testimony and defendant race. Results revealed no significant main effect of expert testimony, $b = -0.04$, OR = 0.96, 95% CI [0.64, 1.45], $W^2(1, N = 591) = 0.03$, $p = 0.861$, or defendant race, $b = -0.12$, OR = 0.89, 95% CI [0.59, 1.34], $W^2(1, N = 591) = 0.33$ $p = 0.568$. Additionally, the expert testimony by defendant race interaction was non-significant, $b = 0.56$, OR = 1.75, 95% CI [0.97, 3.15], $W(1, N = 591) = 3.48$, $p = 0.062$.

Because the watchdog hypothesis specifically involves White individuals (Sargent and Bradfield, 2004), we re-ran our initial regression using only White participants ($N = 482$). As before, we conducted an ordinal regression on verdict using expert testimony, defendant race, and the interaction as the predictor variables. Again, we observed no significant main effects of expert testimony, $b = -0.26$, OR = 0.77, 95% CI [0.49, 1.21], $W^2(1, N = 482) = 1.29$, $p = 0.257$, or defendant race, $b = -0.35$, OR = 0.70, 95% CI [0.44, 1.11], $W^2(1, N = 482) = 2.26$, $p = 0.133$. However, there was a significant interaction between expert testimony and defendant race, $b = 0.93$, OR = 2.56, 95% CI [1.33, 4.92], $W^2(1, N = 482) = 7.9$, $p = 0.005$. We probed this interaction for White participants by running two separate ordinal regressions with expert testimony as the predictor, splitting the data file based on defendant race. Analyses indicated that for White jurors, there was no effect of expert testimony when the defendant was White, $b = -0.26$, OR = 0.77, 95% CI [0.49, 1.21], $W^2(1, N = 245) = 1.28$, $p = 0.258$. In comparison, we observed a significant effect of expert testimony when the defendant was Black, $b = 0.68$, OR = 1.98, 95% CI [1.23, 3.17], $W^2(1, N = 237) = 8.04$, $p = 0.005$. The odds of White jurors rendering a not guilty verdict (versus other verdict options) for the Black defendant were approximately twice as high when given expert testimony as compared to when no such testimony was presented.

Using the hmisc package (Harrell, 2021) in R (R Core Team, 2021), we conducted a *post hoc* sensitivity analysis to provide an estimate of the smallest effect size that we would have sufficient power (i.e., 80%) to detect. Analysis indicated that for an overall N of 590, our design had a power of .80 to detect an odds ratio of 1.52, which is equivalent to a "small" effect size under Cohen's conventions (see Chen et al., 2010). Therefore, we appeared to be sufficiently powered to conduct our ordinal analyses.

Voluntariness of Confession

We conducted an exploratory analysis on participants' perceptions of how voluntary the defendant's confession

was. We were interested in examining effects on voluntariness in particular because we felt this was a purer measure of jurors' perceptions of the confession itself. In comparison, participants' final verdicts could be influenced by a number of factors unrelated to the confession evidence (e.g., the circumstantial evidence presented at trial).

Overall, participants scored near the midpoint on their perceived voluntariness of the defendant's confession ($M = 4.71$, $SD = 2.57$). We ran a 2×2 ANOVA to test the degree to which defendant race and expert testimony influenced this rating. Results revealed a significant main effect for expert testimony [$F(1,586) = 4.76$, $p = 0.03$, $\eta^2_p = 0.008$, $w^2_p = 0.006$]; participants who received expert testimony perceived the defendant's confession to be less voluntary ($M = 4.48$, $SD = 2.50$) than those who did not ($M = 4.94$, $SD = 2.62$). The main effect for defendant race was not significant [$F(1,586) = 0.14$, $p = 0.71$, $partial\eta^2 = 0.001$ $w^2_p \leq 0.001$], nor was the interaction [$F(1,586) = 2.17$, $p = 0.14$, $\eta^2_p = 0.004$, $w^2_p = 0.002$].

As above, we re-ran this analysis with only White participants. This test again revealed a small, significant main effect for expert testimony [$F(1,477) = 4.00$, $p = 0.046$, $\eta^2_p = 0.008$, $w^2_p = 0.006$], qualified by a significant interaction between defendant race and expert testimony [$F(1,477) = 4.20$, $p = 0.041$, $\eta^2_p = 0.009$, $w^2_p = 0.007$]; the main effect for defendant race was not significant [$F(1,477) = 0.46$, $p = 0.50$, $partial\eta^2 = 0.001$, $w^2_p \leq 0.001$]. To probe the interaction, we first compared the effects of expert testimony on voluntariness ratings by defendant race. Simple effects tests demonstrated that for those who read about a White defendant, voluntariness ratings did not differ significantly in the presence ($M = 4.61$, $SD = 2.45$) or absence ($M = 4.60$, $SD = 2.66$) of expert testimony, $t(242) = -0.04$, $p = 0.98$, $d = 0.01$, 95% CI [$-0.25, 0.24$]. However, participants who read about a Black defendant were significantly less likely to perceive his confession as voluntary when they received expert testimony ($M = 4.29$, $SD = 2.52$) as compared to when they did not ($M = 5.23$, $SD = 2.63$), $t(235) = 2.83$, $p = 0.005$, $d = 0.37$, 95% CI [$0.11, 0.63$]. See **Figure 1** for a visual display of this relationship. When probing the interaction the other way, we did not find significant effects for defendant race in either the expert testimony present {$t(240) = 1.01$, $p = 0.315$, $d = 0.13$, 95% CI [$-0.12, 0.38$]} or absent {$t(237) = -1.86$, $p = 0.064$, $d = 0.24$, 95% CI [$-0.50, 0.01$]} conditions.

DISCUSSION

The aim of this study was to explore the combined effects of defendant race and expert testimony on jurors' decision-making in trials involving a recanted confession. In line with the watchdog hypothesis (Sargent and Bradfield, 2004), when the defendant was Black, White jurors were significantly more likely to render a not guilty verdict when given expert testimony. In comparison, White jurors' verdicts were not significantly influenced by expert testimony in conditions involving White defendants. This same pattern was found for perceptions of the confession's voluntariness, although findings relating to the voluntariness measure require further confirmatory testing from future research.

Our results may demonstrate evidence of a sensitivity effect in situations involving a Black defendant and a confession. Given that jurors in the Black defendant condition convicted less often and perceived the confession as less voluntary, these jurors likely attended more to the expert testimony than did those in the White defendant condition. Doing so may have allowed the testimony to sensitize jurors to the issues related to the disputed confession (e.g., long period of time, minimization and maximization techniques employed, lack of independent knowledge of the crime, etc.). However, because we did not manipulate the strength of the confession, it is unclear whether the expert testimony truly sensitized jurors in these conditions. Instead, they may have simply become skeptical of all confession evidence. Given that earlier findings regarding expert testimony and confession evidence suggest that the mere presence of expert testimony (regardless of the presence of coercive interrogation tactics) could reduce reliance on confessions (Woody and Forrest, 2009), this was an important first step in establishing the presence of a watchdog effect. Future research should extend the current study's design with the inclusion of a confession quality (i.e., lower vs. higher number of coercive tactics present in the interrogation) manipulation. Doing so would allow for a better understanding of whether sensitivity or skepticism is occurring in these situations.

Because our sample was predominantly White, we predicted an overall similarity-leniency bias such that Black defendants would be more likely to be convicted than defendants who were White (Devine and Caughlin, 2014). Contrary to predictions,
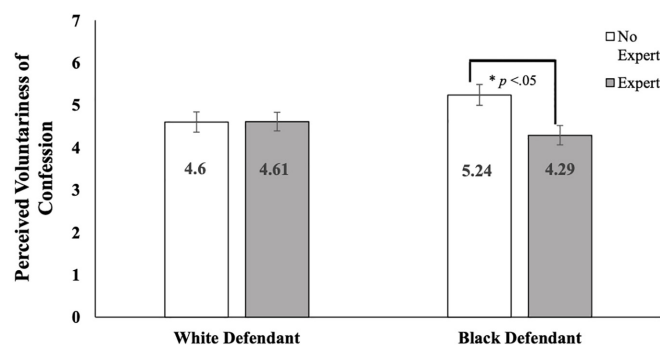

FIGURE 1 | Ratings of voluntariness of confession by defendant race and expert testimony for White participants.

there was no significant main effect of defendant race on jurors' verdicts. Although this finding conflicts with research supporting the similarity-leniency hypothesis (e.g., Ugwuegbu, 1978; Sommers and Ellsworth, 2000; Devine and Caughlin, 2014), other research has also demonstrated null effects relating to defendant race (e.g., Braden-Maguire et al., 2005; Maeder et al., 2012; Yamamoto and Maeder, 2017).

There are a number of potential explanations for why we observed no significant overall effect of defendant race. According to the aversive racism framework (Schweitzer and Nuñez, 2018), the influence of racial bias is greatest in ambiguous situations (Dovidio and Gaertner, 1996 2000, 2004). In the current study, the legal instructions likely removed some of the ambiguity concerning participants' verdict decision, lessening the effect defendant race may have had (Pfeifer and Ogloff, 1991). Furthermore, a recent meta-analysis observed out-group bias in studies involving property crimes or adult sexual assaults, but much smaller (or non-existent) effects in studies using violent cases (Devine and Caughlin, 2014). Because our trial transcript involved a murder, future research should consider replicating the current study using other crimes, such as burglary.

Finally, our data were collected between June and December 2018. During this time, the police's unjust treatment of Black individuals became a salient topic in the media (e.g., Carney, 2016; Lopez, 2018; Scott, 2018). Our participants may therefore have been cognizant of the potential for such discrimination, particularly because the defendant claimed that he was threatened and coerced by police during his interrogation. Previous research has found that White jurors' bias against BIPOC defendants is minimized when racial issues are made salient during the trial (e.g., Cohn et al., 2009; Bucolo and Cohn, 2010). It is a strong possibility that cases involving alleged police misconduct are inherently race salient, leading jurors to correct for racial bias and resulting in a null effect of defendant race (Sommers and Ellsworth, 2000, 2001). It is important to note that following the death of George Floyd in May 2020, the topic of racial discrimination in the United States' justice system received unprecedented attention and media coverage. We encourage researchers to replicate and extend these findings to see what effect these recent events may have had in this context.

Similar to our results concerning defendant race, we observed no significant main effect of expert testimony on jurors' verdicts. This complements the work of Jones and Penrod (2016), as well as Moffa and Platania (2007), but contradicts a number of other studies that did observe an effect of expert testimony in trials involving recanted confessions (Woody and Forrest, 2009; Blandon-Gitlin et al., 2011; Woestehoff and Meissner, 2016). Related research concerning jurors' perceptions of secondary confessions has also observed no significant effect of expert testimony on verdicts (Neuschatz et al., 2012; Maeder and Pica, 2014).

In comparison to our results concerning expert testimony and verdict, there was a significant main effect of expert testimony on perceived voluntariness of the confession. One explanation for this pattern may be that although expert testimony lowered jurors' perceived voluntariness of the confession, they still viewed the confession itself as indicative of guilt. Researchers

have used the fundamental attribution error to explain jurors' reluctance to discount disputed confession (e.g., Kassin and Sukel, 1997; Kassin and Gudjonsson, 2005). In our study, jurors may indeed have perceived the confession as less voluntary following expert testimony, but they still may have believed that overall, the defendant confessed because he was guilty (rather than because of the situational factors present). In similar research, Kassin and Wrightsman (1981) found instructions on the unreliability of coerced confessions significantly decreased participants' perceived voluntariness of the confession, but did not influence verdicts.

In their work, Sargent and Bradfield (2004) manipulated the strength of the defendant's alibi as well as the strength of the prosecutor's cross-examination[4]; future research should continue to examine the watchdog hypothesis by manipulating other types of evidence and/or expert testimony (such as expert testimony concerning police use of force or eyewitness identifications). Because the watchdog effect has now been demonstrated using both direct evidence (i.e., defendant's alibi) as well as trial-level phenomena (i.e., cross-examination and expert testimony), we tentatively predict that our observed effects would likely generalize to these other forms of evidence. Further, our results underscore the notion that there is a complex effect of race in the courtroom that goes beyond a simple similarity-leniency effect; we found White jurors to be more lenient to the racial outgroup when given expert testimony. As discussed above, it may be the case that, due to the increased public attention regarding racial discrimination in the legal system, the similarity-leniency effect is minimized (or outright reduced) in trials involving potential police misconduct. More work examining this issue, particularly sampling from BIPOC jurors, is needed to better understand these complexities. Based on these preliminary results, White jurors appear to either interpret or apply evidence differently depending upon the defendant's race, ultimately leading to different verdict decisions. Specifically, our findings suggest that attorneys should particularly consider the use of expert testimony in trials involving a BIPOC defendant and a recanted confession.

Finally, although we found evidence to suggest a watchdog effect, there are other possible explanations for our findings. In our study, White participants may have been more likely to use expert testimony in their verdict decisions when the defendant is Black as opposed to White not because they are paying more attention to legally relevant factors (as per the watchdog hypothesis), but because they are looking for a reason to acquit the Black defendant. This may be in an attempt to establish non-racist credentials (e.g., Effron and Conway, 2015) – when evaluating a Black defendant, White participants may feel as though their moral standing is uncertain, and so make greater use of the expert testimony and subsequently acquit in order to demonstrate their egalitarianism. Future research could implement a detailed measure of comprehension of the expert's testimony. This would reveal whether participants comprehend

---

[4]In the "weak" cross-examination condition, participants read a case summary where the court reporters indicated the prosecution had presented ineffective cross-examinations of the defense witnesses. In the "strong" condition, the case summary described the prosecution as presenting an effective cross-examination that diminished the defense witnesses' credibility.

the information better when the defendant is Black, or whether they simply use the expert testimony as a reason to acquit the Black defendant.

Limitations

Our study's methodology had a number of limitations. To begin, we used a written trial transcript, which limited ecological validity (Wenger and Bornstein, 2006). However, existing literature suggests that presentation mode does not significantly affect mock jurors' verdict decisions (Bornstein, 1999; Pezdek et al., 2010). Furthermore, our participants were likely aware that their responses had no true consequences, which may have influenced our findings (Bornstein and McCabe, 2005; Bornstein et al., 2017). Studying real jury trials would help to overcome this problem of consequentiality and may have led to different results. However, such a methodology would also introduce a host of additional issues regarding feasibility and internal validity. Like most jury research, we also only used a single trial transcript that had specific evidence and charges. Replications using additional cases would increase the generalizability of our results.

An additional ecological limitation of the current study is the lack of a deliberation component. Although research has demonstrated that the jury's final verdict is often predicted from individual verdicts (Kalven and Zeisel, 1966; Devine et al., 2007), other literature suggests that deliberation can influence jurors' bias (London and Nunez, 2000) and also affect jurors' cognitive processes when trying to reach a decision (Salerno and Diamond, 2010; Salerno et al., 2017). Sommers (2006) has further demonstrated that the racial composition of a jury can influence how jurors talk about race, which may be relevant to our results as our study involved a Black defendant in half of the conditions. Therefore, future research examining the watchdog hypothesis should likely involve a deliberation component.

Because the study was conducted online on MTurk, there was a lack of general control over the environment in which participants responded, which may have produced environmental confounds. However, we implemented manipulation and attention checks to ascertain data quality (e.g., Peer et al., 2014). Using MTurk also allowed for recruitment from a nationwide community sample (rather than relying on an undergraduate sample from a single university), which likely increased the generalizability of our results (Baker et al., 2016). Regardless, we still had a fairly racially homogenous sample, as response rates from BIPOC participants were low. We were therefore unable to do any proper comparisons based on juror race. Although the watchdog hypothesis focuses specifically on White jurors, future research needs to be conducted that explicitly examines BIPOC jurors' perceptions in the courtroom.

CONCLUSION

Our study examined the role of defendant race and the influence of expert testimony in the context of trials involving recanted confessions. To the best of our knowledge, this is the first study to examine the interactive effects of these variables. For White jurors, we observed an interaction between defendant race and the presence of expert testimony. There was no significant effect of expert testimony on verdict when the defendant was White, but White jurors were significantly less likely to find the Black defendant guilty (and perceive his confession as voluntary) when presented with expert testimony concerning false confessions. These findings support the existence of the watchdog hypothesis (Sargent and Bradfield, 2004), such that White jurors are more receptive to legally relevant evidence when the defendant is Black. To gain a stronger understanding of when this effect is elicited, future research should replicate the current study using other types of evidence and expert testimony.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Carleton University Research Ethics Board-B (CUREB-B) Carleton University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LE and EM contributed to the conception and design of the study and conducted the data analysis. LE organized the data collection and wrote the first draft of the manuscript. EM wrote the sections of the manuscript. Both authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

REFERENCES

Alhabash, S., Hales, K., Baek, J. H., and Oh, H. J. (2014). Effects of race, visual anonymity, and social category salience on online dating outcomes. *Comput. Hum. Behav.* 35, 22–32. doi: 10.1016/j.chb.2014.02.016

Appleby, S. C., and Kassin, S. M. (2016). When self-report trumps science: effects of confessions, DNA, and prosecutorial theories on perceptions of guilt. *Psychol. Public Policy Law* 22, 127–140. doi: 10.1037/law0000080

Baker, M. A., Fox, P., and Wingrove, T. (2016). Crowdsourcing as a forensic psychology research tool. *Am. J. For. Psychol.* 34, 37–50.

Bertrand, M., and Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *Am. Econ. Rev.* 94, 991–1013. doi: 10.1257/0002828042002561

Blandon-Gitlin, I., Sperry, K., and Leo, R. (2011). Jurors believe interrogation tactics are not likely to elicit false confessions: will expert witness testimony inform them otherwise? *Psychol. Crime Law* 17, 239–260. doi: 10.1080/10683160903113699

Bodenhausen, G. V., and Wyer, R. S. Jr. (1985). Effects of stereotypes on decision making and information-processing strategies. *J. Pers. Soc. Psychol.* 48, 267–282.

Bornstein, B. H. (1999). The ecological validity of jury simulations: is the jury still out? *Law Hum. Behav.* 23, 75–91. doi: 10.1023/A:1022326807441

Bornstein, B. H., Golding, J. M., Neuschatz, J., Kimbrough, C., Reed, K., Magyarics, C., et al. (2017). Mock juror sampling issues in jury simulation research: a meta-analysis. *Law Hum. Behav.* 41, 13–28. doi: 10.1037/lhb0000223

Bornstein, B. H., and McCabe, S. G. (2005). Jurors of the absurd? The role of consequentiality in jury simulation research. *Florida State Univ. Law Rev.* 32, 443–467.

Braden-Maguire, J., Sigal, J., and Perrino, C. S. (2005). Battered women who kill: variables affecting simulated jurors' verdicts. *J. Fam. Violence* 20, 403–408. doi: 10.1007/s10896-005-7801-0

Bucolo, D. O., and Cohn, E. S. (2010). Playing the race card: making race salient in defence opening and closing statements. *Legal Criminol. Psychol.* 15, 293–303. doi: 10.1348/135532508X400824

Burnham, M. J., Le, Y. K., and Piedmont, R. L. (2018). Who is Mturk? Personal characteristics and sample consistency of these online workers. *Ment. Health Relig. Culture* 21, 934–944. doi: 10.1080/13674676.2018.1486394

Carney, N. (2016). All lives matter, but so does race: black lives matter and the evolving role of social media. *Hum. Soc.* 40, 180–199. doi: 10.1177/0160597616643868

Chen, H., Cohen, P., and Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Commun. Stat. Simul. Comput.* 39, 860–864. doi: 10.1080/03610911003650383

Chojnacki, D. E., Cicchini, M. D., and White, L. T. (2008). An empirical basis for the admission of expert testimony on false confessions. *Arizona State Law J.* 40, 1–46.

Cohn, E. S., Bucolo, D., Pride, M., and Sommers, S. R. (2009). Reducing White juror bias: the role of race salience and racial attitudes. *J. Appl. Soc. Psychol.* 39, 1953–1973. doi: 10.1111/j.1559-1816.2009.00511.x

Costanzo, M., Shaked-Schroer, N., and Vinson, K. (2010). Juror beliefs about police interrogations, false confessions, and expert testimony. *J. Empir. Legal Stud.* 7, 231–247. doi: 10.1111/j.1740-1461.2010.01177.x

Cutler, B. L., Penrod, S. D., and Dexter, H. R. (1989). The eyewitness, the expert psychologist, and the jury. *Law Hum. Behav.* 13, 311–332. doi: 10.1007/BF01067032

Devine, D. J., Buddenbaum, J., Houp, S., Stolle, D. P., and Studebaker, N. (2007). Deliberation quality: a preliminary examination in criminal juries. *J. Empir. Legal Stud.* 4, 273–303. doi: 10.1111/j.1740-1461.2007.00089.x

Devine, D. J., and Caughlin, D. E. (2014). Do they matter? A meta-analytic investigation of individual characteristics and guilt judgments. *Psychol. Public Policy Law* 20, 109–134. doi: 10.2466/pr0.1997.81.2.679

Dovidio, J. F., and Gaertner, S. L. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychol. Sci.* 11, 315–319. doi: 10.1111/1467-9280.00262

Dovidio, J. F., and Gaertner, S. L. (1996). Affirmative action, unintentional racial biases, and intergroup relations. *J. Soc. Issue.* 52, 51–75. doi: 10.1111/j.1540-4560.1996.tb01848.x

Dovidio, J. F., and Gaertner, S. L. (2004). "Aversive racism," in *Advances in Experimental Social Psychology*, Vol. 36, ed. M. P. Zanna (San Diego, CA: Academic Press), 1–52.

Drizin, S., and Leo, R. (2004). The problem of false confessions in the post-DNA world. *N. Carolina Law Rev.* 82, 891–1007.

Duru, N. J. (2003). The Central Park Five, the Scottsboro Boys, and the myth of the bestial black man. *Cardozo Law Rev.* 25, 1315–1366.

Effron, D. A., and Conway, P. (2015). When virtue leads to villainy: advances in research on moral self-licensing. *Curr. Opin. Psychol.* 6, 32–35.

ForsterLee, R., ForsterLee, L., Horowitz, I. A., and King, E. (2006). The effects of defendant race, victim race, and juror gender on evidence processing in a murder trial. *Behav. Sci. Law* 24, 179–198. doi: 10.1002/bsl.675

Fulero, S. M. (2010). Admissibility of expert testimony based on the Grisso and Gudjonsson scales in disputed confession cases. *J. Psychiatry Law* 38, 193–214. doi: 10.1177/009318531003800109

Garcia, D., and Abascal, M. (2016). Colored perceptions: racially distinctive names and assessments of skin color. *Am. Behav. Sci.* 60, 420–441. doi: 10.1177/0002764215613395

Gomes, D. M., Stenstrom, D. M., and Calvillo, D. P. (2016). Examining the judicial decision to substitute credibility instructions for expert testimony on confessions. *Legal Criminol. Psychol.* 21, 319–331. doi: 10.1111/lcrp.12068

Gudjonsson, G. H. (2003). *The Psychology of Interrogations and Confessions: A Handbook*. Chichester: John Wiley & Sons.

Harrell, F. (2021). *HmiscL Version 4.5-0*.

Henderson, K. S., and Levett, L. M. (2016). Can expert testimony sensitize jurors to variations in confession evidence? *Law Hum. Behav.* 40, 638–649. doi: 10.1037/lhb0000204

Henkel, L. A. (2008). Jurors' reactions to recanted confessions: do the defendant's personal and dispositional characteristics play a role? *Psychol. Crime Law* 14, 565–578. doi: 10.1080/10683160801995247

Innocence Project (2021). *Research Resources*. Available online at: https://innocenceproject.org/research-resources/

Jones, A. M., and Penrod, S. (2016). Can expert testimony sensitize jurors to coercive interrogation tactics? *J. For. Psychol. Pract.* 16, 393–409. doi: 10.1080/15228932.2016.1232029

Judicial Council of California Civil Jury Instructions (2020). Available online at: https://www.courts.ca.gov/partners/juryinstructions.htm (accessed January, 2019).

Kalven, H. Jr., and Zeisel, H. (1966). *The American Jury*. Chicago: University of Chicago Press.

Kassin, S. M. (2008). False confessions: causes, consequences, and implications for reform. *Curr. Direct. Psychol. Sci.* 17, 249–253. doi: 10.1111/j.1467-8721.2008.00584.x

Kassin, S. M., Goldstein, C. C., and Savitsky, K. (2003). Behavioral confirmation in the interrogation room: on the dangers of presuming guilt. *Law Hum. Behav.* 27:187. doi: 10.1023/a:1022599230598

Kassin, S. M., and Gudjonsson, G. H. (2005). True crimes, false confessions. *Sci. Am. Mind* 16, 24–31.

Kassin, S. M., and Kiechel, K. L. (1996). The social psychology of false confessions: compliance, internalization, and confabulation. *Psychol. Sci.* 7, 125–128. doi: 10.1111/j.1467-9280.1996.tb00344.x

Kassin, S. M., and McNall, K. (1991). Police interrogations and confessions. *Law Hum. Behav.* 15, 233–251. doi: 10.1007/BF01061711

Kassin, S. M., Meissner, C. A., and Norwick, R. J. (2005). "I'd know a false confession if I saw one": a comparative study of college students and police investigators. *Law Hum. Behav.* 29:211. doi: 10.1007/s10979-005-2416-9

Kassin, S. M., and Neumann, K. (1997). On the power of confession evidence: an experimental test of the fundamental difference hypothesis. *Law Hum. Behav.* 21:469. doi: 10.1023/A:1024871622490

Kassin, S. M., and Sommers, S. R. (1997). Inadmissible testimony, instructions to disregard, and the jury: substantive versus procedural considerations. *Pers. Soc. Psychol. Bull.* 23, 1046–1054. doi: 10.1177/01461672972310005

Kassin, S. M., and Sukel, H. (1997). Coerced confessions and the jury: an experimental test of the "harmless error" rule. *Law Hum. Behav.* 21, 27–46. doi: 10.1023/A:1024814009769

Kassin, S. M., and Wrightsman, L. S. (1981). Coerced confessions, judicial instruction, and mock juror verdicts. *J. Appl. Soc. Psychol.* 11, 489–506. doi: 10.1111/j.1559-1816.1981.tb00838.x

King, L., and Snook, B. (2009). Peering inside a Canadian interrogation room: an examination of the Reid model of interrogation, influence tactics, and coercive strategies. *Crim. Just. Behav.* 36, 674–694. doi: 10.1177/0093854809335142

Leo, R., and Ofshe, R. (1998). The consequences of false confessions: deprivations of liberty and miscarriages of justice in the age of psychological interrogation. *J. Crim. Law Criminol.* 88, 429–496. doi: 10.2307/1144288

Leo, R. A., and Davis, D. (2010). From false confession to wrongful conviction: seven psychological processes. *J. Psychiatry Law* 38, 9–56. doi: 10.1177/009318531003800103

Levett, L. M., and Kovera, M. B. (2008). The effectiveness of opposing expert witnesses for educating jurors about unreliable expert evidence. *Law Hum. Behav.* 32, 363–374. doi: 10.1007/s10979-007-9113-9

Levine, T. R., Kim, R. K., and Blair, J. P. (2010). (In) accuracy at detecting true and false confessions and denials: an initial test of a projected motive model of

veracity judgments. *Hum. Commun. Res.* 36, 82–102. doi: 10.1111/j.1468-2958.2009.01369.x

Lieberman, J. D., Carrell, C. A., Miethe, T. D., and Krauss, D. A. (2008). Gold versus platinum. *Psychol. Public Policy Law* 14, 27–62. doi: 10.1037/1076-8971.14.1.27

Lopez, G. (2018). *There Are Huge Racial Disparities in How US Police Use Force.* Washington, DC: Vox Media.

London, K., and Nunez, N. (2000). The effect of jury deliberations on jurors' propensity to disregard inadmissible evidence. *J. Appl. Psychol.* 85, 932–939. doi: 10.1037/0021-9010.85.6.932

Maeder, E., Dempsey, J., and Pozzulo, J. (2012). Behind the veil of juror decision making: testing the effects of Muslim veils and defendant race in the courtroom. *Crim.Justice Behav.* 39, 666–678. doi: 10.1177/0093854812436478

Maeder, E. M., and Pica, E. (2014). Secondary confessions: the influence (or lack thereof) of incentive size and scientific expert testimony on jurors' perceptions of informant testimony. *Law Hum. Behav.* 38, 560–568. doi: 10.1037/lhb0000106

Mitchell, T. L., Haw, R. M., Pfeifer, J. E., and Meissner, C. A. (2005). Racial bias in mock juror decision-making: a meta-analytic review of defendant treatment. *Law Hum. Behav.* 29, 621–637. doi: 10.1007/s10979-005-8122-9

Moffa, M. S., and Platania, J. (2007). Effects of expert testimony and interrogation tactics on perceptions of confessions. *Psychol. Rep.* 100, 563–570. doi: 10.2466/pr0.100.2.563-570

National Registry of Exonerations (2020). *Exonerations in 2019.* Ann Arbor: University of Michigan Law School.

Neuschatz, J. S., Wilkinson, M. L., Goodsell, C. A., Wetmore, S. A., Quinlivan, D. S., and Jones, N. J. (2012). Secondary confessions, expert testimony, and unreliable testimony. *J. Police Crim. Psychol.* 27, 179–192. doi: 10.1007/s11896-012-9102-x

Paolacci, G., and Chandler, J. (2014). Inside the Turk: understanding Mechanical Turk as a participant pool. *Curr. Direct. Psychol. Sci.* 23, 184–188. doi: 10.1177/0963721414531598

Paolacci, G., Chandler, J., and Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment Decis. Making* 5, 411–419.

Pezdek, K., Avila-Mora, E., and Sperry, K. (2010). Does trial presentation medium matter in jury simulation research? Evaluating the effectiveness of eyewitness expert testimony. *Appl. Cogn. Psychol.* 24, 673–690. doi: 10.1002/acp.1578

Peer, E., Vosgerau, J., and Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behav. Res. Methods* 46, 1023–1031. doi: 10.3758/s13428-013-0434-y

Pfeifer, J. E., and Ogloff, J. R. (1991). Ambiguity and guilt determinations: a modern racism perspective 1. *J. Appl. Soc. Psychol.* 21, 1713–1725.

Pickel, K. L., Warner, T. C., Miller, T. J., and Barnes, Z. T. (2013). Conceptualizing defendants as minorities leads mock jurors to make biased evaluations in retracted confession cases. *Psychol. Public Policy Law* 19:56. doi: 10.1037/a0029308

R Core Team (2021). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.

Ratcliff, J. J., Lassiter, G. D., Jager, V. M., Lindberg, M. J., Elek, J. K., and Hasinski, A. E. (2010). The hidden consequences of racial salience in videotaped interrogations and confessions. *Psychol. Public Policy Law* 16:200. doi: 10.1037/a0018482

Redlich, A. D., Summers, A., and Hoover, S. (2010). Self-reported false confessions and false guilty pleas among offenders with mental illness. *Law Hum. Behav.* 34, 79–90. doi: 10.1007/s10979-009-9194-8

Salerno, J. M., Bottoms, B. L., and Peter-Hagene, L. C. (2017). Individual versus group decision making: jurors' reliance on central and peripheral information to evaluate expert testimony. *PLoS One* 12:e0183580. doi: 10.1371/journal.pone.0183580

Salerno, J. M., and Diamond, S. S. (2010). The promise of a cognitive perspective on jury deliberation. *Psychon. Bull. Rev.* 17, 174–179. doi: 10.3758/PBR.17.2.174

Sargent, M. J., and Bradfield, A. L. (2004). Race and information processing in criminal trials: does the defendant's race affect how the facts are evaluated? *Pers. Soc. Psychol. Bull.* 30, 995–1008. doi: 10.1177/0146167204265741

Scott, E. (2018). *Only One-Third of African Americans Say They Have Confidence in the Police. Killings Like Alton Sterling's are Part of the Reason.* Washington, DC: Washington Post.

Schweitzer, K., and Nuñez, N. (2018). What evidence matters to jurors? The prevalence and importance of different homicide trial evidence to mock jurors. *Psychiatry Psychol. Law* 25, 437–451. doi: 10.1080/13218719.2018.1437666

Sommers, S. R. (2006). On racial diversity and group decision making: identifying multiple effects of racial composition on jury deliberations. *J. Pers. Soc. Psychol.* 90, 597–612. doi: 10.1037/0022-3514.90.4.597

Sommers, S. R., and Ellsworth, P. C. (2000). Race in the courtroom: perceptions of guilt and dispositional attributions. *Pers. Soc. Psychol. Bull.* 26, 1367–1379. doi: 10.1177/0146167200263005

Sommers, S. R., and Ellsworth, P. C. (2001). White juror bias: an investigation of prejudice against Black defendants in the American courtroom. *Psychol. Public Policy Law* 7:201. doi: 10.1037/1076-8971.7.1.201

Sommers, S. R., and Kassin, S. M. (2001). On the many impacts of inadmissible testimony: selective compliance, need for cognition, and the overcorrection bias. *Pers. Soc. Psychol. Bull.* 27, 1368–1377. doi: 10.1177/01461672012710012

Struckman-Johnson, C., Miller, M. G., and Struckman-Johnson, D. (2008). Effects of Native American race, intoxication, and crime severity on judgments of guilt. *J. Appl. Soc. Psychol.* 38, 1981–1992. doi: 10.1111/j.1559-1816.2008.00376.x

Tajfel, H., and Turner, J. C. (1986). "The social identity theory of intergroup behavior," in *Psychology of Intergroup Relations*, eds S. Worchel and W. Austin (Chicago, IL: Nelson-Hall).

Ugwuegbu, D. C. E. (1978). Racial and evidential factors in juror attribution of legal responsibility. *J. Exp. Soc. Psychol.* 15, 133–146. doi: 10.1016/0022-1031(79)90025-8

United States Census Bureau (2019). *ACS Demographic and Housing Estimates.* Available online at: https://data.census.gov/

Watson, S., Appiah, O., and Thornton, C. G. (2011). The effect of name on pre-interview impressions and occupational stereotypes: the case of black sales job applicants. *J. Appl. Soc. Psychol.* 41, 2405–2420. doi: 10.1111/j.1559-1816.2011.00822.x

Wenger, A. A., and Bornstein, B. H. (2006). The effects of victim's substance use and relationship closeness on mock jurors' judgments in an acquaintance rape case. *Sex Roles* 54, 547–555. doi: 10.1007/s11199-006-9014-2

Widner, D., and Chicoine, S. (2011). It's all in the name: employment discrimination against Arab Americans. *Sociol. Forum* 26, 806–823. doi: 10.1111/j.1573-7861.2011.01285.x

Woestehoff, S. A., and Meissner, C. A. (2016). Juror sensitivity to false confession risk factors: dispositional vs. situational attributions for a confession. *Law Hum. Behav.* 40:564. doi: 10.1037/lhb0000201

Woody, W. D., and Forrest, K. D. (2009). Effects of false-evidence ploys and expert testimony on jurors' verdicts, recommended sentences, and perceptions of confession evidence. *Behav. Sci. Law* 27, 333–360.

Yamamoto, S., and Maeder, E. M. (2017). Defendant and juror race in a necessity case: an ultimate attribution error. *J. Ethn. Crim. Justice* 15, 270–284. doi: 10.1080/15377938.2017.1347542

The Importance of Perceived Procedural Justice Among Defendants With a Non-Western Background Involved in Dutch Criminal Cases

Lisa F. M. Ansems[1]*, Kees van den Bos[1,2] and Elaine Mak[1]

[1]School of Law, Utrecht University, Utrecht, Netherlands, [2]Department of Psychology, Utrecht University, Utrecht, Netherlands

This study aims to put perceived procedural justice to a critical test in the context of Dutch criminal court hearings. To that end, we surveyed 198 criminal defendants to examine whether their perceptions of procedural fairness were significantly associated with trust in judges and intentions to protest against judicial rulings, among other variables. We also examine the possibility that sometimes *un*fair procedures may have nice aspects, because they offer opportunities to attribute negative outcomes to external causes. Previous studies conducted in different settings support this line of reasoning by showing that associations between perceived procedural justice and other variables are sometimes attenuated or even reversed, particularly when people feel strongly evaluated. The current study takes these insights into the novel context of Dutch criminal court hearings by focusing on defendants with a non-Western ethnic-cultural background. Some of these defendants may feel negatively evaluated by society, which can manifest as a high level of perceived discrimination. Thus, we examine whether the associations between perceived procedural justice and important other variables may be attenuated or reversed depending on respondents' perceptions of everyday discrimination and their outcome judgments. Our results revealed significant associations between perceived procedural justice on the one hand and trust in judges and protest intentions on the other hand, which remained intact regardless of perceptions of everyday discrimination and outcome judgments. Hence, even in this real-life courtroom context, procedural justice was a relevant concern. Taken together, our findings support the importance of perceived procedural justice, even when it is put to a critical test.

Keywords: perceived procedural justice, outcome judgments, perceived everyday discrimination, attenuation, reversal, critical test, criminal defendants, external attributions

INTRODUCTION

Procedural Justice and the Fair Process Effect

Fair and just procedures are key aspects of law. Issues of procedural fairness, when viewed from a legal perspective, concern the extent to which legal procedures meet standards laid down in statutes, case law, and unwritten legal principles. In contrast, social psychologists empirically study the extent to which procedures correspond with citizens' ideas about fairness and justice. These experiences of being treated fairly by decision-making authorities are referred to as perceived procedural justice[1] (Lind and Tyler, 1988; Tyler and Lind, 1992).

When people perceive procedures as fair, they tend to be more satisfied with the outcomes of their cases and more inclined to accept those outcomes (Thibaut and Walker, 1975; Lind et al., 1993; Van den Bos et al., 2014). They also tend to report higher levels of self-esteem and trust in judges (Koper et al., 1993; Sedikides et al., 2008; Grootelaar, 2018). Other important attitudes and behaviors associated with perceived procedural justice are perceived legitimacy, cooperation with legal authorities, and compliance with the law (Paternoster et al., 1997; Tyler and Huo, 2002; Tyler, 2006). Such favorable responses to perceived procedural justice are generally referred to as the fair process effect (Folger et al., 1979; Van den Bos, 2015).

There are various explanations for people's concern with issues of procedural fairness (see, e.g., Van den Bos, 2005). For instance, in their pioneering research on this topic, Thibaut and Walker (1975) propose an instrumental explanation of the fair process effect by suggesting that people care about procedural justice because fair procedures are more likely to yield fair and favorable outcomes. Others argue that people care about procedural fairness for relational reasons, as being treated fairly communicates to them that they are valued members of society (e.g., Lind and Tyler, 1988; Tyler and Lind, 1992). Furthermore, information about procedural justice may help to make sense of uncertain situations. For instance, when people do not know whether or not they can trust authorities, they may look to whether or not these authorities treat them fairly and use this information as a heuristic substitute (e.g., Lind et al., 1993; Van den Bos and Lind, 2002).

The current study puts procedural justice and the fair process effect to a critical test by assessing whether the associations between procedural justice perceptions and important other variables hold in the real-life courtroom context of Dutch criminal cases. Contrary to the psychological laboratories which provide the research context of many procedural justice studies, criminal court hearings involve actual stakes, with defendants risking sanctions ranging from fines to community service and imprisonment. In addition, convictions can have serious consequences for defendants' positions on the job market.

Hence, one may wonder whether perceived procedural fairness is a relevant concern for defendants in these criminal cases, as they may be much more concerned with their case outcomes.

In an early critique of procedural justice research, Hayden and Anderson (1979; see also Anderson and Hayden, 1981) point to the use of simulation experiments, which necessarily involve a simplification of social situations, and the resulting limitations as to the conclusions that can be drawn from these studies. Others, too, note the importance of considering real-world complexities. For example, results from a study by Berrey et al. (2012) suggest that litigants involved in employment discrimination cases often did not distinguish between procedures and outcomes, and that they defined fairness in terms of whether or not the procedure benefited their own side. In line with this, Jenness and Calavita (2018) argue that, in their sample of incarcerated men in the United States, participants' concerns about procedural fairness were largely subordinate to (or even defined by) their outcome concerns.

In contrast, other studies found that perceptions of procedural fairness matter even to people involved in high stakes cases. For instance, Landis and Goodstein (1986) reported that inmates' perceptions of outcome fairness were associated with both procedural and outcome issues, but that procedural characteristics were dominant in this regard. In addition, Casper et al. (1988) showed that perceived procedural justice was significantly associated with multiple measures of outcome satisfaction among defendants in felony cases. Other findings also indicate the importance of perceived procedural justice in criminal justice contexts (e.g., Tyler, 1984, 1988, 2006; Paternoster et al., 1997; Tyler and Huo, 2002).

To shed further light on this issue, the current paper focuses on the context of real-life criminal court hearings involving actual stakes and examines whether perceptions of procedural fairness are associated with trust in judges and intentions to protest against the judicial ruling, among other variables. More specifically, as our first hypothesis we propose the following:

Defendants with higher levels of perceived procedural justice report more trust in judges, more positive outcome judgments, lower intentions to protest against their outcomes, and higher levels of state self-esteem (i.e., self-esteem at the moment of filling out the questionnaire; Hypothesis 1).

Attenuated or Reversed Fair Process Effects

Our study also adds to current insights into procedural justice in criminal justice contexts by examining the possibility that sometimes *un*fair procedures may have nice aspects (Van den Bos et al., 1999). As explained by Brockner et al. (2009, p. 185), these "reductions in people's desire for higher process fairness" may result in attenuated or even reversed fair process effects. That is, the associations between perceived procedural justice and relevant other variables may be weakened, possibly to the extent that they are no longer statistically significant, or reversed, such that people respond more favorably to perceived procedural unfairness than to perceived procedural fairness (Brockner et al., 2009).

---

[1]Previous research suggests that criminal defendants sometimes find the term "fair" easier to understand and use than "just" and has therefore treated these terms as synonyms (Ansems et al., 2020). Given the similarity in research context, the current paper uses the terms "justice" and "fairness" interchangeably as well.

One explanation for attenuated or reversed fair process effects is people's self-enhancement motive. That is, people often want to feel good about themselves (Leary and Terry, 2013). When people perceive procedures as fair, they are more likely to view themselves as personally responsible for their outcomes and thus make more internal attributions (Leung et al., 2001; Brockner et al., 2009). Unfavorable outcomes may then harm their self-esteem (Weiner, 1985). To protect their self-esteem, people may look for opportunities to attribute these negative outcomes to external causes (Cohen 1982). Procedures that people perceive as unfair offer such external attribution opportunities (Brockner et al., 2003). That is, people may maintain their self-esteem by attributing negative outcomes to the perceived unfairness of procedures rather than to themselves. Hence, for people who receive negative outcomes unfair procedures can have nice aspects, at least under some circumstances (Van den Bos et al., 1999).

This line of reasoning is supported by a number of empirical studies, which show that the positive association between perceived procedural justice on the one hand and self-esteem or related measures on the other hand may sometimes be attenuated or even reversed when outcomes are perceived as unfavorable in work contexts (e.g., Gilliland, 1994; Ployhart et al., 1999; Schroth and Shah, 2000; Brockner et al., 2008, 2009; Brockner, 2010). Some studies have found attenuated or reversed fair process effects when examining different kinds of dependent variables. Thus, possibly, people's reduced threat to sense of self due to perceived procedural unfairness (or, vice versa, their heightened threat to sense of self due to perceived procedural fairness) translates into other kinds of reactions. For instance, one study found an attenuated fair process effect on trust in judges among research participants with relatively high external attribution ratings (see Ansems, 2021). In addition, Holmvall and Bobocel (2008) found a reversed fair process effect on measures of perceived outcome fairness and outcome satisfaction. Furthermore, Van den Bos et al. (1999) found a reversed fair process effect on protest intentions in the face of unfavorable outcomes manipulated in laboratory experiments. In their study, the reversal was triggered by the strength of the evaluative context: Participants who felt strongly evaluated during the decision-making procedure reported lower protest intentions when they perceived the procedure as unfair rather than fair. Thus, feeling strongly evaluated can play an important role in attenuating or reversing the fair process effect.

Perceived Everyday Discrimination

To examine these issues, the current research focuses on perceived procedural justice among defendants with a non-Western ethnic-cultural background[2] involved in Dutch criminal cases. Some

of these defendants may feel negatively evaluated by society, which can manifest as a high level of perceived discrimination (Huijnk and Andriessen, 2016). We propose that, as a result, these defendants might respond differently to perceptions of procedural fairness during their court hearings (Van den Bos et al., 1999). Hence, we assess whether the associations between perceived procedural justice and relevant other variables may be attenuated or reversed depending on how much discrimination defendants experience in their daily lives.

In doing so, we also take into account research on discrimination which examines processes similar to the self-enhancement processes outlined earlier. Perceived discrimination can lead to various problems, including stress and reduced psychological well-being (Major et al., 2002). At the same time, experiencing discrimination may enable people to maintain their self-esteem, as it reduces their sense of personal responsibility and deservingness of negative outcomes (Crocker and Major, 1989; Major, 1994). Hence, attributing negative events to discrimination rather than one's personal qualities is a coping strategy people can use to counter the negative impact these events may otherwise have on their self-esteem (Major et al., 2002). Building on these insights, we formulate our second hypothesis as follows:

There is a two-way interaction between perceived procedural justice and perceived everyday discrimination, such that defendants who experience relatively high levels of everyday discrimination show attenuated or reversed associations between perceived procedural justice and our other variables (i.e., trust in judges, outcome judgments, protest intentions, and state self-esteem; Hypothesis 2).

Outcome Judgments

In addition to perceptions of everyday discrimination, we examine the potentially moderating role of defendants' outcome judgments (i.e., how positively or negatively they judge their outcomes). According to Brockner and Wiesenfeld (1996), receiving negative outcomes triggers sense-making processes. As a result, people may pay more attention to issues of procedural fairness. Thus, the fair process effect may be strengthened when people receive negative outcomes. At the same time, people may look for opportunities to attribute these negative outcomes to external causes in order to protect their self-esteem (Cohen 1982), as explained earlier. Because procedures that are perceived as unfair offer such opportunities, the fair process effect may be attenuated or reversed (Brockner et al., 2009).

Combining these possibilities, our third hypothesis suggests that the associations between perceived procedural justice and our other variables may be moderated by defendants' outcome judgments. When defendants judge their outcomes more negatively, this may either strengthen the associations between perceived procedural justice and our other variables (Brockner and Wiesenfeld, 1996) or, alternatively, attenuate or even reverse these associations (Brockner et al., 2009). Thus, we formulate our third hypothesis as follows:

There is a two-way interaction between perceived procedural justice and outcome judgments, such that defendants who judge

---

[2]In this paper, a "non-Western ethnic-cultural background" refers to being born in a non-Western country, which according to Statistics Netherlands (2018) refers to countries in Africa, Latin America, and Asia (excluding Indonesia and Japan), or Turkey. We also use the term to refer to persons whose parents or other ancestors were born in a non-Western country. We included these latter respondents in our sample because they, too, might experience discrimination in their daily lives due to their ethnic-cultural background.

their outcomes more negatively show stronger, attenuated, or reversed associations between perceived procedural justice and our other variables (i.e., trust in judges, protest intentions, and state self-esteem; Hypothesis 3).[3]

Research Context

To study these issues, we conducted a face-to-face survey among 198 defendants with a non-Western ethnic-cultural background involved in Dutch single judge criminal cases. In the Dutch court system, single judges handle relatively simple criminal cases in which the sanction demanded by the public prosecutor does not exceed 1 year of imprisonment. Typical cases handled by single judges include theft, simple assault, and traffic offenses such as driving under the influence. Defendants can be represented by a lawyer, or they can choose to defend themselves. Usually, single judge criminal court hearings last around 30 min and judges render a verdict directly afterward. The description below provides more details on the Dutch legal context and some of the main differences with the legal system of (for example) the United States.

> "First, Dutch criminal proceedings take place largely "on paper". That is, the emphasis is on the pretrial investigation rather than on court hearings, which generally last around 30 min in small criminal cases and 60–90 min in more severe ones. Second, the Dutch legal system does not have a plea-bargaining system like the United States. Third, the administration of justice is entirely in the hands of professional judges; the Dutch legal system does not have bifurcated proceedings in which defendants' guilt is determined by a jury and their sentences by a judge. Fourth, criminal court hearings in the Netherlands are less adversarial than in the United States. That is, Dutch hearings involve an active role for judges and traditionally treat defendants as subject of the investigation, whereas the United States legal system involves more passive judges and views the court hearing as a clash of parties." (Ansems et al., 2020, p. 648)

Relatively, many defendants in Dutch criminal cases have a non-Western ethnic-cultural background. People with a Moroccan or Antillean background in particular are overrepresented in Dutch crime statistics, which could be due partly to negative stereotypes and ethnic profiling (Huijnk and Andriessen, 2016). Discrimination is a relevant issue in Dutch society more broadly, too, as several studies show

that people with a non-Western migration background report relatively high levels of perceived discrimination (Huijnk et al., 2015; Huijnk and Andriessen, 2016; Andriessen et al., 2020). Indeed, there are signs that Dutch people with a migration background may be discriminated in important life domains (e.g., Thijssen et al., 2019).

Against this backdrop, the present study examines whether experiences of everyday discrimination and outcome judgments may moderate how defendants with a non-Western ethnic-cultural background involved in Dutch criminal cases react to perceived procedural justice during their court hearings. Our study helps to refine current insights into perceived procedural justice by focusing not only on the possible robustness of associations between perceived procedural justice and relevant other variables, but also on the potential attenuation or reversal of these associations. In doing so, we take insights from previous studies conducted in work settings and the psychological laboratory (e.g., Van den Bos et al., 1999; Brockner et al., 2009) and apply them to the novel context of Dutch criminal cases.

In addition, because of our focus on defendants with a non-Western ethnic-cultural background, this study sheds light on experiences of a relatively underinvestigated research population. Contrary to research participants in many other procedural justice studies, respondents in the current study generally have non-WEIRD (Western, Educated, Industrialized, Rich, and Democratic; Henrich et al., 2010) backgrounds. Taken together, we conducted our study in a real-life courtroom context, focusing on defendants with diverse ethnic-cultural backgrounds who might respond differently to perceived procedural justice. In this way, we critically examine the role of perceived procedural justice in Dutch criminal court hearings.

MATERIALS AND METHODS

Sample

Our sample consisted of 198 defendants with a non-Western ethnic-cultural background who appeared before a single judge at the court of the Mid-Netherlands in Utrecht, Lelystad, and Almere.[4] **Table 1** details sample characteristics.

Research Procedure

After gaining the court's permission to conduct the study, we collected our data between January 21 and October 15, 2019. Except for the summer break, the first author went to the court almost every work day during this period to collect data and stayed there for the duration of the criminal court's session that day (most often from 9 to 18 h, sometimes from 9 to 13 h, or from 13 to 18 h). Among the causes for the relatively long duration of data collection were our focus on defendants with a non-Western ethnic-cultural background, the fact that many defendants did not appear for their court hearings, and some defendants' poor command of Dutch.

---

[3]In this paper, we focus on the hypotheses that are central to the line of reasoning presented here. We also performed a small number of additional analyses. For example, we examined whether there was a significant three-way interaction between outcome judgments, perceived everyday discrimination, and perceived procedural justice. This analysis was conducted for exploratory purposes only, however, and is not reported in the current paper. After all, a power analysis (Faul et al., 2007) showed that, to achieve sufficient statistical power of 0.80 (Cohen et al., 2003) to detect the three-way interaction, with $\alpha = 0.05$ and a relatively small effect size ($f^2 = 0.02$), at least 387 respondents were needed. Complete details and results are available with the first author on request.

[4]Like some other Dutch courts, the court of the mid-Netherlands has buildings in multiple cities (including Utrecht, Lelystad, and Almere).

TABLE 1 | Sample description.

Categorical variables			
Variable	Category	N	%
Location of court hearing	Utrecht	190	96.0
	Lelystad	6	3.0
	Almere	2	1.0
Gender	Male	178	90.4
	Female	19	9.6
Highest completed level of education	Primary school	14	7.4
	Secondary school	81	42.6
	Secondary vocational education	62	32.6
	Higher professional education	24	12.6
	University	6	3.2
	Other	3	1.6
	Special needs education	*1*	*0.5*
	None at all	*2*	*1.1*
Ethnic-cultural background	Moroccan	85	42.9
	Surinam	25	12.6
	Turkish	20	10.1
	Antillean	12	6.1
	Other (e.g., Somalian, Iraqi, Afghan)	58	29.3
Offense	Assault or violence	57	30.0
	Theft, embezzlement, fencing, or breaking and entering	45	23.7
	Traffic offense (e.g., driving under the influence)	43	22.6
	Threatening someone	19	10.0
	Drug offense	17	8.9
	Insulting someone	13	6.8
	Destruction	12	6.3
	Scam or fraud	5	2.6
Case outcome	Convicted with imposition of sanction or measure (conditional or unconditional)	152	79.2
	Community service	*108*	*65.1*
	Fine	*51*	*30.7*
	Prison sentence	*27*	*16.3*
	Acquitted	26	13.5
	Found guilty without imposition of sanction or measure	13	6.8
	Discharged from further prosecution	2	1.0
Legal assistance	By lawyer	136	70.1
	By someone else	5	2.6
	None	53	27.3
Number of previous criminal court hearings	None	69	35.0
	One	48	24.4
	Two to ten	66	33.5
	More than ten	14	7.1
Continuous variables			
Variable	Range (years)	Average (years)	SD
Age	18–66	30.10	10.75

Defendants appeared before the court of the mid-Netherlands if they were accused of a crime that had been committed in that geographic region or if they were living there.

The first author approached defendants in the court hallway after they had made their presence known at the counter to ask whether they were willing to participate in a study about how fairly and justly they felt they were treated during their court hearings, indicating that they would be thanked for their participation with a small token of appreciation. Seventeen respondents (8.6% of the sample) were approached by a research assistant. We approached respondents before the start of their court hearings as much as possible to ask whether they were willing to participate in the study once their court hearings had ended. When it was not possible to approach respondents before the start of their court hearings – for instance, because they appeared for their hearings only very last minute or because they were consulting with their lawyers – respondents were approached immediately after their court hearings.

Our study procedures were approved by the ethical board of the Faculty of Law, Economics, Governance, and Organization at Utrecht University. Following these approved procedures, and because our study focused on how people with non-Western ethnic-cultural backgrounds would respond to issues of procedural justice, we approached defendants who appeared to have a non-Western ethnic-cultural background for participation in our study. Therefore, based on their names and physical appearance, we made an initial assessment whether people appeared to have a non-Western ethnic-cultural background and invited those people to take part in our study.[5] At the start of the questionnaire, we informed respondents that we were interested in how people who were born in a different country than the Netherlands and people whose parents or other ancestors were born in a different country than the Netherlands would evaluate how they were treated during their court hearings, their trust in Dutch judges, and how they felt treated in their daily lives, among other things. We ensured that we always treated people respectfully throughout the entire study. In fact, while filling out the questionnaire or afterward, multiple respondents indicated that they appreciated studies like ours, as these are needed to help to understand discrimination in Dutch society.

We also note that, due to our way of sampling, we may have missed people who were eligible for participation in our study but whose physical appearance or name was not clearly non-Western. Although we cannot rule out that this may have affected our results, we do not think this was a big problem in the current study. After all, these people may be less likely to feel discriminated against based on their ethnic-cultural background, whereas we were particularly interested in defendants

---

[5]After having signed a confidentiality agreement, at the beginning of each week of data collection the first author received an overview of single judge criminal cases that would be heard that week, detailing times, charges, and defendants' names. We used these overviews to decide in which hallway to wait for potential respondents when there were multiple court hearings taking place at different floors at the same time (to be able to approach as many eligible defendants as possible) and we shredded these overviews at the end of each week for reasons of confidentiality.

who experience relatively high levels of discrimination in Dutch society and might therefore respond differently to perceived procedural justice. Our impression is that we were successful in conducting our study in responsible and sound ways. In the Discussion section, we note limitations of our study that may inspire future research.

In addition to having a non-Western ethnic-cultural background, our other inclusion criteria were that defendants had received the outcome of their case and that they had a sufficient command of Dutch. We immediately filtered out respondents who, when starting to fill out the questionnaire in the court hallway, turned out not to meet our inclusion criteria and thus turned out to be ineligible for participation in our study. In total, we approached 447 defendants (excluding defendants who, based on this initial screening, turned out to be ineligible for participation). Of those 447 defendants, 210 filled out our questionnaire, resulting in a response rate of 47.0%. In a later stage, before conducting our analyses, we filtered out the questionnaires that did not indicate the respondent's ethnic-cultural background or that had a very large number of missing values. Thus, the final sample consisted of 198 respondents. A power analysis (Faul et al., 2007) showed that, to achieve sufficient statistical power of 0.80 (Cohen et al., 2003) to detect the two-way interaction between perceived everyday discrimination and perceived procedural justice, with $\alpha = 0.05$ and a relatively small effect size ($f^2 = 0.04$), at least 191 respondents were needed.

Most respondents completed the questionnaire directly. Six respondents (3.0% of the sample) filled it out at home and sent it to us in an envelope with prepaid postage stamps. The respondents who filled out the questionnaire directly often did so themselves, while 25 respondents (12.8% of the sample) preferred having the questions read out loud by the researcher. Before respondents filled out the questionnaire, we explained that the research focused on persons who were born in a different country and persons whose parents or other ancestors were born in a different country. In addition, we told respondents that participation was voluntary and anonymous, and that the research was conducted independently of the court and the Public Prosecution Service.

After they completed the questionnaire, we thanked respondents for their participation and offered to send them a summary of our research results, which we sent to interested respondents later. During the entire period of data collection, we kept an extensive logbook detailing relevant background information to the research, such as information obtained through informal conversations with defendants and defense lawyers.

Measures
Our main variables were perceived procedural justice, outcome judgments, perceived everyday discrimination, trust in Dutch judges, protest intentions, and state self-esteem. The questionnaire started with those variables relating to the court hearing (perceived procedural justice, outcome judgments, protest intentions, and trust in Dutch judges) and then assessed variables targeting respondents' perceptions more generally (state self-esteem and perceived everyday discrimination).[6]

We measured perceived procedural justice with a six-item scale based on the findings of a recent qualitative interview study conducted in the same criminal courtroom context as the current study (Ansems et al., 2020). Our survey items corresponded with the six core components of perceived procedural justice among defendants as revealed by the interview study. Specifically, we asked respondents to indicate, on a Likert scale from 1 (*totally disagree*) to 7 (*totally agree*), to what extent they agreed with the following six statements: "During the court hearing, I was treated in a pleasant way," "During the court hearing, I was treated in an unprejudiced manner," "During the court hearing, I was sufficiently able to tell my side of the story," "During the court hearing, my side of the story was listened to," "During the court hearing, everything important has been taken into account," and "During the court hearing, my case was treated in a careful manner." Together, these items formed a reliable scale ($\alpha = 0.82$) on which higher scores reflect higher levels of perceived procedural justice. Therefore, we report the results of our analyses without the additional 11 items which we included as backup in case the six-item scale would turn out to be unreliable and which were based on previous work in other courtroom settings (Grootelaar and van den Bos, 2018).[7]

We also assessed respondents' outcome judgments, which in this study include outcome satisfaction, perceived outcome fairness, and perceived outcome favorability. Our outcome judgments scale was largely based on previous research in a similar context (Grootelaar and van den Bos, 2018) and consisted of six items: "I find this ruling fair," "I find this ruling favorable," "I am satisfied with the judge's ruling," "I find this ruling just," "The judge's ruling has positive consequences for me," and "I agree with the judge's ruling." Again, respondents indicated on a scale from 1 to 7 to what extent they agreed with these statements, and for each respondent, we took the average of their scores on these items to calculate their scores on our outcome judgments scale ($\alpha = 0.97$). Higher scores on this scale indicate that respondents judged their outcomes more positively.

We examined perceived everyday discrimination with the 10-item version of the everyday discrimination scale (Williams et al., 1997, 2008). We asked respondents to indicate on a

---

[6]Because we report all measures used in our questionnaire, we note that we also measured respondents' external attribution ratings and the extent to which they identified with their ethnic-cultural subgroup. The items we used to measure respondents' external attribution ratings yielded a very low Cronbach's alpha ($\alpha = 0.17$), rendering this variable unsuitable for analysis. We included the items on subgroup identification in our questionnaire for potential additional analyses, as previous work suggests that attributions to discrimination might not protect the self-esteem of people who strongly identify with their ethnic-cultural subgroup (McCoy and Major, 2003). In the end, we did not perform these analyses because of power issues. Hence, we decided to drop these variables from the current paper. Complete details and results are available on request.

[7]We conducted all analyses involving perceived procedural justice with both this six-item scale and the entire 17-item scale. In the Results section, we explicitly note when these analyses yielded (slightly) different results with regard to our main findings.

scale from 1 (*never*) to 6 (*almost every day*) how often they encountered the following events in their daily lives: "In my day-to-day life, I am treated with less courtesy than other people are," "In my day-to-day life, I am treated with less respect than other people are," "In my day-to-day life, I receive poorer service than other people at restaurants or stores," "In my day-to-day life, people act as if they think I am not smart," "In my day-to-day life, people act as if they are afraid of me," "In my day-to-day life, people act as if they think I am dishonest," "In my day-to-day life, people act as if they are better than I am," "In my day-to-day life, I am called names or insulted," "In my day-to-day life, I am threatened or harassed," and "In my day-to-day life, I am followed around in stores." Together, these items formed a reliable perceived everyday discrimination scale ($\alpha = 0.91$). Higher scores on this scale reflect higher levels of perceived everyday discrimination. In addition, respondents who answered "a few times a year" (score 3 on the six-point scale) or more often to at least one question were asked to indicate what they thought was the main reason for these experiences: their gender, their age, their religion, their ethnic-cultural background, their level of education, their level of income, and/or some other reason (which they could then write down). In this way, we assessed perceived grounds of discrimination.

We solicited their trust in Dutch judges with items that target this construct in a way that we deemed as direct and straightforward as possible (see also Grootelaar and van den Bos, 2018). Specifically, we asked respondents to indicate on a scale from 1 (*totally disagree*) to 7 (*totally agree*) to what extent they agreed with the following five statements: "I have faith in Dutch judges," "I deem Dutch judges trustworthy," "I trust Dutch judges," "I do not trust Dutch judges" (reverse coded), and "I feel like Dutch judges cannot be trusted" (reverse coded). Respondents' answers on these items were averaged into a reliable trust in Dutch judges scale ($\alpha = 0.90$) on which higher scores reflect higher levels of trust. We also included an additional sixth item asking respondents to express their trust in Dutch judges with a grade between 1 (*lowest*) and 10 (*highest*), in line with the grading system used in Dutch schools.

Following Stahl et al. (2008), we assessed protest intentions by asking respondents to indicate on a scale from 1 (*not at all*) to 7 (*very much*) to what extent they would like to criticize the ruling and to what extent they would like to protest against the ruling. Respondents' answers on these two items were averaged into a reliable protest intentions scale ($\alpha = 0.85$). Higher scores on this scale represent stronger protest intentions.

Finally, to measure respondents' state self-esteem at the moment they filled out our questionnaire we adapted the global self-esteem scale of Rosenberg (1965) to measure state global self-esteem. Hence, respondents were asked to indicate on a scale from 1 (*totally disagree*) to 7 (*totally agree*) to what extent they agreed with the following 10 statements: "Now, at this moment, I am satisfied with myself," "Now, at this moment, I think I am no good at all" (reverse coded), "Now, at this moment, I feel that I have a number of good qualities," "Now, at this moment, I am able to do things as well as most other people," "Now, at this moment, I feel like I do not have much

to be proud of" (reverse coded), "Now, at this moment, I feel useless" (reverse coded), "Now, at this moment, I feel that I am a person of worth, at least on an equal plane with others," "Now, at this moment, I wish I could have more respect for myself" (reverse coded), "Now, at this moment, I feel like I am a failure" (reverse coded), and "Now, at this moment, I take a positive attitude toward myself." Respondents' answers on these items were averaged into a reliable state self-esteem scale ($\alpha = 0.83$) on which higher scores reflect higher state self-esteem.

We also assessed relevant background variables, asking respondents to indicate whether they had legal assistance during their court hearings, their number of previous court hearings before a criminal judge, their highest completed level of education, their gender, and their age. At the end of the questionnaire, respondents could write down remarks or issues they deemed important that had not been the subject of our questions.[8]

RESULTS

Descriptive Statistics and Bivariate Correlations

All statistical analyses were conducted with IBM SPSS software. **Table 2** presents means, standard deviations, and bivariate correlations for our main variables and background variables.

As shown in **Table 2**, there were statistically significant relationships between some of our background variables (i.e., legal assistance, number of previous court hearings, and age) and some of our main variables (i.e., perceived procedural justice, trust in judges, and protest intentions). Hence, we controlled for legal assistance, number of previous court hearings, and age in the hierarchical regression analyses reported below by entering them in Step 1 of the analysis. Main effects were entered in Step 2, and two-way interactions were entered in Step 3. Following recommendations by Cohen et al. (2003), all continuous independent variables (including quasi-interval variables) were standardized before being entered into the equation when the equation involved an interaction effect. When reporting the results of these hierarchical regression analyses, we focus on the last step in the analysis that significantly added to the amount of explained variance in the dependent variables in our regression equations.

Reacting to Procedural Justice

Hypothesis 1 predicted that defendants with higher levels of perceived procedural justice report more trust in judges, more

---

[8]There were missing values for perceived ground of discrimination (51 missing values), trust in judges (one missing value), grade for trust in judges (12 missing values), and self-esteem (one missing value). There were also some missing values for gender (one missing value), age (three missing values), highest completed level of education (eight missing values), offense (eight missing values), type of verdict (six missing values), sanction received (32 missing values), legal assistance (four missing values), and number of previous court hearings (one missing value).

TABLE 2 | Means, standard deviations, and correlations for the main variables and background variables.

Variable	M	SD	1	2	3	4	5	6	7	8	9	10	11	12
1. Procedural justice	5.38	1.27	–											
2. Outcome judgments	4.55	2.18	0.59***	–										
3. Discrimination	2.48	1.16	−0.06	−0.12	–									
4. Protest intentions	3.31	2.02	−0.45***	−0.60***	0.26***	–								
5. Trust in judges	5.09	1.55	0.50***	0.42***	−0.28***	−0.34***	–							
6. Trust in judges grade	6.83	2.06	0.47***	0.40***	−0.22**	−0.34***	0.79***	–						
7. Self-esteem	5.62	1.09	0.19**	0.16*	−0.22**	−0.28***	0.26***	0.15*	–					
8. Legal assistance (0=no, 1=yes)	–	–	−0.19**	−0.06	0.07	0.05	−0.14	−0.17*	0.00	–				
9. Previous hearings	2.14	1.00	−0.06	0.02	0.11	−0.03	−0.15*	−0.18*	−0.13	0.11	–			
10. Level of education	5.37	2.26	−0.05	−0.11	−0.08	0.05	0.02	0.02	−0.01	−0.01	−0.33***	–		
11. Gender (0=female, 1=male)	–	–	−0.03	−0.11	0.05	0.06	−0.05	0.03	−0.04	−0.09	0.17	−0.02	–	
12. Age	30.10	10.75	0.10	0.10	−0.10	−0.15*	0.17*	0.14	0.02	−0.14	0.11	−0.19*	−0.06	–

The number of previous court hearings was measured on a five-point scale (1=0, 2=1, 3=2–10, 4=11–20, 5=more than 20). Highest completed level of education was measured on a nine-point scale, ranging from primary school (coded as 1) to university (coded as 9). Perceived grounds for discrimination included respondents' ethnic-cultural background (N=93; 63.3% of the sample), religion (N=56; 38.1% of the sample), gender (N=25; 17.0% of the sample), age (N=21; 14.3% of the sample), level of income (N=17; 11.6% of the sample), and level of education (N=17; 11.6% of the sample). *p<0.05; **p<0.01; ***p<0.001.

positive outcome judgments, lower intentions to protest against their outcomes, and higher levels of state self-esteem. This hypothesis was supported by our results. That is, respondents who felt treated more fairly during their court hearings showed more trust in judges ($\beta=0.50$) and gave their trust in judges higher grades ($\beta=0.44$), judged their outcomes more positively ($\beta=0.59$), indicated lower protest intentions ($\beta=-0.45$), and reported higher state self-esteem ($\beta=0.22$). Further details are presented in **Tables 3–6**.

Adding Perceived Everyday Discrimination

Hypothesis 2 proposed that there is a two-way interaction between perceived procedural justice and perceived everyday discrimination, such that defendants who experience relatively high levels of everyday discrimination show attenuated or reversed associations between perceived procedural justice and our other variables (i.e., trust in judges, outcome judgments, protest intentions, and state self-esteem). This hypothesis was not supported by our results, as we did not find significant interaction effects between perceived everyday discrimination and perceived procedural justice. We did find significant main effects of perceived procedural justice, sometimes in addition to significant main effects of perceived everyday discrimination. That is, perceived procedural justice was positively associated with respondents' trust in judges ($\beta=0.48$) and the grades they gave their trust in judges ($\beta=0.42$), their outcome judgments ($\beta=0.58$), and their state self-esteem ($\beta=0.21$) and was negatively related to respondents' protest intentions ($\beta=-0.43$). In addition, perceived everyday discrimination was negatively associated with trust in judges ($\beta=-0.22$) and the grades respondents gave their trust in judges ($\beta=-0.16$) as well as respondents' state self-esteem ($\beta=-0.17$) and was positively associated with protest intentions ($\beta=0.24$). Further details are shown in **Tables 7–10**.

Adding Outcome Judgments

Hypothesis 3 suggested that there is a two-way interaction between perceived procedural justice and outcome judgments, such that defendants who judge their outcomes more negatively show stronger, attenuated, or reversed associations between perceived procedural justice and our other variables (i.e., trust in judges, protest intentions, and state self-esteem). This hypothesis was not supported by the results, as our analyses did not reveal significant interaction effects between outcome judgments and perceived procedural justice. Our analyses did yield significant main effects of perceived procedural justice, sometimes in addition to significant main effects of outcome judgments. More specifically, we found a positive association between perceived procedural justice and trust in judges ($\beta=0.37$) and the grade respondents gave their trust in judges ($\beta=0.32$), a marginally significant association between perceived procedural justice and self-esteem ($\beta=0.18$), and a negative association between perceived procedural justice and protest intentions ($\beta=-0.15$). We also found a positive association between outcome judgments and trust in judges ($\beta=0.22$) and the grade respondents gave their trust in judges ($\beta=0.19$) and a negative

association between outcome judgments and protest intentions ($\beta = -0.50$). Entering the entire 17-item perceived procedural justice scale into the regression equation rather than the six-item scale yielded the same (non-significant) result regarding the interaction between outcome judgments and perceived procedural justice, the only difference being that the association between outcome judgments and trust in judges was no longer statistically significant. **Tables 11–13** present further details.

DISCUSSION

The present study critically examines the role of perceived procedural justice, and other important variables, in Dutch criminal court hearings. We think the message of what we learn from the reported findings is twofold. First, perceived procedural justice matters. That is, our findings showed robust associations between perceived procedural justice and trust in judges, outcome judgments, protest intentions, and state self-esteem. Second, processes of self-enhancement did not have the effects found by studies conducted in organizational contexts or laboratory settings. That is, outcome judgments and perceptions of everyday discrimination did not significantly moderate the associations between perceived procedural justice, on the one hand, and trust in judges, protest intentions, and state self-esteem on the other hand. In what follows, we deepen these conclusions. We then discuss the limitations of the present study, suggestions for future research that follow from these limitations, and practical implications of our findings.

The Importance of Fair Procedures

An important finding of this study is that respondents who felt treated more fairly during their court hearings reported higher levels of trust in judges, judged their outcomes more positively, showed lower protest intentions, and displayed higher state self-esteem. These favorable reactions to perceived procedural justice indicate that, even in the real-life courtroom context of our study in which respondents risked actual sanctions, respondents cared not only about their outcomes but also about the way they were treated during their court hearings. Thus, the current study contributes to the ongoing academic debate on the relative importance of perceived procedural justice in real-life cases. Our findings support the argument by Casper et al. (1988) that the positive associations between perceived procedural justice and relevant other variables represent real-world phenomena that can also be observed outside the artificial settings of psychological laboratories, including criminal justice contexts (see also, for instance, Tyler, 1984, 1988, 2006; Paternoster et al., 1997; Tyler and Huo, 2002).

Our study also complements current insights into perceived procedural justice in criminal justice contexts by examining whether defendants' reactions to perceived procedural fairness may be moderated by experiences of everyday discrimination and outcome judgments. Our results indicated that none of the interaction effects we examined were statistically significant. In other words, rather than being attenuated or reversed, the

associations between perceived procedural justice and our other variables remained intact regardless of the extent to which respondents experienced discrimination in their daily lives and how positively or negatively they judged their outcomes. This might be interpreted as an indication of the robustness of the fair process effect. These findings also fit with other studies, which show that people belonging to ethnic minorities respond equally favorably to perceived procedural justice as do people from majority groups (Tyler, 2001; Sunshine and Tyler, 2003; Higgins and Jordan, 2005; Johnson et al., 2017).

Different Cases and Contexts

We note that there may also be other possible explanations for the lack of statistically significant interaction effects in our study. These explanations may relate, for instance, to the type of cases examined and the context of our research. In the organizational and performance-oriented settings of previous studies examining attenuated or reversed fair process effects (e.g., Van den Bos et al., 1999; Brockner et al., 2009), negative outcomes were likely to threaten people's self-esteem and thus make them look for opportunities to attribute these outcomes to external causes. In the courtroom context of the current study, negative case outcomes may not have posed a similar threat to respondents' sense of self-worth. Hence, the lack of significant interaction effects might be explained by the legal context of our study. This indicates, we think, that more research is needed into the operations of self-enhancement processes in relevant legal contexts, such as criminal court hearings. Our study provides an important first step in this regard.

Similarly, the interaction between outcome judgments and perceived procedural justice has generally been found in work contexts or in other settings with different types of respondents than we examined in the current study (for overviews, see Brockner and Wiesenfeld, 1996; Brockner, 2010). Previous studies examining perceptions of actual defendants in criminal cases (Grootelaar and van den Bos, 2018) or undergraduates putting themselves in the position of defendants (Walker et al., 1974) did not find an interaction effect between outcomes and procedural justice. Our findings thus provide further support for the argument by Grootelaar and Van den Bos (2018), who did find interactions between perceived procedural justice and outcome favorability for motoring fine cases, that the type of case may play an important role in shaping people's reactions to perceived procedural justice and outcome favorability in legal contexts.

Conflicting Psychological Processes

Another potential explanation for not finding interactive effects of outcome judgments or perceived everyday discrimination and perceived procedural justice might be that conflicting psychological processes are at work. That is, the self-enhancement processes underlying attenuated or reversed fair process effects in other studies might play a role in the courtroom context of our study, but their effects may have been canceled out or overridden by other psychological processes (see also Brockner et al., 2009).

TABLE 3 | Trust in judges regressed on procedural justice and relevant background variables.

Variable	Step 1				95% CI for b	Step 2				95% CI for b
	b	β	t	p		b	β	t	p	
Legal assistance (0=no, 1=yes)	−0.37	−0.11	−1.48	0.141	−0.86, 0.12	−0.08	−0.02	−0.36	0.718	−0.51, 0.35
Previous hearings	−0.21	−0.14	−1.91	0.058	−0.43, 0.01	−0.18	−0.12	−1.86	0.065	−0.36, 0.01
Age	0.03	0.18	2.50	0.013	0.01, 0.05	0.02	0.15	2.44	0.016	0.00, 0.04
Procedural justice						0.61	0.50	7.98	0.000	0.46, 0.76
df			3					4		
F		4.31, p =0.006					20.23, p =0.000			
F change		4.31, p =0.006					63.62, p =0.000			
R²			0.07					0.31		
Adjusted R²			0.05					0.29		
N			189					189		

The number of previous court hearings was measured on a five-point scale (1=0, 2=1, 3=2–10, 4=11–20, 5=more than 20).

TABLE 4 | Outcome judgments regressed on procedural justice and relevant background variables.

Variable	Step 1				95% CI for b	Step 2				95% CI for b
	b	β	t	p		b	β	t	p	
Legal assistance (0=no, 1=yes)	−0.27	−0.06	−0.75	0.455	−0.98, 0.44	0.21	0.04	0.72	0.475	−0.38, 0.80
Previous hearings	0.05	0.02	0.33	0.745	−0.26, 0.37	0.11	0.05	0.83	0.406	−0.15, 0.37
Age	0.02	0.08	1.08	0.282	−0.01, 0.05	0.01	0.05	0.77	0.442	−0.02, 0.03
Procedural justice						1.03	0.59	9.69	0.000	0.82, 1.24
df			3					4		
F		0.72, p =0.541					24.27, p =0.000			
F change		0.72, p =0.541					93.85, p =0.000			
R²			0.01					0.34		
Adjusted R²			−0.00					0.33		
N			190					190		

The number of previous court hearings was measured on a five-point scale (1=0, 2=1, 3=2–10, 4=11–20, 5=more than 20).

TABLE 5 | Protest intentions regressed on procedural justice and relevant background variables.

Variable	Step 1				95% CI for b	Step 2				95% CI for b
	b	β	t	p		b	β	t	p	
Legal assistance (0=no, 1=yes)	0.16	0.04	0.48	0.630	−0.50, 0.82	−0.18	−0.04	−0.60	0.550	−0.78, 0.42
Previous hearings	−0.03	−0.02	−0.20	0.842	−0.32, 0.26	−0.07	−0.04	−0.52	0.601	−0.33, 0.19
Age	−0.02	−0.13	−1.71	0.088	−0.05, 0.00	−0.02	−0.10	−1.52	0.130	−0.04, 0.01
Procedural justice						−0.72	−0.45	−6.74	0.000	−0.94, −0.51
df			3					4		
F		1.21, p =0.308					12.49, p =0.000			
F change		1.21, p =0.308					45.46, p =0.000			
R²			0.02					0.21		
Adjusted R²			0.00					0.20		
N			190					190		

The number of previous court hearings was measured on a five-point scale (1=0, 2=1, 3=2–10, 4=11–20, 5=more than 20).

For instance, defendants who experience much discrimination in their daily lives may be pleasantly surprised by how fairly they feel treated during their court hearings, leading them to respond even *more* favorably to perceived procedural justice than defendants who experience little everyday discrimination. In the current paper, we explored whether defendants who

TABLE 6 | Self-esteem regressed on procedural justice and relevant background variables.

Variable	Step 1				95% CI for *b*	Step 2				95% CI for *b*
	b	*β*	*t*	*p*		*b*	*β*	*t*	*p*	
Legal assistance (0 = *no*, 1 = *yes*)	0.05	0.02	0.26	0.793	−0.31, 0.40	0.14	0.06	0.77	0.440	−0.21, 0.49
Previous hearings	−0.14	−0.13	−1.78	0.077	−0.30, 0.02	−0.13	−0.12	−1.68	0.094	−0.28, 0.02
Age	0.00	0.03	0.39	0.696	−0.01, 0.02	0.00	0.02	0.23	0.821	−0.01, 0.02
Procedural justice						0.19	0.22	3.07	0.002	0.07, 0.32
df			3					4		
F		1.06, *p* = 0.366					3.19, *p* = 0.015			
F change		1.06, *p* = 0.366					9.41, *p* = 0.002			
*R*²			0.02					0.07		
Adjusted *R*²			0.00					0.04		
N			189					189		

The number of previous court hearings was measured on a five-point scale (1 = 0, 2 = 1, 3 = 2–10, 4 = 11–20, 5 = more than 20).

experience much discrimination in their daily lives might respond *less* favorably to perceived procedural justice because of self-enhancement processes. These potential moderating effects of perceived everyday discrimination may have canceled each other out, resulting in non-significance of the interaction effect.

Defendants' desire for fair treatment may also simply have overridden their self-enhancement motive. After all, perceived procedural justice may be desirable for various instrumental and non-instrumental reasons, as explained in the Introduction. These beneficial aspects of perceived procedural justice may have been stronger than defendants' self-enhancement processes, resulting in favorable responses to procedures that defendants perceive as fair rather than unfair.

Levels of Perceived Everyday Discrimination

Respondents' relatively low levels of perceived everyday discrimination ($M = 2.48$, SD = 1.16, measured on a six-point scale) may be relevant as well. After all, respondents who scored one standard deviation above the mean level of perceived everyday discrimination (i.e., a score of 3.64) encountered negative treatment between a few times a year (score 3) and a few times a month (score 4). These experiences of discrimination may not have been sufficiently frequent to make defendants respond favorably to perceived procedural unfairness during their court hearings for self-enhancement reasons. Hence, we recommend that future studies examining these issues use samples in which levels of perceived everyday discrimination are likely to be higher.

Limitations

An engaging aspect of our study, we think, is that we were able to study perceptions of actual defendants in single judge criminal cases after a 9-month period of data collection at the court of the mid-Netherlands. The flip side of this approach is that our sample is sufficiently large, yet smaller than we would have wanted ideally. For instance, a larger sample would have

enabled us to robustly examine the three-way interaction between outcome judgments, perceived everyday discrimination, and perceived procedural justice. We would expect attenuated or reversed associations between perceived procedural justice and relevant other variables in particular among respondents with both relatively high levels of perceived everyday discrimination and relatively negative outcome judgments. Thus, future studies with larger samples are needed to better understand the issues examined in the current paper. Follow-up studies with larger samples could also examine, for instance, the possible role of being found guilty and seriousness of the sanction imposed. After all, more serious sanctions might pose a greater threat to defendants' self-esteem, thereby making attenuated or reversed fair process effects more likely to occur.

We also note that we conducted our study at only one court and included only single judge criminal cases. Furthermore, the first author – who collected the bulk of the data – is a White and university-based researcher. As a result, interviewer effects may have played a role in our study. For instance, respondents may have concealed their levels of distrust in Dutch judges, as they may have considered the researcher as belonging to their outgroup (Hulst, 2017). Thus, we propose that it is important to replicate our study in other courts with different researchers and different types of court cases. Follow-up studies could also include defendants whom we were not able to include in our current sample, such as defendants in pre-trial detention.

In addition, the correlational design of this study does not allow for conclusions about any causal relationships between our variables. Thus, although we believe the field work element to be a strength of the paper, this also has methodological limitations. For instance, perceptions of procedural justice may influence defendants' outcome judgments, and vice versa, which renders the analysis pertaining to Hypothesis 3 difficult to interpret.

Furthermore, a possibility that cannot be ruled out in our correlational design is that some of the effects that we examined with this hypothesis might already be present in the variation in the independent variables, making the validity of the interaction

TABLE 7 | Regression results for procedural justice, discrimination, and their interaction on trust in judges.

Variable	Step 1				95% CI for b	Step 2				95% CI for b	Step 3				95% CI for b
	b	β	t	p		b	β	t	p		b	β	t	p	
Legal assistance (0 = no, 1 = yes)	−0.37	−0.11	−1.48	0.141	−0.86, 0.12	−0.05	−0.02	−0.25	0.804	−0.47, 0.36	−0.04	−0.01	−0.18	0.858	−0.45, 0.38
Previous hearings	−0.21	−0.14	−1.91	0.058	−0.43, 0.01	−0.14	−0.09	−1.54	0.125	−0.33, 0.04	−0.14	−0.09	−1.46	0.146	−0.32, 0.05
Age	0.28	0.18	2.50	0.013	0.06, 0.50	0.20	0.13	2.11	0.036	0.01, 0.38	0.19	0.13	2.05	0.042	0.01, 0.38
Procedural justice						0.75	0.48	7.94	0.000	0.57, 0.94	0.76	0.49	8.00	0.000	0.57, 0.95
Discrimination						−0.33	−0.22	−3.56	0.000	−0.51, −0.15	−0.35	−0.23	−3.71	0.000	−0.53, −0.16
Discrimination × Procedural justice											0.11	0.07	1.14	0.258	−0.08, 0.30
df		3					5					6			
F		4.31, p = 0.006					19.75, p = 0.000					16.70, p = 0.000			
F change		4.31, p = 0.006					40.17, p = 0.000					1.29, p = 0.258			
R²		0.07					0.35					0.36			
Adjusted R²		0.05					0.33					0.33			
N		189					189					189			

The number of previous court hearings was measured on a five-point scale (1 = 0, 2 = 1, 3 = 2–10, 4 = 11–20, 5 = more than 20).

TABLE 8 | Regression results for procedural justice, discrimination, and their interaction on outcome judgments.

Variable	Step 1				95% CI for b	Step 2				95% CI for b	Step 3				95% CI for b
	b	β	t	p		b	β	t	p		b	β	t	p	
Legal assistance (0 = no, 1 = yes)	−0.27	−0.06	−0.75	0.455	−0.98, 0.44	0.23	0.05	0.78	0.438	−0.36, 0.82	0.24	0.05	0.79	0.430	−0.35, 0.83
Previous hearings	0.05	0.02	0.33	0.745	−0.26, 0.37	0.13	0.06	0.99	0.322	−0.13, 0.39	0.13	0.06	1.01	0.315	−0.13, 0.39
Age	0.17	0.08	1.08	0.282	−0.14, 0.49	0.08	0.04	0.60	0.552	−0.18, 0.34	0.08	0.04	0.58	0.565	−0.19, 0.34
Procedural justice						1.29	0.58	9.58	0.000	1.02, 1.55	1.29	0.58	9.56	0.000	1.02, 1.55
Discrimination						−0.20	−0.09	−1.53	0.127	−0.46, 0.06	−0.21	−0.10	−1.56	0.122	−0.47, 0.06
Discrimination × Procedural Justice											0.04	0.02	0.29	0.774	−0.23, 0.31
df		3					5					6			
F		0.72, p = 0.541					20.03, p = 0.000					16.62, p = 0.000			
F change		0.72, p = 0.541					48.44, p = 0.000					0.08, p = 0.774			
R²		0.01					0.35					0.35			
Adjusted R²		−0.00					0.34					0.33			
N		190					190					190			

The number of previous court hearings was measured on a five-point scale (1 = 0, 2 = 1, 3 = 2–10, 4 = 11–20, 5 = more than 20).

TABLE 9 | Regression results for procedural justice, discrimination, and their interaction on protest intentions.

Variable	Step 1				95% CI for b	Step 2				95% CI for b	Step 3				95% CI for b
	b	β	t	p		b	β	t	p		b	β	t	p	
Legal assistance (0 = no, 1 = yes)	0.16	0.04	0.48	0.630	−0.50, 0.82	−0.22	−0.05	−0.76	0.447	−0.80, 0.36	−0.24	−0.05	−0.82	0.414	−0.82, 0.34
Previous hearings	−0.03	−0.02	−0.20	0.842	−0.32, 0.26	−0.12	−0.06	−0.93	0.354	−0.38, 0.14	−0.13	−0.06	−0.99	0.323	−0.38, 0.13
Age	−0.25	−0.13	−1.71	0.088	−0.55, 0.04	−0.15	−0.07	−1.15	0.253	−0.41, 0.11	−0.14	−0.07	−1.09	0.277	−0.40, 0.12
Procedural justice						−0.88	−0.43	−6.68	0.000	−1.14, −0.62	−0.89	−0.44	−6.73	0.000	−1.15, −0.63
Discrimination						0.48	0.24	3.71	0.000	0.22, 0.73	0.50	0.25	3.82	0.000	0.24, 0.76
Discrimination × Procedural justice											−0.13	−0.06	−0.98	0.328	−0.40, 0.13
df	3					5					6				
F	1.21, p = 0.308					13.43, p = 0.000					11.35, p = 0.000				
F change	1.21, p = 0.308					31.17, p = 0.000					0.96, p = 0.328				
R^2	0.02					0.27					0.27				
Adjusted R^2	0.00					0.25					0.25				
N	190					190					190				

The number of previous court hearings was measured on a five-point scale (1 = 0, 2 = 1, 3 = 2–10, 4 = 11–20, 5 = more than 20).

TABLE 10 | Regression results for procedural justice, discrimination, and their interaction on self-esteem.

Variable	Step 1				95% CI for b	Step 2				95% CI for b	Step 3				95% CI for b
	b	β	t	p		b	β	t	p		b	β	t	p	
Legal assistance (0 = no, 1 = yes)	0.05	0.02	0.26	0.793	−0.31, 0.40	0.15	0.06	0.87	0.386	−0.19, 0.50	0.16	0.07	0.91	0.365	−0.19, 0.51
Previous hearings	−0.14	−0.13	−1.78	0.077	−0.30, 0.02	−0.11	−0.10	−1.45	0.149	−0.26, 0.04	−0.11	−0.10	−1.40	0.164	−0.26, 0.05
Age	0.03	0.03	0.39	0.696	−0.13, 0.19	−0.00	−0.00	−0.04	0.971	−0.16, 0.15	−0.01	−0.01	−0.08	0.940	−0.16, 0.15
Procedural justice						0.23	0.21	2.92	0.004	0.08, 0.39	0.23	0.21	2.95	0.004	0.08, 0.39
Discrimination						−0.19	−0.17	−2.43	0.016	−0.34, −0.04	−0.20	−0.18	−2.51	0.013	−0.35, −0.04
Discrimination × Procedural justice											0.05	0.05	0.67	0.502	−0.10, 0.21
df	3					5					6				
F	1.06, p = 0.366					3.80, p = 0.003					3.23, p = 0.005				
F change	1.06, p = 0.366					7.78, p = 0.001					0.45, p = 0.502				
R^2	0.02					0.09					0.10				
Adjusted R^2	0.00					0.07					0.07				
N	189					189					189				

The number of previous court hearings was measured on a five-point scale (1 = 0, 2 = 1, 3 = 2–10, 4 = 11–20, 5 = more than 20).

TABLE 11 | Regression results for procedural justice, outcome judgments, and their interaction on trust in judges.

Variable	Step 1				95% CI for b	Step 2				95% CI for b	Step 3				95% CI for b
	b	β	t	p		b	β	t	p		b	β	t	p	
Legal assistance (0 = no, 1 = yes)	−0.37	−0.11	−1.48	0.141	−0.86, 0.12	−0.12	−0.03	−0.55	0.585	−0.54, 0.30	−0.12	−0.04	−0.56	0.579	−0.54, 0.31
Previous hearings	−0.21	−0.14	−1.91	0.058	−0.43, 0.01	−0.20	−0.13	−2.09	0.038	−0.38, −0.01	−0.20	−0.13	−2.08	0.039	−0.38, −0.01
Age	0.28	0.18	2.50	0.013	0.06, 0.50	0.22	0.14	2.34	0.020	0.04, 0.41	0.22	0.15	2.33	0.021	0.03, 0.41
Procedural justice						0.58	0.37	4.91	0.000	0.34, 0.81	0.57	0.37	4.75	0.000	0.33, 0.81
Outcome judgments						0.34	0.22	2.94	0.004	0.11, 0.56	0.34	0.22	2.94	0.004	0.11, 0.56
Outcome judgments × Procedural justice											−0.02	−0.01	−0.14	0.888	−0.23, 0.20
df		3					5					6			
F		4.31, p = 0.006					18.59, p = 0.000					15.42, p = 0.000			
F change		4.31, p = 0.006					37.47, p = 0.000					0.02, p = 0.888			
R²		0.07					0.34					0.34			
Adjusted R²		0.05					0.32					0.32			
N		189					189					189			

The number of previous court hearings was measured on a five-point scale (1 = 0, 2 = 1, 3 = 2–10, 4 = 11–20, 5 = more than 20).

TABLE 12 | Regression results for procedural justice, outcome judgments, and their interaction on protest intentions.

Variable	Step 1				95% CI for b	Step 2				95% CI for b	Step 3				95% CI for b
	b	β	t	p		b	β	t	p		b	β	t	p	
Legal assistance (0 = no, 1 = yes)	0.16	0.04	0.48	0.630	−0.50, 0.82	−0.08	−0.02	−0.30	0.763	−0.61, 0.45	−0.09	−0.02	−0.32	0.751	−0.62, 0.45
Previous hearings	−0.03	−0.02	−0.20	0.842	−0.32, 0.26	−0.02	−0.01	−0.16	0.875	−0.25, 0.22	−0.02	−0.01	−0.15	0.882	−0.25, 0.22
Age	−0.25	−0.13	−1.71	0.088	−0.55, 0.04	−0.16	−0.08	−1.31	0.192	−0.39, 0.08	−0.15	−0.08	−1.26	0.210	−0.39, 0.09
Procedural justice						−0.31	−0.15	−2.09	0.038	−0.61, −0.02	−0.32	−0.16	−2.08	0.039	−0.62, −0.02
Outcome judgments						−1.01	−0.50	−7.02	0.000	−1.30, −0.73	−1.01	−0.50	−7.01	0.000	−1.30, −0.73
Outcome judgments × Procedural justice											−0.03	−0.01	−0.19	0.849	−0.29, 0.24
df		3					5					6			
F		1.21, p = 0.308					22.47, p = 0.000					18.63, p = 0.000			
F change		1.21, p = 0.308					53.33, p = 0.000					0.04, p = 0.849			
R²		0.02					0.38					0.38			
Adjusted R²		0.00					0.36					0.36			
N		190					190					190			

The number of previous court hearings was measured on a five-point scale (1 = 0, 2 = 1, 3 = 2–10, 4 = 11–20, 5 = more than 20).

TABLE 13 | Regression results for procedural justice, outcome judgments, and their interaction on self-esteem.

Variable	Step 1					Step 2					Step 3				
	b	β	t	p	95% CI for b	b	β	t	p	95% CI for b	b	β	t	p	95% CI for b
Legal assistance (0 = no, 1 = yes)	0.05	0.02	0.26	0.793	−0.31, 0.40	0.13	0.05	0.72	0.474	−0.22, 0.48	0.12	0.05	0.69	0.492	−0.23, 0.48
Previous hearings	−0.14	−0.13	−1.78	0.077	−0.30, 0.02	−0.14	−0.13	−1.74	0.083	−0.29, 0.02	−0.14	−0.13	−1.72	0.086	−0.29, 0.02
Age	0.03	0.03	0.39	0.696	−0.13, 0.19	0.01	0.01	0.18	0.857	−0.14, 0.17	0.02	0.02	0.22	0.826	−0.14, 0.18
Procedural justice						0.19	0.18	1.96	0.051	−0.00, 0.39	0.19	0.17	1.85	0.066	−0.01, 0.38
Outcome judgments						0.09	0.08	0.92	0.360	−0.10, 0.28	0.09	0.08	0.91	0.363	−0.10, 0.28
Outcome judgments × Procedural justice											−0.02	−0.02	−0.26	0.793	−0.20, 0.15
df	3					5					6				
F	1.06, p = 0.366					2.72, p = 0.022					2.26, p = 0.039				
F change	1.06, p = 0.366					5.12, p = 0.007					0.07, p = 0.793				
R²	0.02					0.07					0.07				
Adjusted R²	0.00					0.04					0.04				
N	189					189					189				

The number of previous court hearings was measured on a five-point scale (1 = 0, 2 = 1, 3 = 2–10, 4 = 11–20, 5 = more than 20).

between the two terms difficult to assess. For example, after having received a negative case outcome, some defendants may have re-evaluated their sense of procedural fairness because of their need for self-enhancement. Thus, part of what might be going on in our analyses could be motivated reasoning regarding procedural fairness once defendants received a negative case outcome. This would fit with findings reported by Lilly and Wipawayangkool (2018) obtained in a non-courtroom setting. They found that external self-serving bias and self-threat following unfavorable outcomes were negatively related to procedural justice perceptions. Thus, studies using experimental control can clarify issues of causality and are therefore a viable avenue for future research into the issues examined here.

Practical Implications

While recognizing these limitations, we think the findings of our study can have some important practical implications. Trust in judges, for instance, is an issue that has the Dutch judiciary's ongoing attention. Although the level of trust in the Dutch judiciary is relatively high compared to trust in other Dutch governmental institutions and judiciaries in other European countries (Ridder et al., 2019; Bovens, 2020), safeguarding this trust is considered important (Grimmelikhuijsen, 2018). Fair procedures, in terms of objective legal standards as well as people's subjective perceptions, can play an important role in this regard. This is relevant not only with a view to maintaining and possibly increasing trust in judges as an end in itself, but also because trust in judges is related to other important attitudes and behaviors, such as perceived legitimacy and compliance with the law (Grootelaar and van den Bos, 2018).

Our finding that perceived procedural justice is negatively associated with protest intentions can be of interest to legal policymakers and judges as well. Although reporting protest intentions is not the same as actually appealing a verdict, the two are likely to be related. It is noteworthy in this regard that more than 90% of appeals to criminal verdicts are initiated by defendants (Croes, 2016). Promoting procedural justice could therefore be a way to decrease the number of appeals and the social costs associated therewith. These social costs may concern not only financial costs but also costs in terms of quality of adjudication, as judges' workload is considered a threat for impartial adjudication by one out of five Dutch judges (Weijers, 2019). Taken together, we think our findings regarding the importance of procedural justice in the criminal courtroom context are relevant for both their contribution to procedural justice theory and their possible implications for legal practice.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://doi.org/10.24416/UU01-74OU94.

ETHICS STATEMENT

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Anderson, J. K., and Hayden, R. M. (1981). Questions of validity and drawing conclusions from simulation studies in procedural justice: a comment. *Law Soc. Rev.* 15, 293–303. doi: 10.2307/3053606

Andriessen, I., Dijkhof, J. H., van der Torre, A., van den Berg, E., Pulles, I., Iedema, J., et al. (2020). Ervaren discriminatie in Nederland II. Retrieved from the Netherlands Institute for Social Research. Available at: http://www.scp.nl (Accessed May 14, 2020).

Ansems, L. F. M. (2021). Procedural justice on trial: a critical test of perceived procedural justice from the perspective of criminal defendants. dissertation. Utrecht University. Ridderprint: Alblasserdam, Utrecht University.

Ansems, L. F. M., van den Bos, K., and Mak, E. (2020). Speaking of justice: a qualitative interview study on perceived procedural justice among defendants in Dutch criminal cases. *Law Soc. Rev.* 54, 643–679. doi: 10.1111/lasr.12499

Berrey, E., Hoffman, S. G., and Nielsen, L. B. (2012). Situated justice: a contextual analysis of fairness and inequality in employment discrimination litigation. *Law Soc. Rev.* 46, 1–36. doi: 10.1111/j.1540-5893.2012.00471.x

Bovens, M. (2020). "Het gezag van de juristen: Een empirische verkenning," in *Handelingen Nederlandse Juristen-Vereniging: De toekomst van de jurist, de jurist van de toekomst.* eds. A. Berlee, M. Bovens, J. S. Buiting, A. R. Mackor, E. Mak, J. Silvis and E. T. T. Tai (Deventer: Wolters Kluwer), 31–55.

Brockner, J. (2010). *A Contemporary Look at Organizational Justice: Multiplying Insult Times Injury.* New York, NY: Routledge.

Brockner, J., de Cremer, D., Fishman, A. Y., and Spiegel, S. (2008). When does high process fairness reduce self-evaluations following unfavorable outcomes? The moderating effect of prevention focus. *J. Exp. Soc. Psychol.* 44, 187–200. doi: 10.1016/j.jesp.2007.03.002

Brockner, J., Heuer, L., Magner, N., Folger, R., Umphress, E., van den Bos, K., et al. (2003). High procedural fairness heightens the effect of outcome favorability on self-evaluations: an attributional analysis. *Organ. Behav. Hum. Decis. Process.* 91, 51–68. doi: 10.1016/S0749-5978(02)00531-9

Brockner, J., and Wiesenfeld, B. M. (1996). An integrative framework for explaining reactions to decisions: interactive effects of outcomes and procedures. *Psychol. Bull.* 120, 189–208. doi: 10.1037/0033-2909.120.2.189

Brockner, J., Wiesenfeld, B. M., and Diekmann, K. A. (2009). Towards a "fairer" conception of process fairness: why, when and how more may not always be better than less. *Acad. Manag. Ann.* 3, 183–216. doi: 10.5465/194165 20903047228

Casper, J. D., Tyler, T., and Fisher, B. (1988). Procedural justice in felony cases. *Law Soc. Rev.* 22, 483–507. doi: 10.2307/3053626

Cohen, R. L. (1982). "Perceiving justice: An attributional perspective," in *Equity and justice in social behavior.* eds. J. Greenberg and R. L. Cohen (New York: Academic Press), 119–160.

Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2003). *Applied Multiple Regression/Correlational Analysis for the Behavioral Sciences. 3rd Edn.* Mahwaw, NJ: Erlbaum.

Crocker, J., and Major, B. (1989). Social stigma and self-esteem: the self-protective properties of stigma. *Psychol. Rev.* 96, 608–630. doi: 10.1037/0033-295X.96.4.608

Croes, M. T. (2016). De staat van het hoger beroep in het strafrecht. Factsheet Retrieved from the Research and Documentation Centre. Available at: http://www.wodc.nl (Accessed May 14, 2020).

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146

Folger, R., Rosenfield, D., Grove, J., and Corkran, L. (1979). Effects of "voice" and peer opinions on responses to inequity. *J. Pers. Soc. Psychol.* 37, 2253–2261. doi: 10.1037/0022-3514.37.12.2253

Gilliland, S. W. (1994). Effects of procedural and distributive justice on reactions to a selection system. *J. Appl. Psychol.* 79, 691–701. doi: 10.1037/0021-9010.79.5.691

Grimmelikhuijsen, S. (2018). Van gegeven naar verdiend gezag: Hoe kan transparantere rechtspraak (blijvend) bijdragen aan legitimiteit? *Rechtstreeks* 15, 13–35.

Grootelaar, H. A. M. (2018). Interacting with procedural justice in courts. *Dissertation,* Amsterdam: Off Page.

Grootelaar, H. A. M., and van den Bos, K. (2018). How litigants in Dutch courtrooms come to trust judges: The role of perceived procedural justice, outcome favorability, and other socio-legal moderators. *Law Soc. Rev.* 52, 234–268. doi: 10.1111/lasr.12315

Hayden, R. M., and Anderson, J. K. (1979). On the evaluation of procedural systems in laboratory experiments: a critique of Thibaut and Walker. *Law Hum. Behav.* 3, 21–38. doi: 10.1007/BF01039147

Henrich, J., Heine, S. J., and Ara Norenzayan, A. (2010). The weirdest people in the world? *Behav. Brain Sci.* 33, 61–83. doi: 10.1017/S0140525X0999152X

Higgins, G. E., and Jordan, K. L. (2005). Race and gender: an examination of the models that explain evaluations of the court system for differences. *Crim. Justice Stud.* 18, 81–97. doi: 10.1080/14786010500071188

Holmvall, C. M., and Bobocel, D. R. (2008). What fair procedures say about me: Self-construal and reactions to procedural fairness. *Organ. Behav. Hum. Decis. Process.* 105, 147–168. doi: 10.1016/j.obhdp.2007.09.001

Huijnk, W., and Andriessen, I. (2016). Integratie in zicht? De integratie van migranten in Nederland op acht terreinen nader bekeken. Retrieved from

the Netherlands Institute for Social Research Available at: http://www.scp.nl (Accessed May 14, 2020).

Huijnk, W., Dagevos, J., Gijsberts, M., and Iris Andriessen, I. (2015). Werelden van verschil. over de sociaal-culturele afstand en positie van migrantengroepen in Nederland. Retrieved from the Netherlands Institute for Social Research. Available at: http://www.scp.nl (Accessed May 14, 2020).

Hulst, L. (2017). Experimental legal studies on perceived procedural justice and trust in law and society. dissertation. Amsterdam: Off Page.

Jenness, V., and Calavita, K. (2018). It depends on the outcome: prisoners, grievances, and perceptions of justice. *Law Soc. Rev.* 52, 41–72. doi: 10.1111/lasr.12312

Johnson, D., Wilson, D. B., Maguire, E. R., and Lowrey-Kinberg, B. V. (2017). Race and perceptions of police: experimental results on the impact of procedural justice. *Justice Q.* 34, 1184–1212. doi: 10.1080/07418825.2017.1343862

Koper, G., van Knippenberg, D., Bouhuijs, F., Vermunt, R., and Wilke, H. (1993). Procedural fairness and self-esteem. *Eur. J. Soc. Psychol.* 23, 313–325. doi: 10.1002/ejsp.2420230307

Landis, J. M., and Goodstein, L. L. (1986). When is justice fair? An integrated approach to the outcome versus procedure debate. *Am. Bar Found. Res. J.* 11, 675–707. doi: 10.1111/j.1747-4469.1986.tb00262.x

Leary, M. R., and Terry, M. L. (2013). "Self-evaluation and self-esteem," in *Oxford Handbook of Social Cognition*. ed. D. Carlston (New York: Oxford University Press), 534–547.

Leung, K., Steven, S., and Morris, M. W. (2001). When is criticism not constructive? The roles of fairness perceptions and dispositional attributions in employee acceptance of critical supervisory feedback. *Hum. Relat.* 54, 1155–1187. doi: 10.1177/0018726701549002

Lilly, J. D., and Wipawayangkool, K. (2018). When fair procedures don't work: a self-threat model of procedural justice. *Curr. Psychol.* 37, 680–691. doi: 10.1007/s12144-016-9555-7

Lind, E. A., Kulik, C. T., Ambrose, M., and de Vera, M. V. Park (1993). Individual and corporate dispute resolution: using procedural fairness as a decision heuristic. *Adm. Sci. Q.* 38, 224–251. doi: 10.2307/2393412

Lind, E. A., and Tyler, T. R. (1988). *The Social Psychology of Procedural Justice.* New York: Plenum.

Major, B. (1994). "From social inequality to personal entitlement: the role of social comparisons, legitimacy appraisals, and group membership," in *Advances in Experimental Social Psychology. Vol. 26.* ed. M. P. Zanna (San Diego: Academic Press), 293–355.

Major, B., Quinton, W. J., and McCoy, S. K. (2002). "Antecedents and consequences of attributions to discrimination: theoretical and empirical advances," in *Advances in Experimental Social Psychology. Vol. 34.* ed. M. P. Zanna (New York: Academic Press), 251–330.

McCoy, S. K., and Major, B. (2003). Group identification moderates emotional responses to perceived prejudice. *Personal. Soc. Psychol. Bull.* 29, 1005–1017. doi: 10.1177/0146167203253466

Paternoster, R., Brame, R., Bachman, R., and Sherman, L. W. (1997). Do fair procedures matter? The effect of procedural justice on spouse assault. *Law Soc. Rev.* 31, 163–204. doi: 10.2307/3054098

Ployhart, R. E., Ryan, A. M., and Bennett, M. (1999). Explanations for selection decisions: applicants' reactions to informational and sensitivity features of explanations. *J. Appl. Psychol.* 84, 87–106. doi: 10.1037/0021-9010.84.1.87

Ridder, D., Josje, E. M., Huijnk, W., and van Rijnberk, S. (2019). Continu onderzoek burgerperspectieven 2019/4. Retrieved from the Netherlands Institute for Social Research. Available at: http://www.scp.nl (Accessed May 14, 2020).

Rosenberg, M. (1965). *Society and the Adolescent Self-Image.* Princeton, NJ: Princeton University Press.

Schroth, H. A., and Shah, P. P. (2000). Procedures: do we really want to know them? An examination of the effects of procedural justice on self-esteem. *J. Appl. Psychol.* 85, 462–471. doi: 10.1037/0021-9010.85.3.462

Sedikides, C., Hart, C. M., and David de Cremer, D. (2008). The self in procedural fairness. *Soc. Personal. Psychol. Compass* 2, 2107–2124. doi: 10.1111/j.1751-9004.2008.00156.x

Stahl, T., Vermunt, R., and Ellemers, N. (2008). Reactions to outgroup authorities' decisions: the role of expected bias, procedural fairness, and outcome favorability. *Group Process. Intergroup Relat.* 11, 281–299. doi: 10.1177/1368430208090643

Statistics Netherlands (2018). Jaarrapport integratie 2018. Available at: https://www.cbs.nl (Accessed May 14, 2020).

Sunshine, J., and Tyler, T. R. (2003). The role of procedural justice and legitimacy in shaping public support for policing. *Law Soc. Rev.* 37, 513–548. doi: 10.1111/1540-5893.3703002

Thibaut, J., and Walker, L. (1975). *Procedural Justice: A Psychological Analysis.* Hillsdale, NJ: Erlbaum.

Thijssen, L., Coenders, M., and Lancee, B. (2019). Etnische discriminatie op de Nederlandse arbeidsmarkt. Verschillen tussen etnische groepen en de rol van beschikbare informatie over sollicitanten. *Mens en Maatschappij* 94, 141–176. doi: 10.5117/MEM2019.2.002.THIJ

Tyler, T. R. (1984). The role of perceived injustice in defendants' evaluations of their courtroom experience. *Law Soc. Rev.* 18, 51–74. doi: 10.2307/3053480

Tyler, T. R. (1988). What is procedural justice? Criteria used by citizens to assess the fairness of legal procedures. *Law Soc. Rev.* 22, 103–136. doi: 10.2307/3053563

Tyler, T. R. (2001). Public trust and confidence in legal authorities: what do majority and minority group members want from the law and legal institutions? *Behav. Sci. Law* 19, 215–235. doi: 10.1002/bsl.438

Tyler, T. R. (2006). *Why People Obey the Law.* New Haven: Yale University Press.

Tyler, T. R., and Huo, Y. J. (2002). *Trust in the Law: Encouraging Public Cooperation with the Police and Courts.* New York: Russell Sage Foundation.

Tyler, T. R., and Lind, E. A. (1992). "A relational model of authority in groups," in *Advances in Experimental Social Psychology, Vol. 25.* ed. M. P. Zanna (San Diego, CA: Academic Press), 115–191.

Van den Bos, K. (2005). "What is responsible for the fair process effect?" in *Handbook of Organizational Justice.* eds. J. Greenberg and J. A. Colquitt (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 273–300.

Van den Bos, K. (2015). "Humans making sense of alarming conditions: psychological insight into the fair process effect," in *Oxford Handbook of Justice in Work Organizations.* eds. R. S. Cropanzano and M. L. Ambrose (New York: Oxford University Press), 403–417.

Van den Bos, K., Bruins, J., Wilke, H. A. M., and Dronkert, E. (1999). Sometimes unfair procedures have nice aspects: on the psychology of the fair process effect. *J. Pers. Soc. Psychol.* 77, 324–336. doi: 10.1037/0022-3514.77.2.324

Van den Bos, K., and Lind, E. A. (2002). "Uncertainty management by means of fairness judgments," in *Advances in Experimental Social Psychology, Vol. 34.* ed. M. P. Zanna (San Diego, CA: Academic Press), 1–60.

Van den Bos, K., van der Velden, L., and Lind, E. A. (2014). On the role of perceived procedural justice in citizens' reactions to government decisions and the handling of conflicts. *Utrecht Law Rev.* 10, 1–26. doi: 10.18352/ulr.287

Walker, L., Stephen, L. T., Allan Lind, E., and Thibaut, J. (1974). Reactions of participants and observers to modes of adjudication. *J. Appl. Soc. Psychol.* 4, 295–310. doi: 10.1111/j.1559-1816.1974.tb02809.x

Weijers, J. (2019). De opbrengst van de rechterlijke macht. *Tréma* 1, 5–14.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychol. Rev.* 92, 548–573. doi: 10.1037/0033-295X.92.4.548

Williams, D. R., González, H. M., Williams, S., Mohammed, S. A., Moomal, H., and Stein, D. J. (2008). Perceived discrimination, race and health in South Africa: findings from the South Africa stress and health study. *Soc. Sci. Med.* 67, 441–452. doi: 10.1016/j.socscimed.2008.03.021

Williams, D. R., Yu, Y., Jackson, J. S., and Anderson, N. B. (1997). Racial differences in physical and mental health: socioeconomic status, stress, and discrimination. *J. Health Psychol.* 2, 335–351. doi: 10.1177/135910539700200305

Likeability and Expert Persuasion: Dislikeability Reduces the Perceived Persuasiveness of Expert Evidence

Mariam Younan* and Kristy A. Martire

School of Psychology, The University of New South Wales, Kensington, NSW, Australia

With the use of expert evidence increasing in civil and criminal trials, there is concern jurors' decisions are affected by factors that are irrelevant to the quality of the expert opinion. Past research suggests that the likeability of an expert significantly affects juror attributions of credibility and merit. However, we know little about the effects of expert likeability when detailed information about expertise is provided. Two studies examined the effect of an expert's likeability on the persuasiveness judgments and sentencing decisions of 456 jury-eligible respondents. Participants viewed and/or read an expert's testimony (lower vs. higher quality) before rating expert persuasiveness (via credibility, value, and weight), and making a sentencing decision in a Capitol murder case (death penalty vs. life in prison). Lower quality evidence was significantly less persuasive than higher quality evidence. Less likeable experts were also significantly less persuasive than either neutral or more likeable experts. This "penalty" for less likeable experts was observed irrespective of evidence quality. However, only perceptions of the foundational validity of the expert's discipline, the expert's trustworthiness and the clarity and conservativeness of the expert opinion significantly predicted sentencing decisions. Thus, the present study demonstrates that while likeability does influence persuasiveness, it does not necessarily affect sentencing outcomes.

Keywords: expert, juror decision-making, evidence evaluation, likeability, credibility, persuasion

INTRODUCTION

Expert evidence is ubiquitous in modern civil and criminal trials (Gross, 1991; Diamond, 2007; Jurs, 2016). Jurors involved in legal proceedings must assess the value of expert opinions to inform consequential decisions affecting lives and liberty. However, these assessments are sometimes mistaken, threatening the administration of justice, and contributing to unsafe trial outcomes (Innocence Project, 2021).

The Elaboration Likelihood Model (ELM) of persuasion is an information-processing model that has been used to understand jury decision-making about expert evidence (Petty and Cacioppo, 1986; McAuliff et al., 2003). This model suggests that jurors may struggle to accurately distinguish between low- and high-quality expert opinions because of the cognitive demands involved in the task (Petty and Cacioppo, 1986; Greene and Gordan, 2016). According to ELM, limited cognitive resources and insufficient knowledge increase reliance on readily accessible but potentially irrelevant, peripheral aspects of a message (Petty and Cacioppo, 1984, 1986; San José-Cabezudo et al., 2009; Salerno et al., 2017). This theory is supported by evidence suggesting that when information is unfamiliar, highly technical or complex—as is often the case for expert

opinions—juror evaluations of credibility and persuasiveness may be swayed by superficial features of the expert and their evidence (Chaiken, 1980; Heuer and Penrod, 1994; Shuman et al., 1994; Cooper et al., 1996; Schuller et al., 2005; Ivković and Hans, 2006; Daftary-Kapur et al., 2010; Bornstein and Greene, 2011; Neal, 2014; Maeder et al., 2016). Expert likeability is one peripheral cue that may affect perceptions of persuasiveness.

"Likeability" refers to the extent to which an expert presents as friendly, respectful, well-mannered, and warm (McAdams and Powers, 1981; Kerns and Sun, 1994; Levin et al., 1994; Gladstone and Parker, 2002; Neal and Brodsky, 2008; Brodsky et al., 2009, 2010; Neal et al., 2012). The likeability of the expert is a prominent social cue that is readily accessible to jurors. It is considered important because likeability increases juror connection, attention and receptiveness (McGaffey, 1979; Schutz, 1997), thereby fostering perceptions of credibility and merit (Chaiken, 1980; Brodsky et al., 2009; Neal et al., 2012). The importance of likeability for expert credibility assessment is supported by evidence that the Witness Credibility Scale (Brodsky et al., 2010) accounts for ∼70% of the observed variance in credibility using just four factors: likeability, confidence, knowledge, and trustworthiness. On its own likeability accounts for ∼7% of the variance within this model.

Although likeability is clearly not the sole determinant of jurors' credibility assessments, experimental research further supports the significance of likeability in expert persuasion. For example, Brodsky et al. (2009) presented mock jurors one of two videos of the testimony of an expert who was a licenced clinical psychologist, with an established private practice, 14 years of experience conducting over 100 forensic risk evaluations, and a history of providing expert testimony in over 50 cases. The only difference between the two videos was the level of expert likeability, which was manipulated to be either "low" or "high" using verbal and non-verbal cues such as smiling, body language and deferential speech. The results showed that the likeable expert was rated as more credible and trustworthy than the less likeable expert. Thus, the more likeable expert was more persuasive than a less likeable expert of the same quality.

Adapting the materials used by Brodsky et al. (2009) and Neal et al. (2012) examine the effect of likeability and expert knowledge on perceptions of persuasiveness. In their study, mock jurors watched the testimony of a high or low likeability expert who was either a "high knowledge" experienced clinical psychologist, or a "low knowledge" inexperienced general psychologist. The results showed that the more knowledgeable expert was more credible to jurors than the low knowledge expert. They also found that likeability had a consistent effect, boosting the credibility of both high and low knowledge experts. Taken together, these findings show that likeability does influence perceptions of expert credibility. Yet there is no evidence that a more likeable expert provides evidence that is more scientifically sound, logically coherent, or empirically justified than a less likeable expert (Chaiken, 1980; Petty and Cacioppo, 1986; Greene and Gordan, 2016). Thus, a reliance on likeability may misdirect or misinform juror evaluations and contribute to unjust trial outcomes, especially when a highly likeable expert provides a low-quality opinion. However, it is important to consider

the limitations of past research when assessing the potentially negative effects of expert likeability on juror assessments of credibility and persuasiveness.

To-date studies have typically conceptualised and manipulated expert evidence quality in simplistic ways, for example, using abridged trial vignettes, decontextualised expert extracts, and few or basic indicators of quality (e.g., years of experience or prestige of credentials; Petty et al., 1981; Swenson et al., 1984; Guy and Edens, 2003; McAuliff and Kovera, 2008; Brodsky et al., 2009; Neal et al., 2012; Parrott et al., 2015; Salerno et al., 2017). Given these somewhat impoverished materials, it is possible that the information that was available—including about likeability—may gain undue prominence in decision-making. Where peripheral cues are available, they may even "stand in" for useful but unavailable information (Petty and Cacioppo, 1986; Shuman et al., 1994; Sporer et al., 1995; Cooper et al., 1996; Ivković and Hans, 2006; Tenney et al., 2008). For example, there is evidence that likeability is used to make inferences about expert trustworthiness (Neal et al., 2012). Thus, it remains unclear how likeability may impact jurors' assessments of credibility and persuasiveness, when more realistic indicators of expertise are available to inform decision-making.

Another related limitation is the tendency to conflate expert evidence quality and likeability in experimental manipulations. For example, in previous studies, *likeability* manipulations also altered aspects of the evidence quality. Specifically, modest statements that acknowledge limited certainty and the potential for error used in studies such as Brodsky et al. (2009) and Neal et al. (2012) are generally considered to be higher quality than overstated conclusions that fail to acknowledge uncertainty (Koehler, 2012; Edmond et al., 2016). Thus, the influence of likeability on judgments of credibility might not have been entirely attributable to likeability, but rather, may partially be a response to differences in evidence quality. Consequently, it is unclear how influential peripheral cues such as likability are to credibility judgements when they are made in more realistic contexts where expert opinion quality is operationalised in more subtle and realistic ways.

Recent attempts to address this gap in the persuasion literature have used richer representations of expert opinion quality. Martire et al. (2020) operationalised expert opinion quality using the Expert Persuasion Expectancy (ExPEx) Framework. The ExPEx Framework specifies eight attributes that are logically relevant to the quality of an expert opinion: foundation, field, specialty, ability, opinion, support, consistency, and trustworthiness. *Foundation* refers to the empirical validity and reliability of the field in which the expert is opining (e.g., the discipline's error rate). *Field* relates to expert's training, study, and experience in an area generally relevant to their opinion (e.g., clinical psychology training). *Specialty* concerns whether the testifying expert has training, study or experience that is specifically relevant to the assertions they are making (e.g., risk assessment training). *Ability* relates to the expert's track record and their ability to form accurate and reliable opinions (e.g., personal proficiency). *Opinion* concerns the substantive opinion or judgment conveyed by the expert, its clarity, and the acknowledgement of limitations. *Support* concerns the presence

and quality of evidence underpinning the opinion (e.g., the results of psychometric testing). *Consistency* relates to the level of agreement amongst different suitable experts. *Trustworthiness* refers to the experts' conscientiousness, objectivity, and honesty.

When information about all ExPEx attributes was available to jury-eligible respondents, participants were more persuaded by objectively high- compared to low-quality forensic gait expert evidence (Martire et al., 2020). Jurors were also particularly influenced by information about the experts' track record (ability), their impartiality (trustworthiness), and the acceptability of their conclusion to other experts (consistency). However, the nuanced operationalisation of expert evidence quality used in this research did not extend to the use of realistic trial materials. Participants were merely presented an eight-statement description of the expert and their opinion and were not given any information about peripheral cues such as likeability. Thus, the influence of likeability on the assessment of expert evidence quality, especially in information-rich decision scenarios, remains unknown. Our research addresses this gap.

Across two studies, jury-eligible participants rated the persuasiveness of an expert opinion and provided a sentencing decision in a Capitol murder case after viewing and/or reading ExPEx-enriched high- or low-quality expert testimony. The materials were adapted from Neal et al. (2012) and Parrott et al. (2015) and included versions of the testimony from a high- or low-likeability expert (Study 1) with a neutral likeability control (Study 2).

In line previous research using the ExPEx framework, we expect that jurors will regard higher quality expert evidence as more persuasive than lower quality evidence, and that sentencing decision will be affected by evidence quality. We also expect that persuasiveness ratings will predict sentencing decisions. In addition, if as previously observed, likeability does influence perceptions of expert credibility and persuasiveness, then we would expect to find that more likeable experts are more persuasive than less likeable experts irrespective of evidence quality. However, if the previous effects of likeability were a result of the simplistic or confounded experimental materials rather than the persuasiveness of likeability *per se*, then we would not expect an effect of likeability because jurors will instead rely on the numerous valid quality indicators available in the trial scenario. Main effects of both quality and likeability, and any interactions between quality and likeability, would suggest that both likeability and indicators of evidence quality affect the persuasiveness of an expert opinion and/or sentencing decisions.

STUDY 1

Method
Design
Study 1 used a two (expert evidence quality: low, high) × two (likeability: low, high) between-subjects factorial design. Expert evidence quality was operationalised using either eight "high-quality" or eight "low-quality" ExPEx attributes. Low-vs. high-likeability was operationalised using the trial materials and verbal components from Neal et al. (2012). The primary dependent variables were persuasiveness rating and sentencing decision.

Persuasiveness was measured by averaging ratings of expert credibility, evidence value and evidence weight (all rated from 0 to 100). Sentencing decision was a binary choice between life in prison or death sentence per Neal et al. (2012). This study was pre-registered (AsPredicted#: 65017) and materials, data and analyses are available at [link for blind review to be updated if accepted: https://osf.io/yfgke/].

Participants
Participants were recruited from Amazon Mechanical Turk (MTurk). All participants resided in the United States and were aged 18 years or older. To maximise data quality, participation was limited to those who had not been involved in our similar studies and who had a 99% MTurk approval rating. Participants also completed attention checks and a reCAPTCHA to exclude non-human respondents (Von Ahn et al., 2008). Two-hundred and forty participants were recruited and were compensated US\$2.00 for their time. Participants who either failed the age check, were ineligible to serve on a jury or failed the attention checks ($n = 22$), were excluded from the final sample per our pre-registered exclusion criteria. The final sample consisted of 218 jury-eligible participants randomly allocated to condition as follows: high-quality, high-likeability: $n = 55$: high-quality, low-likability: $n = 57$; low-quality, high-likeability: $n = 50$; low-quality, low-likeability: $n = 56$.

Materials and Measures
Trial Materials
The trial materials used in this study were adapted from Neal et al. (2012) with the permission of the author. Departures from the original materials and procedures are specified below.

Pre-trial Instructions
Participants read written jury instructions indicating that the defendant had been found guilty of first-degree murder and that they were to return either a sentence of life in prison, or death, based on whether it could be shown "*beyond a reasonable doubt that there is a probability that the defendant would commit criminal acts of violence that would constitute a continuing danger in society*" (Neal et al., 2012). This jury instruction was adapted from the Texas Criminal Procedure Code, Article 370.071b-f (1985) by Krauss and Sales (2001).

Expert Evidence
The expert evidence transcript used by Neal et al. (2012) was based on an actual jury sentencing proceeding and portrayed the examination-in-chief and cross-examination of a forensic psychologist testifying about the likelihood that a convicted murderer would commit future violence (Krauss and Sales, 2001). The expert provided inculpatory evidence, ultimately stating that there is a "*high probability that he will commit future acts of dangerousness*" (Neal et al., 2012).

Participants were presented the original examination-in-chief and cross-examination of the expert used by Neal et al. (2012) without modifications. This transcript contained information about the experts' educational credentials, experience, method for conducting violence risk assessment, and their opinion about

the defendants' future risk of violence. This information related to the field, specialty, and support attributes in the ExPEx framework, which together formed the manipulation of expert "knowledge" (see Evidence Quality Manipulation for further detail). Ability, foundation and opinion were also addressed though in a limited way. Specifically, in all conditions, the expert had ultimately concluded that the defendant posed a *"continuing danger to society"* and that despite research showing clinical psychologists can be inaccurate, as far they knew, they had *"never been wrong"* in their evaluations.

To ensure that there was information available about all ExPEx attributes, a three-page 're-examination' was added to enrich the trial transcript. In this supplementary material participants were told that the prosecution and defence have recalled the expert for further testimony, were reminded of the jury instructions before reading the three new pages of written testimony. During the re-direct and cross-examination, the expert provided additional detail about their educational credentials, experience, methodology, and clarified their opinion. They also provided new information about the scientific basis for risk assessment (foundation), their own proficiency conducting risk assessments (ability), whether other experts agreed with their conclusions (consistency), and their track record working for the defence and prosecution (trustworthiness).

Evidence Quality Manipulation. All eight ExPEx attributes were manipulated in the transcript to produce either a high- or low-quality opinion (see OSF for evidence quality manipulations).

In the high-quality condition, participants read the materials developed by Neal et al. (2012) presenting the testimony of a clinical psychologist, educated at Yale, with a PhD, who was a Board-certified Forensic Psychologist with several academic publications in forensic risk assessment (field and specialty). The expert had 14 years of specialist training and experience in dangerousness and violence risk assessment and had used multiple clinical interviews totaling 15 h with the defendant to assess risk utilising the Violence Risk Assessment Guide (specialty and support). In the enriched re-examination, participants were also given information that the expert was highly proficient in conducting violence risk assessments, with an average performance of 90–94% accuracy (ability), that clinical psychology is a discipline that equips professionals to make accurate risk judgments, and that the V-RAG is an empirically supported and validated assessment method (foundation). The high-quality condition also read that the clinical psychologist managed the potential for bias in their opinion, had testified equally for the prosecution and defence, and did not know the defendant previously (trustworthiness). They also acknowledged the limits of their conclusion by suggesting that even experts do not always have perfect judgment and that risk assessments are not 100% accurate (opinion). The clinical psychologists' opinion was based on collateral information, interview and addressed the relevant risk factors (support). The opinion was also verified by independent experts in the same specialist field (consistency).

By contrast, adopting Neal et al. (2012) manipulations, participants in the low-quality condition read the testimony of a non-specialist psychologist, with 2 years of experience

as a psychotherapist in private practice, who did not provide their educational credentials (field), had no specialisation or experience in violence risk assessment (specialty) and had only completed a 30-min interview with the defendant before using the V-RAG (support). In the enriched re-examination, participants also learned that the psychotherapist had not had their risk assessment performance tested but nevertheless reported they were highly proficient (ability), relied on unvalidated clinical judgment and modified the VRAG to assess risk (foundation and support). Those in the low-quality condition also learned that the psychotherapist had worked mostly for the prosecution, had known the defendant previously, and had only considered information they considered relevant in their personal opinion (trustworthiness). They communicated no uncertainty or limitations around their conclusions when re-clarifying their expert judgment (opinion). The psychotherapist's opinion was based only on information from the 30-min interview, did not refer to empirical literature (support), and was verified by a law enforcement official who was also working on the same case rather than an expert in risk assessment (consistency).

Likeability Manipulation. Likeability was manipulated in the original materials using verbal cues (Neal et al., 2012; see OSF for likeability manipulations). The same high and low likeability manipulations were applied throughout the enriched re-examination transcript to ensure consistency throughout the scenario. These likeability manipulations have been shown to be successful at differentiating an expert high in likeability from an expert low in likeability (Brodsky et al., 2009; Neal et al., 2012).

In the high-likeability conditions, participants read a version of the expert who used terms such as "we" or "us" when referring to themselves or others, used informal speech (e.g., referring to an individual by name), was genuine, humble and deferential (e.g., commended the work of others), showed considerate and respectful disagreement, agreeableness to requests and questions (e.g., stating "of course" when asked to repeat something), and had a pleasant and friendly interpersonal style.

In the low-likeability conditions, participants read a version of the expert who used individualistic pronouns (i.e., I, me), was disingenuous, arrogant, and non-deferential (e.g., displayed superiority relative to others, was self-complimenting), showed aggressive contradiction and disagreement, disagreeableness in response to requests, and questions (e.g., pointing out repetitiveness and labelling questions as redundant), and had an unfriendly and condescending interpersonal style.

Primary Dependant Measures

Persuasiveness. The persuasiveness measure comprised three questions. Participants rated the credibility of the expert ("*how credible is Dr. Morgan Hoffman?*") from 0 "not at all" to 100 "definitely credible," the value of the expert's evidence ("*how valuable was Dr. Morgan Hoffman's testimony?*") from 0 "not at all" to 100 "definitely valuable," and the weight of the expert's evidence ("*how much weight do you give to Dr. Morgan Hoffman's testimony?*") on a scale from 0 "none at all" to 100 "the most possible." Question order was randomised. These items have been previously found to be highly correlated (all r's > 0.847) and

have high internal consistency (Cronbach's α = 0.954; Martire et al., 2020).

Sentencing Decision. Participants were asked "*Considering all the evidence provided to you, what is the sentence you would recommend for the defendant?*" and were required to answer either "I would recommend that the defendant receive a death sentence" or "I would recommend that the defendant receive a sentence of life in prison."

Secondary Measures & Manipulation Checks

ExPEx Attribute Ratings. Eight items were used to assesses decision-makers' perceptions of whether or not the expert opinion had a high-quality foundation, field, specialty, ability, opinion, support, consistency and trustworthiness from 0 "not at all" to 100 "definitely." Question order was randomised. See **Table 1** for verbatim wording and format.

Witness Credibility Scale. The Witness Credibility Scale (WCS) is a 20-item measure assessing expert credibility (Brodsky et al., 2010). Each item contains bipolar adjectives on a 10-point Likert scale [e.g., *not confident* (1) to *confident* (10)]. The presentation of the items was randomised. The highest possible score for overall credibility is 200, with higher scores indicating higher credibility ratings. The WCS also yields a sub-scale score for four credibility domains: knowledge, trustworthiness, confidence, and likeability. The highest possible score for each domain is 50, with a higher score indicating higher rating in a domain. The WCS has good validity and reliability—it can successfully differentiate between expert displaying varying levels of the four sub-domains (Brodsky et al., 2010). The WCS was included to provide an embedded measure of expert likeability. Participants were asked to rate the expert on the following bipolar adjectives: *unfriendly* (1) to *friendly* (10); *unkind* (1) to *kind* (10); *disrespectful* (1) to *respectful* (10); *ill-mannered* (1) to *well-mannered* (10) and *unpleasant* (1) to *pleasant* (10). Collectively, these items produced the WCS-Likeability subscale.

Agreement. Participants were asked "*If Dr. Hoffman reported that the defendant will commit a violent offence and pose a danger to society, would you agree with that opinion?*" and were required to answer either "yes" or "no." Analyses involving agreement are available on OSF.

Likeability. Participants were asked to "*rate how likeable Dr. Morgan Hoffman is to you, with zero being 'not at all likeable' and ten being 'extremely likeable'.*"

Expert Testimony Comprehension. Comprehension of the expert evidence was measured using 6 multiple-choice items to assess engagement with the testimony and understanding of its substantive content. Higher comprehension scores (out of 6) indicated greater recall and comprehension of the expert testimony. Analyses involving comprehension are available on OSF.

Demographic Information. Participants were asked to provide information about their age, gender, education level, cultural background, English proficiency, religiosity, political orientation, views of the death penalty, experience and familiarity with the expert's discipline and jury eligibility and experience.

Procedure

This study was approved by the UNSW Human Advisory Ethics Panel C—Behavioural Sciences (Approval #3308) and pre-registered. The study was advertised on MTurk and completed by participants online in Qualtrics. Before commencing the study, participants were asked to provide informed consent, complete age eligibility and reCAPTCHA, before random allocation to condition. Participants read the instructions and the version of the expert testimony transcript as determined by allocation to condition. Next, participants completed the ExPEx, WCS and likeability measure, in randomised order; completed the persuasiveness measure, and made their sentencing decision. Finally, participants completed the comprehension and demographic items. At the conclusion of the study, participants were given a completion code, were debriefed, and thanked. The average study completion time was 23.5 min.

Results

Participant Demographics

Participants were aged between 18 and 71 years (M = 39.36, SD =12.34) and 48.6% were male. Most participants reported that their highest level of completed education was college/university (54.1%) or a Masters degree (27.5%). Most identified as White/Caucasian (71.6%), followed by Asian (10.6%), African American (8.3%), and Hispanic (5%). Almost all participants (95.9%) were native English speakers.

About half of participants (53.2%) considered themselves more than "moderately" religious (on a 10-point scale from "not at all" to "very" religious). The largest proportion of participants (45.5%) rated themselves as conservative (on a 10-point scale from "very liberal" to "very conservative;" 42.2% were liberal; and 12.4% were neutral). The largest proportion of participants (47.7%) were in favour of the death penalty (on a 10-point scale from "strongly opposed" to "strongly in favor;" 45% were not in favour; 7.3% were neutral). A majority (61.5%) were unfamiliar with dangerousness and violence risk assessments ("none" to "some" familiarity), but approximately half (52.8%) reported being familiar with psychology/clinical psychology (from "some" to "extensive" familiarity). More than half of the sample (56.9%) had been called for jury duty; 46.8% of these participants had served on a jury, and 7.3% (n = 9) had served on a murder trial.

Manipulation Checks

All assumptions were tested before conducting the planned analyses. The analytic approach reported here either satisfies the relevant assumptions or is robust to violations.

Evidence Quality

A two-way (Pillai's Trace) MANOVA was conducted comparing the ratings of each of the eight ExPEx expert attributes between the low- and high-quality expert evidence conditions. There was a significant main effect of expert evidence quality overall [$F_{(8,207)}$ = 11.9, $p < 0.001$, $\eta p^2 = 0.315$] and for each attribute [all

TABLE 1 | Expert persuasion expectancy (ExPEx) quality items.

ExPEx Attribute	Question
Foundation	Does training, study, and experience in clinical psychology **support assertions** that a *defendant will commit a violent offence and pose a danger to society*?
Field	Does Dr. Morgan Hoffman have **training, study, and/or experience** in *clinical* psychology?
Specialty	Does Dr. Morgan Hoffman **have training, study and/or experience specific** to making assertions that *defendant will commit a violent offence and pose a danger to society*?
Ability	Does Dr. Morgan Hoffman make assertions that the *defendant will commit a violent offence and pose a danger to society* **accurately and reliably**?
Opinion	Did Dr. Morgan Hoffman convey their assertion that the *defendant will commit a violent offence and pose a danger to society* **clearly**, and with **necessary qualifications/limitations**?
Support	Did Dr. Morgan Hoffman **rely on evidence** when forming their assertion that the *defendant will commit a violent offence and pose a danger to society*?
Consistency	Is Dr. Morgan Hoffman's assertion that *defendant will commit a violent offence and pose a danger to society*? **consistent** with what other experts in clinical psychology would assert?
Trustworthiness	Do you believe that Dr. Morgan Hoffman is **fair, impartial,** and **objective**?

The bolded writing reflects the definitional component of the ExPEx attributes and the italicised writing reflects the key statement (i.e., the expert's conclusion) from which the ExPEx attribute is being rated in accordance with.

TABLE 2 | Table of marginal means and inferential statistics for expert persuasion expectancy (ExPEx) attributes by evidence quality condition.

ExPEx Attribute	ExPEx Attribute Rating							
	High-Quality Mean (SE)	95% CI	Low-Quality Mean (SE)	95% CI	*F*	*p*	η^2
Foundation	76.62 (2.18)	72.32, 80.92	60.70 (2.25)	56.28, 65.13	25.83	<0.001	0.108
Field	92.64 (1.82)	89.04, 96.23	72.86 (1.88)	69.17, 76.56	57.10	<0.001	0.211
Specialty	88.16 (1.99)	84.24, 92.09	63.08 (2.05)	59.04, 67.12	77.04	<0.001	0.265
Ability	78.93 (2.34)	74.32, 83.53	57.24 (2.41)	52.50, 61.98	41.82	<0.001	0.163
Opinion	82.87 (2.39)	78.16, 87.59	64.33 (2.46)	59.47, 69.18	29.18	<0.001	0.120
Support	76.48 (2.34)	71.87, 81.10	58.22 (2.41)	53.47, 62.97	29.56	<0.001	0.121
Consistent	78.74 (2.08)	74.64, 82.84	57.97 (2.14)	53.74, 62.19	48.40	<0.001	0.184
Trustworthy	73.07 (2.61)	67.93, 78.21	51.05 (2.69)	45.76, 56.34	34.59	<0.001	0.139

$F's_{(1, 214)} \geq 25.83$, all $p's < 0.001$, all $\eta p^2 \geq 0.108$] such that, on average, participants in the high-quality condition rated each ExPEx attribute as higher quality than those in the low-quality condition (see **Table 2**).

Likeability

Independent samples Welch *t*-tests showed a significant difference between high- and low-likeability experts on the WCS-likeability sub-scale score [$t_{(163.79)} = -10.99$, 95% CI (-20.78, -14.45), $p < 0.001$]. On average participants in the high-likeability condition rated the expert 39.3 out of 50 (SD = 7.3) compared to 21.7 out of 50 (SD = 15.3) in the low-likeability condition.

The single item rating subjective likeability was strongly and positively correlated with the WCS-Likeability subscale score ($r = 0.921$, $p < 0.001$). Accordingly, we report all subsequent analyses using the validated WCS-Likeability scores rather than the single likeability item.

Persuasiveness Ratings

Consistent with Martire et al. (2020), ratings of expert credibility, value and weight were all strongly and positively correlated

($r_{credibility/weight} = 0.905$; $r_{credibility/value} = 0.894$; $r_{value/weight} = 0.913$, all $p's < 0.001$), and had high internal consistency (Cronbach's $\alpha = 0.966$), so were combined into a single measure of persuasiveness.

Effect of Expert Evidence Quality and Likeability on Persuasiveness

Average persuasiveness ratings by condition are shown in **Figure 1**. The mean persuasiveness ratings by condition were: high-quality, high-likeability $M = 84.5$ (SD = 11): high-quality, low-likability $M = 72.8$ (SD = 24.2); low-quality, high-likeability $M = 62.7$ (SD= 24.4); low-quality, low-likeability $M = 47.9$ (SD = 32). A two-way ANOVA showed a significant main effect of evidence quality [$F_{(1, 214)} = 50.76$, $p < 0.001$, $\eta p^2 = 0.192$] and a significant main effect of likeability [$F_{(1, 214)} = 16.39$, $p < 0.001$, $\eta p^2 = 0.071$]. The high-quality expert was significantly more persuasive than the low-quality expert. The high-likeability expert was also more persuasive than the low-likeability expert. There was no significant interaction between evidence quality and likeability indicating that the effect of

FIGURE 1 | Persuasiveness as a function of expert evidence quality and likeability (Study 1). Figure depicts four raincloud plots showing the distribution of persuasiveness ratings observed in each condition. From left to right, each raincloud plot shows the: jittered individual data points, box-and-whisker plots (middle bar within the box is the median, the box represents the interquartile range of persuasiveness ratings, and the whiskers represent persuasiveness ratings no further than 1.5× the interquartile range), and the distributions showing the frequency of persuasiveness ratings. Mean persuasiveness ratings differed by evidence quality and likeability conditions.

likeability was consistent for high- and low-quality evidence [$F_{(1, 214)} = 0.234$, $p = 0.629$, $\eta p^2 = 0.001$].

Multiple regressions were conducted to examine whether continuous subjective ratings of the eight ExPEx attributes and WCS-likeability predicted persuasiveness ratings. The overall model was significant [$F_{(9, 208)} = 111.06$, $p < 0.001$] and accounted for 82% of the variance in persuasiveness ratings (adjusted $R^2 = 0.82$). Ratings of trustworthiness, consistency, support, ability, and specialty were all significant independent predictors (all p's ≤ 0.022), while foundation, field, opinion, and likeability were not (all p's ≥ 0.054; see **Table 3**). For example, holding all else constant, a one unit increase in perceptions of the trustworthiness of the expert was associated with a 0.347 unit increase in persuasiveness ratings.

Relationship Between Persuasiveness and Sentencing Decision

A binominal logistic regression was used to examine the relationship between persuasiveness and sentencing decision. The overall model was a good fit and significant [$\chi^2_{(1)} = 56.14$, $p < 0.001$]. Persuasiveness accounted for 31.7% of the variance in sentencing decision [Nagelkerke $R^2 = 0.317$; Wald $\chi^2_{(1)} = 29.79$, $p < 0.001$], with a one unit increase in persuasiveness increasing the odds of the decision-maker choosing a death sentence by 1.063 (Exp B).

Effect of Expert Evidence Quality and Likeability on Sentencing Decision

The proportion of participants giving death sentences by condition is shown in **Table 4**. A binominal logistic regression was used to predict sentencing decision from expert quality condition, likeability condition, and their interaction. The overall model was a good fit but not significant [$\chi^2_{(3)} = 6.84$, p

= 0.077] and accounted for only 4.3% of the variance in sentencing decision (Nagelkerke $R^2 = 0.043$). Neither expert evidence quality, likeability, nor their interaction were significant independent predictors of sentencing decision (all p's ≥ 0.158; see **Table 5**).

Another binominal logistic regression conducted to examine whether subjective continuous ExPEx ratings and WCS-likeability scores predicted sentencing decision. The overall model was a good fit and was significant [$\chi^2_{(9)} = 53.35$, $p < 0.001$], accounting for 30.4% of the variance in sentencing decision (Nagelkerke $R^2 = 0.304$). Ratings of foundation ($p = 0.022$) and trustworthiness ($p = 0.004$) uniquely predicted sentencing decision, while the remaining ExPEx attributes and likeability scores did not (all p's ≥ 0.104; see **Table 6**).

Discussion

Study 1 examined whether expert quality and likeability affected jury-eligible participants' perceptions of expert persuasiveness and sentencing decisions. We found that participants' perceptions of persuasiveness were significantly affected by evidence quality and expert likeability whereby higher quality and higher likeability experts were more persuasive than lower quality and lower likeability experts. However, there was no interaction between evidence quality and likeability. We also found that subjective perceptions of the eight ExPEx attributes and likeability together accounted for ∼80% of the variance in persuasiveness scores, which demonstrates these attributes have strong predictive power.

These results suggest that evidence quality impacts understanding, but persuasiveness is determined by both the underlying quality of the evidence, and superficial aspects of the experts' interpersonal style. Our results also suggest that previously observed effects of likeability on perceptions of credibility or persuasiveness were not merely an artefact of simplified evidence quality materials and manipulations. The expert evidence presented in this study was detailed and included extensive information about the quality of the opinion, yet the effect of likeability persisted and appeared to provide a boost to the persuasiveness of both lower and higher quality evidence. Thus, concerns about juror reliance on peripheral information in their decision-making remain.

However, it is important to note that neither likeability nor quality affected sentencing decisions in the same way that they affected persuasiveness. There were no significant associations between evidence quality or likeability conditions on sentencing outcomes. Continuous subjective likeability ratings also did not predict sentencing outcome, but perceptions of expert trustworthiness and foundation did. Thus, although likeability affected perceptions of persuasiveness, and persuasiveness affected sentencing outcomes, likeability did not directly affect the final sentencing outcome. This was not the case for evidence quality—elements of which remained influential for both evidence evaluation and sentencing decisions. Taken together, this suggests that lay decision-makers consider elements of expert evidence quality more so than peripheral likeability information when making their sentencing decisions.

TABLE 3 | Multiple regression predicting persuasiveness from continuous expert persuasiveness expectancy (ExPEx) ratings and witness credibility score (WCS) for likeability.

		95% CI for B						
	B	**Lower**	**Upper**	**SE B**	**β**	**p**	**R²**	**Adj. R²**
Model						<0.001	0.828	0.82
Foundation	−0.048	−0.143	0.047	0.048	−0.042	0.324		
Field	0.000	−0.097	0.098	0.05	0.000	0.993		
Specialty	0.18***	0.071	0.288	0.055	0.161***	0.001		
Ability	0.237***	0.141	0.334	0.049	0.232***	<0.001		
Opinion	0.021	−0.063	0.105	0.042	0.021	0.619		
Support	0.148***	0.069	0.226	0.04	0.142***	<0.001		
Consistent	0.119*	0.017	0.221	0.052	0.107*	0.022		
Q0Trustworthy	0.347***	0.254	0.44	0.047	0.395***	<0.001		
WCS-Likeability	0.15	−0.002	0.303	0.077	0.081	0.054		

B, unstandardised regression coefficient; CI, confidence interval; SE B, standard error of the coefficient; β, standardised coefficient. *p < 0.05, **p < 0.01, ***p ≤ 0.001.

These results raise further questions that should be explored. First, it is important to establish whether these effects are reliable by attempting to replicate the results. It is also important to consider whether our results are generalisable, especially given the lower ecological validity of trial transcript studies. Perceptions of likeability are strongly affected by non-verbal cues such as smiling, nodding, eye contact and open posture (Kleinke, 1986; Leathers, 1997; Gladstone and Parker, 2002). These cues were not available in our materials. Thus, it is important to examine whether the effects of expert likeability are replicated when more realistic video manipulations of likeability are used. Finally, we were interested to inform our general understanding of the relationship between likeability and persuasion by considering likeability's directional impact on persuasiveness. It is unclear whether being likeable *increases* persuasiveness, or if it is being *dis*liked that *decreases* persuasiveness, or both. Study 2 is designed to tease apart these possibilities.

STUDY 2

Method

Design

Study 2 used a 2 (expert evidence quality: high, low) × 3 (likeability: neutral, low, high) between-subjects factorial design. The dependent variables and evidence quality manipulations were the same as Study 1. Details about the likeability manipulations are described below. Study 2 was pre-registered (AsPredicted#: 39310) and material, data and analyses are available at [blind link to OSF: https://osf.io/yfgke/].

Participants

Participants were recruited using two methods: (1) online via MTurk, with the same quality assurance methods as in Study 1, and (2) via the UNSW first-year psychology undergraduate student pool. Research suggests that online and undergraduate participant samples are generally comparable and there is little evidence of significant differences in the decisions made between

TABLE 4 | Proportion of participants selecting death sentence by evidence quality and likeability condition.

Likeability	Low-Quality Evidence %	High-Quality Evidence %
Low	26.8	43.9
High	22	34.5

student and non-student samples in mock jury decision-making research (Bornstein and Greene, 2011; Buhrmester et al., 2011).

MTurk participants were compensated $US4.00 for their time, while the first-year psychology students received course credit. Participants who did not consent, failed the audio check, failed the attention checks, or were ineligible to serve on a jury (n = 110) were excluded from the final sample as per pre-registered exclusion criteria. The final sample consisted of 238 jury-eligible participants (164 from MTurk and 74 from the undergraduate pool), randomly allocated to condition as follows: high-quality, neutral likeability n = 39; high-quality, high-likeability n = 44: high-quality, low-likability n = 37; low-quality, neutral likeability n = 37; low-quality, high-likeability n = 44; low-quality, low-likeability n = 37.

Materials and Measures

The same trial scenario, expert evidence quality manipulations, and manipulation checks from Study 1 were used in Study 2 except as described below. Changes were made to the likeability manipulation to incorporate the new neutral likeability condition.

Expert Quality and Likeability

To increase realism of the likeability manipulation the transcript of the examination-in-chief and cross-examination of the high- and low-quality and high- and low-likeability expert evidence was replaced with video reenactments also produced by Neal et al. (2012). In these videos participants saw a White middle-aged male providing testimony from a courtroom, with a US

TABLE 5 | Logistic regression predicting sentencing decision from evidence quality condition, likeability condition, and their interaction.

	B	SE	Wald	*df*	*p*	Odds Ratio	95% CI for Odds Ratio	
							Lower	Upper
Evidence quality	−0.63	0.44	1.99	1	0.158	0.53	0.22	1.28
Likeability	0.39	0.39	1.01	1	0.314	1.48	0.69	3.18
Evidence quality* likeability	−0.13	0.6	0.05	1	0.826	0.88	0.27	2.84

*Life in prison was coded as zero and death was coded as one. B, unstandardised regression coefficient; SE B, standard error of the coefficient; CI, confidence interval. *p < 0.05, **p < 0.01, ***p ≤ 0.001.*

TABLE 6 | Logistic regression predicting sentencing decision from continuous expert persuasion expectancy (ExPEx) ratings and witness credibility score (WCS) for likeability.

	B	SE	Wald	*df*	*p*	Odds Ratio	95% CI for Odds Ratio	
							Lower	Upper
Foundation*	0.028*	0.012	5.25	1	0.022	1.03	1.00	1.05
Field	−0.009	0.012	0.6	1	0.439	0.99	0.97	1.01
Specialty	−0.004	0.014	0.095	1	0.758	0.996	0.97	1.02
Ability	0.002	0.013	0.03	1	0.869	1.00	0.98	1.03
Opinion	0.016	0.012	1.83	1	0.177	1.02	0.99	1.04
Support	<0.001	0.008	0.001	1	0.974	1	0.98	1.02
Consistent	−0.006	0.012	0.293	1	0.588	0.99	0.97	1.02
Trustworthy	0.034**	0.012	8.46	1	0.004	1.03	1.01	1.06
WCS-Likeability	−0.027	0.017	2.64	1	0.104	0.97	0.94	1.01

*Life in prison was coded as zero and death was coded as one. B, unstandardised regression coefficient; SE B, standard error of the coefficient; CI, confidence interval. *p < 0.05, **p < 0.01, ***p ≤ 0.001.*

TABLE 7 | Table of marginal means and inferential statistics for expert persuasion expectancy (ExPEx) attributes by evidence quality condition.

ExPEx Attribute	ExPEx Attribute Rating				*F*	*p*	η^2
	High-Quality Mean (SE)	95% CI	Low-Quality Mean (SE)	95% CI			
Foundation	74.79 (2.03)	70.79, 78.78	65.51 (2.05)	61.47, 69.54	10.37	0.001	0.043
Field	88.75 (1.78)	85.25, 92.25	76.85 (1.79)	73.32, 80.38	22.27	<0.001	0.088
Specialty	84.72 (2.04)	80.69, 88.74	66.46 (2.06)	62.40, 70.52	39.59	<0.001	0.146
Ability	75.88 (2.23)	71.48, 80.28	62.23 (2.25)	57.79, 66.66	18.54	<0.001	0.074
Opinion	76.80 (2.24)	72.38, 81.22	62.20 (2.26)	61.74, 70.66	11.06	0.001	0.045
Support	71.66 (2.32)	67.09, 76.22	62.01 (2.34)	57.41, 66.61	8.6	0.004	0.036
Consistent	76.06 (2.07)	71.98, 80.13	62.37 (2.09)	58.26, 66.48	21.71	<0.001	0.086
Trustworthy	73.66 (2.41)	68.91, 78.40	57.37 (2.43)	52.58, 62.15	22.67	<0.001	0.089

flag in the background. The videos were between 4.5 and 6 min long and displayed the same actor to control for between-person characteristics (i.e., attractiveness). In addition to the verbal likeability cues from Study 1, participants in the high likeability condition saw an expert who showed moderate levels of smiling, consistent eye contact, open body language and a modest presentation style. Those in the low likeability condition saw an expert who did not smile, had inconsistent eye contact, closed body language and a conceited presentation style. These videos were followed by the same ExPEx enriched transcript developed for Study 1.

The materials for the new neutral likeability condition were presented in transcript format to minimise all visual likeability

cues (e.g., smiling, eye contact). Participants in this condition read a transcribed version of the same examination-in-chief and cross-examination video. The transcript was developed by Parrott et al. (2015) and removed or neutralised the likeability cues contained in the original Neal et al. (2012) materials. For example, phrases such as "I take this responsibility very seriously," "of course" or "feeble-minded people think they know everything" were removed leaving only the essential substantive content. Participants also read a neutral version of the enriched transcript stripped of the likeability cues added for Study 1.

The manipulation checks, dependent and secondary measures were the same as in Study 1, except for the comprehension measures which were modified to reflect the testimony as

presented in Study 2. Participants also completed the Scientific Reasoning Scale (Drummond and Fischhoff, 2017) and Need for Cognition measure (Cacioppo and Petty, 1982), however analysis of these data was beyond the scope of this study and so is not reported here.

Procedure

This study was approved by the UNSW Human Advisory Ethics Panel C—Behavioural Sciences (Approval #3308) and pre-registered. The study was advertised on MTurk and undergraduate recruitment system and was completed online by all participants in Qualtrics. Before commencing the study, participants were asked to provide informed consent, complete age eligibility, a reCAPTCHA and were randomly allocated to condition. Participants read the study instructions, completed an audiovisual check, and watched/read the version of the expert testimony as determined by quality and likeability condition. Next, participants completed the ExPEx, WCS and likeability measures, in randomised order. Participants then completed the persuasiveness measures and made their sentencing decision. Finally, all participants completed the comprehension items, attention checks, secondary measures, and demographic questions. At the conclusion of the study, participants were given a completion code, were debriefed, and thanked. The average study completion time was 41 min.

Results

Before conducting the planned analyses, the assumptions were tested for all the statistical procedures employed and were robust. Initial analyses were conducted separately for undergraduate and MTurk participants. The results for these two groups varied in minor ways due to the disparate sample sizes but were broadly consistent, so we present the combined analysis here. Data and the primary persuasion analysis for each sample are available on OSF.

Participant Demographics

Overall, participants were aged between 18 and 72 years (M = 31.8, SD =12.2) and 51.3% were male. Most participants reported that college/university (43.3%) or high/secondary school (37.4%) was their highest level of completed education. Most participants identified as White/Caucasian (69.7%), followed by Asian (13.9%), African American (5.5%), and Other (4.6%). Almost all participants (95.4%) were native English speakers.

About half of participants (52.5%) considered themselves more than "moderately" religious. The largest proportion of participants (45%) rated themselves as conservative (39.9% were liberal; 15.1% were neutral). Just over half of participants (52.1%) were against the death penalty (38.7% were in favour; 9.2% were neutral). About half (50.4%) reported they were unfamiliar with clinical psychology and two-thirds (64.3%) were unfamiliar with dangerousness and violence risk assessment. One-third (32.4%) of the sample had been called up for jury duty; 51.9% of these participants had served on a jury, and 8% (n = 3) had served on a murder trial.

Manipulation Checks

Evidence Quality

A two-way (Pillai's Trace) MANOVA was conducted comparing the ratings of each of the eight ExPEx attributes between the low- and high-quality expert evidence conditions. There was a significant main effect of expert evidence quality overall [$F_{(8,225)}$ = 5.46, $p < 0.001$, $\eta p^2 = 0.163$] and for each ExPEx attribute (all F's $_{(1, 232)} \geq 8.6$, all p's ≤ 0.004, all $\eta p^2 \geq 0.036$) such that, on average, participants in the high-quality expert evidence conditions rated each ExPEx attributes as higher quality compared to those in the low-quality condition (see **Table 7**).

Likeability

On average participants in the neutral likeability condition rated the expert 38 out of 50 for likeability (SD = 6.8) compared to 40 (SD = 6.7) in the high-likeability condition and 22.5 (SD = 13.5) in the low-likeability condition. A one-way (Welch) ANOVA showed a significant difference in likeability scores by condition [Welch's $F_{(2,142.321)}$ = 51.87, $p < 0.001$]. Follow-up Games-Howell comparisons showed that ratings in the neutral and high-likeability conditions did not differ from each other [M_{diff} high vs. neutral = 2.03, 95% CI (−0.48, 4.53), $p = 0.138$], though likeability was significantly higher in both these conditions than in the low likeability condition [M_{diff} neutral vs. low = 15.5, 95% CI (11.33, 19.67), $p < 0.001$; M_{diff} high vs. low] = 17.53, 95% CI (13.43, 21.63), $p <0.001$]. Thus, it appeared that adding visual and verbal likeability cues did not significantly increase likeability perceptions beyond the transcript. However, visual, and verbal cues to decrease likeability were effective.

Persuasiveness Ratings

Consistent with Study 1, ratings of credibility, weight and value were strongly and positively correlated (r credibility/weight = 0.849; r credibility/value = 0.879; r value/weight = 0.899, all p's < 0.001), and had high internal consistency (Cronbach's α = 0.955).

Effect of Expert Evidence Quality and Likeability on Persuasiveness

Average persuasiveness ratings by condition are shown in **Figure 2**. The mean persuasiveness ratings by condition were: high-quality, neutral likeability M = 83.5 (SD = 11.9); high-quality, high-likeability M = 80.6 (SD = 17.2); high-quality, low-likability M = 69.1 (SD= 22.4); low-quality, neutral likeability M = 62.8 (SD = 23.7); low-quality, high-likeability M = 66.9 (SD= 22.7); low-quality, low-likeability M = 53.2 (SD = 29).

A two-way ANOVA showed a significant main effect of evidence quality [$F_{(1, 232)}$ = 35.58, $p < 0.001$, $\eta p^2 = 0.133$] whereby higher quality evidence resulted in higher persuasiveness ratings on average compared to lower quality evidence. There was also a significant main effect of likeability [$F_{(2,232)}$ = 8.23, $p < 0.001$, $\eta p^2 = 0.066$]. Follow-up main effects (Tukey HSD) analysis showed that across evidence quality conditions, there was no significant difference in persuasiveness between the high and neutral likeability conditions [M_{diff} high and neutral = 0.28, 95% CI (−7.72, 8.29), $p = 0.996$], however, both the high and neutral conditions resulted in significantly higher persuasiveness ratings than the low-likeability condition
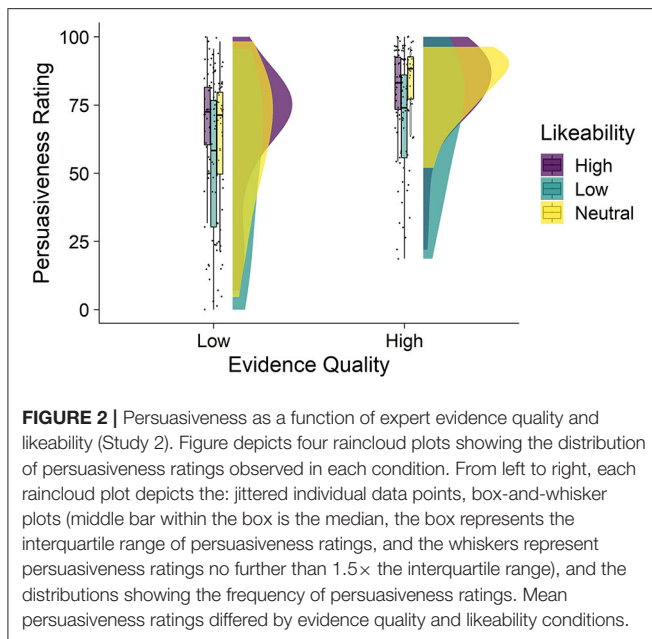
FIGURE 2 | Persuasiveness as a function of expert evidence quality and likeability (Study 2). Figure depicts four raincloud plots showing the distribution of persuasiveness ratings observed in each condition. From left to right, each raincloud plot depicts the: jittered individual data points, box-and-whisker plots (middle bar within the box is the median, the box represents the interquartile range of persuasiveness ratings, and the whiskers represent persuasiveness ratings no further than 1.5× the interquartile range), and the distributions showing the frequency of persuasiveness ratings. Mean persuasiveness ratings differed by evidence quality and likeability conditions.

[M_{diff} neutral and low = 12.28, 95% CI (3.93, 20.62), $p = 0.002$; M_{diff} high and low = 12.56, 95% CI (4.5, 20.62), $p = 0.001$].There was no significant interaction between evidence quality and likeability [$F_{(2,232)} = 0.557$, $p = 0.574$, $\eta p^2 = 0.005$] indicating that the effect of likeability was the same across both evidence quality conditions.

Multiple regressions were conducted to examine whether ratings of the eight ExPEx attributes and WCS-likeability predicted persuasiveness ratings. The overall model was significant [$F_{(9,228)} = 104.74$, $p < 0.001$] and accounted for 79.8% of the variance in persuasiveness ratings (adjusted $R^2 = 0.798$). Ratings of trustworthiness, specialty, opinion, and likeability were significant independent predictors of persuasiveness (all p's ≤ 0.003); foundation, field, ability, support, and consistency were not (all p's ≥ 0.059; see **Table 8**).

Relationship Between Persuasiveness and Sentencing Decision

The binominal logistic regression testing the relationship between persuasiveness and sentencing decision was a good fit and was significant [$\chi^2_{(1)} = 20.72$, $p < 0.001$]. Persuasiveness accounted for 12.2% of the variance in sentencing decision [Nagelkerke $R^2 = 0.122$; Wald $\chi^2_{(1)} = 15.44$, $p < 0.001$], with a one unit increase in persuasiveness increasing the odds of a death sentence by 1.036 (Exp B).

Effect of Expert Evidence Quality and Likeability on Sentencing Decision

The proportion of death sentences by condition is shown in **Table 9**. The binominal logistic regression predicting sentencing decision from expert quality and likeability conditions and their interaction produced a good fit for the data but was not significant [$\chi^2_{(5)} = 1.52$, $p = 0.911$], accounting for just 0.9%

of the variance in sentencing decision (Nagelkerke $R^2 = 0.009$). Neither quality, likeability, nor their interactions were significant independent predictors of sentencing decision (all $ps \geq 0.407$; see **Table 10**).

The binominal logistic regression testing whether continuous ratings of the eight ExPEx attributes (i.e., the ExPEx attribute items) and WCS likeability scores predicted sentencing decision was a good fit and was significant [$\chi^2_{(9)} = 36.09$, $p < 0.001$], accounting for 20.5% of the variance in sentencing decisions (Nagelkerke $R^2 = 0.205$). Ratings of the opinion attribute was the only independent predictor of sentencing decision ($p = 0.045$). The remaining predictors were not significant (all $ps \geq 0.202$ see **Table 11**).

Study 2 Discussion

Study 2 further examined the effect of expert likeability and quality on jurors' perception of expert persuasiveness and sentencing decisions. As in Study 1, we found that participants' perceptions of the persuasiveness of expert evidence were significantly affected by evidence quality and expert likeability. There were also no interactions between evidence quality and likeability. We also found that subjective perceptions of the eight ExPEx attributes and likeability again accounted for approximately 80% of the variance in persuasiveness scores.

While higher quality experts were more persuasive than lower quality experts, Study 2 suggests that adding negative likeability cues reduces perceived likeability and persuasiveness, while adding positive likeability cues did not increase either likeability or persuasiveness. These results replicate the Study 1 finding that the persuasiveness of an expert opinion is determined by both its underlying scientific quality, and superficial aspects of the experts' interpersonal style, but go further to suggest that it may be an unfriendly, arrogant, and conceited style that is particularly influential on persuasiveness.

However, as in Study 1, likeability did not affect sentencing decisions while aspects of evidence quality did. Participants' perceptions of the clarity and conservativeness of the experts' opinion uniquely predicted sentencing outcomes. Likeability condition and ratings did not directly impact sentencing decisions. Thus, concerns about impact of likeability on jurors' sentencing outcomes may be misplaced.

GENERAL DISCUSSION

Two studies examined the effect of expert quality and likeability on potential jurors' perceptions of the persuasiveness expert evidence and sentencing decisions in a Capitol case. Across both our studies we found that higher quality experts were regarded as more persuasive than lower quality experts. We also found that less likeable experts were considered less persuasive than more likeable experts, irrespective of evidence quality. Moreover, models predicting persuasiveness from continuous ratings of expert quality attributes and likeability were significant and accounted for ~80% of the variance in persuasiveness ratings. This result is particularly impressive considering participants were evaluating detailed trial transcripts and videos. Even so, likeability did not significantly affect sentencing

TABLE 8 | Multiple regression predicting persuasiveness from continuous expert persuasiveness expectancy (ExPEx) ratings and witness credibility score (WCS) for likeability.

Persuasiveness	B	CI$_{95\%}$ for B		SE B	β	p	R²	Adj. R²
		Lower	Upper					
Model						<0.001	0.805	0.798
Foundation	0.056	−0.031	0.144	0.044	0.053	0.206		
Field	−0.064	−0.154	0.027	0.046	−0.054	0.166		
Specialty	0.202***	0.113	0.291	0.045	0.203***	<0.001		
Ability	0.09	−0.004	0.184	0.048	0.097	0.059		
Opinion	0.131**	0.048	0.214	0.042	0.14**	0.002		
Support	−0.016	−0.092	0.061	0.039	−0.017	0.686		
Consistent	0.066	−0.027	0.158	0.047	0.065	0.162		
Trustworthy	0.383***	0.299	0.467	0.042	0.462***	<0.001		
WCS-Likeability	0.232**	0.08	0.385	0.077	0.119**	0.003		

B, unstandardised regression coefficient; CI, confidence interval; SE B, standard error of the coefficient; β, standardised coefficient. *p < 0.05, **p < 0.01, ***p ≤ 0.001.

outcomes, whereas various elements of expert quality did (i.e., trustworthiness and foundation in Study 1; opinion in Study 2). Models predicting sentencing decisions from continuous ratings of quality and likeability accounted for a smaller but significant 20–30% of the variance.

Expert Persuasiveness

This research is the first to show that jurors' perceptions of persuasiveness are influenced by expert likeability even in scenarios where very rich information is available about expert evidence quality. This suggests that previously observed likeability effects were not merely an artefact of simplistic or sparse decision-making scenarios. Rather, likeability appears to be genuinely influential in determining how persuasive expert evidence will be.

We also found evidence that the effects of being a dislikeable expert are more impactful than the effects of being likeable. Specifically, we found that a video of an arrogant, conceited, disagreeable expert reduced both likeability and persuasiveness compared to a neutral transcript. But a video of a smiling, modest, open expert did not increase either likeability or persuasiveness compared to a neutral transcript. Thus, irrespective of evidence quality–we saw clear evidence of a dislikeability cost, but we were not able to produce an equivalent likeability benefit. Indeed, in our scenario the cost of being dislikeable was substantial, and in descriptive terms resulted in a low-likeability but *high-quality* expert being treated similarly to a high-likeability but *low-quality* expert.

Our finding that it may be dislikeability rather than likeability that affects persuasiveness is somewhat inconsistent with past research suggesting that likability *boosts* credibility and persuasiveness (Brodsky et al., 2009; Neal et al., 2012). However, this may be because previous studies did not include a neutral likeability control condition as a baseline to gage the effect of likeability manipulations. When this control condition was added, the data clearly suggested the effect of likeability cues was asymmetric and driven by negative rather than positive expert likeability attributes. In fact, our data suggest that likeability ratings may have been at ceiling even in the neutral likeability condition whereby participants seemed to assume the expert

TABLE 9 | Proportion of participants selecting death sentence by evidence quality and likeability condition.

Likeability	Low-Quality Evidence %	High-Quality Evidence %
Neutral	21.6	30.8
Low	27	21.6
High	29.5	27.3

was likeable, until proven otherwise. This suggests it may be impractical or at least very difficult for experts to make themselves more likeable than jurors expect but can easily fall short of existing high expectations.

Across both studies we also found strong and consistent evidence that higher quality evidence is more persuasive than lower quality evidence. This result fits with previous research using rich representations of expert opinion quality (Martire et al., 2020) but is somewhat inconsistent with concerns about juror insensitivity to evidence quality (Cooper et al., 1996; Diamond and Rose, 2005; Hans et al., 2007, 2011; McAuliff and Kovera, 2008; McAuliff et al., 2009; Koehler et al., 2016; Eldridge, 2019). In our studies, jurors were provided with information about an expert's field, their specialist background, the validity of their practicing domain, their trustworthiness, consistency with other experts, their supporting evidence and opinion clarity. This level of information is arguably necessary for an informed evaluation of expert quality and exceeds the level of information provided in previous studies that have typically found that jurors struggle to determine evidence quality (Martire et al., 2020). Our results suggest that jurors can appropriately evaluate evidence quality when they have access to more of the relevant information that they need for the task. This interpretation is in line with the ELM perspective of information processing (Petty and Cacioppo, 1984, 1986) whereby decision-makers are more likely to systematically process information if they have sufficient knowledge and capacity (Petty and Cacioppo, 1984, 1986). However, it remains to be seen whether jurors can also use detailed information about expert quality to differentiate between more marginal or subtle differences in evidence quality

TABLE 10 | Logistic regression predicting sentencing decision from evidence quality condition, likeability condition, and their interaction.

	B	SE	Wald	df	p	Odds Ratio	95% CI for Odds Ratio	
							Lower	Upper
Expert evidence quality	0.11	0.47	0.06	1	0.813	1.12	0.44	2.83
Likeability (1)	0.17	0.49	0.12	1	0.726	1.19	0.46	3.07
Likeability (2)	−0.31	0.52	0.34	1	0.558	0.74	0.26	2.05
Expert evidence quality (1) * likeability (1)	−0.59	0.71	0.69	1	0.41	0.56	0.14	2.23
Expert evidence quality (1) * likeability (2)	0.18	0.72	0.06	1	0.8	1.2	0.29	4.94

Life in prison was coded as zero and death was coded as one. B, unstandardised regression coefficient; SE B, standard error of the coefficient; CI, confidence interval. *p < 0.05, **p < 0.01, ***p ≤ 0.001.

TABLE 11 | Logistic regression predicting sentencing decision from continuous expert persuasion expectancy (ExPEx) ratings and witness credibility score (WCS) for likeability.

	B	SE	Wald	df	p	Odds Ratio	95% CI for Odds Ratio	
							Lower	Upper
Foundation	0.008	0.011	0.52	1	0.473	1.01	0.99	1.03
Field	−0.015	0.012	1.49	1	0.223	0.99	0.96	1.01
Specialty	0.004	0.012	0.08	1	0.775	1.00	0.98	1.03
Ability	−0.002	0.012	0.02	1	0.889	0.998	0.98	1.02
Opinion	0.025*	0.013	4.02	1	0.045	1.03	1.00	1.05
Support	−0.012	0.01	1.46	1	0.227	0.99	0.97	1.01
Consistent	0.012	0.012	0.93	1	0.334	1.01	0.99	1.04
Trustworthy	0.013	0.012	1.28	1	0.258	1.01	0.99	1.04
WCS-Likeability	0.026	0.021	1.63	1	0.202	1.03	0.99	1.07

Life in prison was coded as zero and death was coded as one. B, unstandardised regression coefficient; SE B, standard error of the coefficient; CI, confidence interval. *p < 0.05, **p < 0.01, ***p ≤ 0.001.

than those we used in our manipulations. Future research is needed to examine this possibility.

Even so, it is important to note that sensitivity to evidence quality did not remove the effects of dislikeability. When jurors had the information and knowledge to effectively evaluate expert evidence quality, they still used information about expert likeability to determine how much credibility, value, and weight to give the expert evidence. Given that likeability is not related to expert quality or merit, the fact there is a persuasion cost of dislikeability remains problematic, particularly when we see that high-quality evidence is viewed similarly to low-quality evidence from a more likeable expert. Thus, our data show that likeability has the potential to undermine the effects of evidence quality in an undesirable way.

More broadly, across both studies, we found that subjective ratings of evidence quality and likeability impacted persuasiveness. In both studies, we found that subjective ratings of the eight ExPEx attributes and likeability accounted for approximately 80% of the variance in persuasiveness. This suggests that jurors' perceptions of these markers collectively provide a good account of persuasiveness judgments. Impressions of the expert's trustworthiness, their specialist background, their opinion, the consistency of their judgment with other experts, their supporting evidence, and ability were all unique predictors of persuasiveness. This also indicates that

jurors use relevant indicators of evidence quality to determine how persuasive an expert will be.

Sentencing Decisions

Although allocation to expert evidence quality and likeability condition significantly influenced ratings of persuasiveness, this did not translate into a direct impact on sentencing decision. The finding that expert likeability condition did not predict sentencing decision is consistent with the literature examining expert likeability (Brodsky et al., 2009; Neal et al., 2012; Parrott et al., 2015). Therefore, while likeability is considered in judgments of expert persuasiveness, and may make jurors more inclined to agree with the expert, it does not appear to materially affect the final sentencing outcome. Sentencing decisions are consequential and require jurors to consider a wider range of trial considerations relating to the defendant, the sentencing options, and the expert (Greene et al., 2007). Therefore, jurors may pay less attention to likeability in this context, instead focusing on more relevant information (i.e., the expert's opinion, trustworthiness, and foundation).

Although the absence of an effect of quality condition on sentencing can also be attributed to the same broad complexity of sentencing decisions, that explanation seems unsatisfactory in this case. The quality of the expert opinion *should* be a key determinant of the sentencing decision in this trial scenario, even

when considering the broader trial context. Specifically, a high-quality expert opinion that is consistent with the application of the death penalty, should result in more death sentences than a low-quality version of the same opinion. The fact that this did not happen even though jurors were more persuaded by high- than low-quality opinions suggests that jurors may not know how to apply low- and high-quality evidence in their sentencing decisions. This might explain why much of the literature suggests that expert evidence is universally persuasive—jurors may be influenced by the expert–but they also may struggle to incorporate evidence quality into their final judgments (Cutler et al., 1989; Ivković and Hans, 2006; Daftary-Kapur et al., 2010; Bornstein and Greene, 2011).

The idea that jurors' may not know how to incorporate evidence quality into their sentencing decisions is further supported by regression models considering continuous ratings of expert quality and likeability. In both studies, we found that subjective ratings of the eight ExPEx attributes and likeability accounted for between 20 and 30% of the variance in sentencing decision. This was substantially less than the variance accounted for by quality and likeability in persuasiveness ratings (∼80%). Therefore, perceptions of quality and likeability were less influential for sentencing decisions. This suggests other factors became important or increased in prominence for sentencing that were less relevant to the evaluation of persuasiveness. Indeed, since these new factors appear to be de-emphasising valid quality indicators, it is important to understand what these other factors might be, why they are being used and whether they are logically relevant to sentencing determinations or not. This would form a fruitful line of research to pursue in future studies.

Despite this, we did find that the continuous subjective ratings of foundational validity (Study 1), trustworthiness (Study 1), and opinion (Study 2) were unique predictors of sentencing decision. Likeability ratings were not. This suggests that jurors are incorporating some relevant markers of expert evidence quality into their sentencing decisions. However, these indicators were not consistent across studies, and many valid indicators of quality were not significant independent predictors. Therefore, there is substantial scope for quality information to take a larger role in sentencing decisions and future research should look at methods to improve utilisation of quality information in jurors' sentencing decisions.

Implications

Altogether, these findings suggest that likeability impacts perceptions of persuasiveness but not by increasing persuasiveness, rather by decreasing it. Experts already appear to be assumed to be likeable at baseline, so attempts to be more likeable may not be effective. Instead, experts should consider whether their highly confident, authoritative, or self-assured interpersonal style could come across as arrogant, disagreeable, or conceited, because being seen in these ways may result in high-quality evidence being discounted to the point where its impact is akin to lower quality evidence provided by a more likeable expert.

More significantly, higher quality evidence was more persuasive than lower quality evidence, irrespective of how likeable an expert was. Sentencing decisions were also affected by perceptions of opinion clarity, discipline foundational validity and trustworthiness. Thus, these results suggest that experts can make their evidence more compelling, and influential by increasing the objective quality of their evidence and communicating that quality to decision-makers.

Limitations and Future Directions

One limitation of this research is that both studies involved the same Capitol murder case and sentencing decisions. Sentencing is not the only kind of legal decision made by jurors and so it remains to be investigated whether likeability influences jurors' decision-making in other cases and for other types of decisions (e.g., verdicts, liability, damages). Such information is vital to determine the generalisability of our findings. Further, our participants were not assessed for death-qualification (Witherspoon v. Illinois, 391 U.S. 510, 1968). Participants in our study may therefore be more or less willing to impose a death-sentence than real jurors deciding the same case. For this reason, it would be valuable for future research to include questions to establish a death-qualification. However, we note that this is unlikely to affect our results because we were interested in between-group differences in persuasiveness, rather than verdict frequency *per se*.

Another limitation relates to the ecological validity of our study. The mock trial in our first study was a transcript, participation took ∼20–40 min including post-trial decision-making, and there was no deliberation phase. This does not reflect real criminal trials which typically are conducted in-person, and can last for weeks or months. Our participants were also predominantly recruited via MTurk and may therefore differ from real jurors in terms of demographic characteristics and investment in the task. To improve the ecological validity, we used videoed trial materials rather than a transcript in Study 2 because it has been suggested that trial videos improve the ecological validity of experimental juror studies (Studebaker et al., 2002). We further ensured higher data quality standards by implementing multiple attention and manipulation check measures, constraining the time allocated to complete the study and narrowing participating criteria to higher-quality respondents. Indeed, per Lieberman et al. (2016), the methodology of the current studies surpasses the acceptable criteria for juror decision-making paradigms. Nonetheless, future studies should consider using longer in-person trials involving more types of evidence and involving jury deliberation.

Finally, it is worth mentioning that participants in our study were asked to complete the evidence quality and likeability measures prior to rating persuasiveness and making sentencing decisions. It is therefore possible that jurors were primed with quality and likeability information that they might not have otherwise considered in their assessments of persuasiveness and sentencing options. We included these measures before the persuasiveness judgment to measure the maximum impact our manipulations might have on perceptions of persuasiveness and sentencing decisions. We reasoned that if there is no effect of quality or likeability under these conditions, then we could not reasonably expect a larger effect of either likeability or quality in real-world settings. That is, we wanted to have the best

possible chance of detecting any influence of evidence quality or likeability considerations. The fact that we did not see effects of likeability on sentencing under these conditions strongly suggests that likeability does not significantly affect sentencing decisions. Conversely, the significance of quality attributes suggests that quality may affect real world sentencing decisions. Indeed, the magnitude of the quality effects in our study were consistent with those obtained in other studies where quality was not primed prior to measuring persuasiveness (Martire et al., 2020). Even so, future research could randomise question order to remove any possible priming effects.

CONCLUSION

Our results suggest that expert evidence quality and likeability both impact perception of expert persuasiveness. Specifically, dislikeability reduces persuasiveness irrespective of evidence quality. However, only subjective impressions of the foundational validity, trustworthiness, and clarity of the expert opinion significantly predicted Capitol sentencing decisions. Thus, concerns about juror reliance on the peripheral likeability cue may be most relevant to evaluations of the expert evidence in isolation, rather than to trial outcomes. Our results also strongly suggest that likeability does little to boost persuasion while being disliked has a clear cost. Thus, experts can take comfort from the fact that weak evidence is not bolstered by an affable interpersonal style, but they may rightly be concerned that this superficial attribute has the potential to weaken the persuasive power of otherwise high-quality evidence. Thus, care should be taken to ensure that confidence does not become conceit if experts want their evidence to be given its merited value.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://osf.io/yfgke/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of New South Wales Human Research Ethics Approval Panel C – Behavioural Sciences (#3308). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MY and KM contributed to the study concept, experimental design, and reporting. MY led the data collection and analysis. Both authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Bornstein, B. H., and Greene, E. (2011). Jury decision making: Implications for and from psychology. *Curr. Direct. Psychol. Sci.* 20, 63–67. doi: 10.1177/0963721410397282

Brodsky, S. L., Griffin, M. P., and Cramer, R. J. (2010). The witness credibility scale: an outcome measure for expert witness research. *Behav. Sci. Law* 28, 892–907. doi: 10.1002/bsl.917

Brodsky, S. L., Neal, T. M. S., Cramer, R. J., and Ziemke, M. H. (2009). Credibility in the courtroom: how likeable should an expert witness be? *J. Am. Acad. Psychiatry Law* 37, 525–532.

Buhrmester, M., Kwang, T., and Gosling, S. D. (2011). Amazon's mechanical turk : a new source of inexpensive, yet high-quality, data? *Perspect. Psychol. Sci.* 6, 3–5. doi: 10.1177/1745691610393980

Cacioppo, J. T., and Petty, R. E. (1982). The need for cognition. *J. Pers. Soc. Psychol.* 42, 116–131. doi: 10.1037/0022-3514.42.1.116

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *J. Pers. Soc. Psychol.* 39, 752–766. doi: 10.1037/0022-3514.39.5.752

Cooper, J., Bennett, E. A., and Sukel, H. L. (1996). Complex scientific testimony: How do jurors make decisions? *Law Hum. Behav.* 20, 379–394. doi: 10.1007/BF01498976

Cutler, B. L., Dexter, H. R., and Penrod, J. D. (1989). Expert testimony and jury decision making: an empirical analysis. *Behav. Sci. Law* 7, 215–225. doi: 10.1002/bsl.2370070206

Daftary-Kapur, T., Dumas, R., and Penrod, S. D. (2010). Jury decision-making biases and methods to counter them. *Legal Criminol. Psychol.* 15, 133–154. doi: 10.1348/135532509X465624

Diamond, S. S. (2007). How jurors deal with expert testimony and how judges can help. *J. Law Policy* 16:47. Available online at: https://brooklynworks.brooklaw.edu/jlp/vol16/iss1/4

Diamond, S. S., and Rose, M. R. (2005). Real juries. *Ann. Rev. Law Soc. Sci.* 1, 255–284. doi: 10.1146/annurev.lawsocsci.1.041604.120002

Drummond, C., and Fischhoff, B. (2017). Development and validation of the scientific reasoning scale. *J. Behav. Decis. Mak.* 30, 26–38. doi: 10.1002/bdm.1906

Edmond, G., Found, B., Martire, K., Ballantyne, K., Hamer, D., Searston, R., et al. (2016). Model forensic science. *Austra. J. For. Sci.* 48, 496–537. doi: 10.1080/00450618.2015.1128969

Eldridge, H. (2019). Juror comprehension of forensic expert testimony: A literature review and gap analysis. *Forensic Science International: Synergy* 1, 24–34. doi: 10.1016/j.fsisyn.2019.03.001

Gladstone, G., and Parker, G. B. (2002). When you're smiling does the whole world smile for you? *Austra. Psychiatry* 10, 144–146. doi: 10.1046/j.1440-1665.2002.00423.x

Greene, E., and Gordan, N. (2016). Can the "hot tub" enhance jurors understanding and use of expert testimony? *Wyoming Law Rev.* 16, 359–385. Available online at: https://scholarship.law.uwyo.edu/wlr/vol16/iss2/6

Greene, E., Heilbrun, K., Fortune, W. H., and Nietzel, M. T. (2007). *Wrightsman's Psychology and the Legal System, 6th Edn.* Belmont, CA: Thomson/Wadsworth.

Gross, S. R. (1991). Expert evidence. *Wis. L. Rev.* 1991, 1113–1232.

Guy, L. S., and Edens, J. F. (2003). Juror decision-making in a mock sexually violent predator trial: gender differences in the impact of divergent types of expert testimony. *Behav. Sci. Law* 21, 215–237. doi: 10.1002/bsl.529

Hans, V. P., Kaye, D. H., Farley, E., Albertson, S., and Dann, B. M. (2007). Science in the jury box: jurors' views and understanding of mitochondrial DNA evidence. *Cornell Law Fac. Public.* 82, 1–46. doi: 10.2139/ssrn.1025582

Hans, V. P., Kaye, D. H., Farley, E., Albertson, S., and Dann, B. M. (2011). Science in the jury box: Jurors' views and understanding of mitochondrial DNA evidence. *Law Hum. Behav.* 35, 60–71. doi: 10.1007/s10979-010-9222-8

Heuer, L., and Penrod, S. (1994). Trial complexity: a field investigation of its meaning and its effects. *Law Hum. Behav.* 18, 29–51. doi: 10.1007/BF01499142

Innocence Project (2021). *Pervis Payne Will not be Executed; DA Concedes he is a Person with Intellectual Disability.* Available online at: https://www.innocenceproject.org/

Ivković, S. K., and Hans, V. P. (2006). Jurors' evaluations of expert testimony: judging the messenger and the message. *Law Soc. Inquiry* 28, 441–482. doi: 10.1111/j.1747-4469.2003.tb00198.x

Jurs, A. W. (2016). Expert prevalence, persuasion, and price: what trial participants really think about experts. *Indiana Law J.* 91, 353–392. Available online at: https://www.repository.law.indiana.edu/ilj/vol91/iss2/4

Kerns, M. H., and Sun, C.-R. (1994). Narcissim and reactions of interpersonal feedback. *J. Res. Pers.* 28, 4–13. doi: 10.1006/jrpe.1994.1002

Kleinke, C. L. (1986). Gaze and eye contact: a research review. *Psychol. Bull.* 100, 78–100. doi: 10.1037/0033-2909.100.1.78

Koehler, J. J. (2012). Proficiency tests to estimate error rates in the forensic sciences. *Law Probabil. Risk* 12, 89–98. doi: 10.1093/lpr/mgs013

Koehler, J. J., Schweitzer, N. J., Saks, M. J., and McQuiston, D. E. (2016). Science, technology, or the expert witness: what influences jurors' judgments about forensic science testimony. *Psychol. Public Policy Law* 22, 401–413. doi: 10.1037/law0000103

Krauss, D. A., and Sales, B. D. (2001). The effects of clinical and scientific expert testimony on juror decision making in capital sentencing. *Psychol. Public Policy Law* 7, 267–310. doi: 10.1037/1076-8971.7.2.267

Leathers, D. G. (1997). *Successful Nonverbal Communication, 3rd Edn.* Boston, MA: Allyn and Bacon.

Levin, H., Giles, H., and Garrett, P. (1994). The effects of lexical formality and accent on trait attributions. *Lang. Commun.* 14, 265–274. doi: 10.1016/0271-5309(94)90004-3

Lieberman, J., Krauss, D., Heen, M., and Sakiyama, M. (2016). The good, the bad, and the ugly: professional perceptions of jury decision-making research practices. *Behav. Sci. Law* 34, 495–514. doi: 10.1002/bsl.2246

Maeder, E. M., McManus, L. A., McLaughlin, K. J., Yamamoto, S., and Stewart, H. (2016). Jurors' perceptions of scientific testimony: the role of gender and testimony complexity in trials involving DNA evidence. *Cogent Psychol.* 3, 1–14. doi: 10.1080/23311908.2016.1264657

Martire, K. A., Edmond, G., and Navarro, D. (2020). Exploring juror evaluations of expert opinions using the expert persuasion expectancy framework. *Legal Criminol. Psychol.* 25, 90–110. doi: 10.1111/lcrp.12165

McAdams, D. P., and Powers, J. (1981). Themes of intimacy in behavior and thought. *J. Pers. Soc. Psychol.* 40, 573–587. doi: 10.1037/0022-3514.40.3.573

McAuliff, B. D., and Kovera, M. B. (2008). Juror need for cognition and sensitivity to methodological flaws in expert evidence. *J. Appl. Soc. Psychol.* 38, 385–408. doi: 10.1111/j.1559-1816.2007.00310.x

McAuliff, B. D., Kovera, M. B., and Nunez, G. (2009). Can jurors recognize missing control groups, confounds, and experimenter bias in psychological science? *Law Hum. Behav.* 33, 247–257. doi: 10.1007/s10979-008-9133-0

McAuliff, B. D., Nemeth, R. J., Bornstein, B. H. and Penrod, S. D. (2003). "Juror decision making in the 21st century: Confronting science and technology in court," in *Handbook of Psychology in Legal Contexts, 2nd Edn*, eds D. Carson and R. Bull (Chichester: Wiley), 303–327.

McGaffey, R. (1979). The expert witness and source credibility - the communication perspective. *Am. J. Trial Advoc.* 1, 58–73.

Neal, T. M. S. (2014). Women as expert witnesses: a review of the literature. *Behav. Sci. Law* 32, 164–179. doi: 10.1002/bsl.2113

Neal, T. M. S., and Brodsky, S. L. (2008). Expert witness credibility as a function of eye contact behavior and gender. *Crim. Justice Behav.* 35, 1515–1526. doi: 10.1177/0093854808325405

Neal, T. M. S., Guadagno, R. E., Eno, C. A., and Brodsky, S. L. (2012). Warmth and competence on the witness stand: implications for the credibility of male and female expert witnesses. *J. Am. Acad. Psychiatry Law* 40, 488–497.

Parrott, C. T., Neal, T. M. S., Wilson, J. K., Brodsky, S. L., Parrott, C. T., and Neal, T. M. S. (2015). Differences in expert witness knowledge: do mock jurors notice and does it matter? *J. Am. Acad. Psychiatry Law* 43, 69–81.

Petty, R. E., and Cacioppo, J. T. (1984). The effects of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. *J. Pers. Socl. Psychol.* 46, 69–81. doi: 10.1037/0022-3514.46.1.69

Petty, R. E., and Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Adv. Exp. Soc. Psychol.* 19, 123–205. doi: 10.1016/S0065-2601(08)60214-2

Petty, R. E., Cacioppo, J. T., and Goldman, R. (1981). Personal involvement as a determinant of argument-based persuasion. *J. Pers. Soc. Psychol.* 41, 847–855. doi: 10.1037/0022-3514.41.5.847

Salerno, J. M., Bottoms, B. L., and Peter-Hagene, L. C. (2017). Individual versus group decision making: Jurors' reliance on central and peripheral information to evaluate expert testimony. *PLoS ONE* 12:e0183580. doi: 10.1371/journal.pone.0183580

San José-Cabezudo, R., Gutiérrez-Arranz, A. M., and Gutiérrez-Cillán, J. (2009). The combined influence of central and peripheral routes in the online persuasion process. *CyberPsychol. Behav.* 12, 299–308. doi: 10.1089/cpb.2008.0188

Schuller, R. A., Terry, D., and Mckimmie, B. (2005). The impact of expert testimony on jurors' decisions: Gender of the expert and testimony complexity. *J. Appl. Soc. Psychol.* 35, 1266–1280. doi: 10.1111/j.1559-1816.2005.tb02170.x

Schutz, J. S. (1997). Expert witness and jury comprehension: an expert's perspective. *J. Law Public Policy* 7, 107–119.

Shuman, D. W., Whitaker, E., and Champagne, A. (1994). An empirical examination of the use of expert witnesses in the courts – part II : a three city study. *Jurimetrics* 34, 193–208.

Sporer, S. L., Penrod, S., Read, D., and Cutler, B. (1995). Choosing, confidence, and accuracy: a meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychol. Bull.* 118, 315–327. doi: 10.1037/0033-2909.118.3.315

Studebaker, C. A., Robbennolt, J. K., Penrod, S. D., Pathak-Sharma, M. K., Groscup, J. L., and Devenport, J. L. (2002). Studying pretrial publicity effects: new methods for improving ecological validity and testing external validity. *Law Hum. Behav.* 26, 19–41. doi: 10.1023/A:1013877108082

Swenson, R. A., Nash, D. L., and Roos, D. C. (1984). Source credibility and perceived expertness of testimony in a simulated child-custody case. *Prof. Psychol. Res. Pract.* 15, 891–898. doi: 10.1037/0735-7028.15.6.891

Tenney, E. R., Spellman, B. A., and MacCoun, R. J. (2008). The benefits of knowing what you know (and what you don't): how calibration affects credibility. *J. Exp. Soc. Psychol.* 44, 1368–1375. doi: 10.1016/j.jesp.2008.04.006

Texas Criminal Procedure Code, Article 370.071b-f. (1985).

Von Ahn, L., Maurer, B., McMillen, C., Abraham, D., and Blum, M. (2008). reCAPTCHA: human-based character recognition via web security measures. *Science* 321, 1465–1468. doi: 10.1126/science.1160379

Witherspoon v. Illinois, 391 U.S. 510. (1968).

"Sounding Black": Speech Stereotypicality Activates Racial Stereotypes and Expectations About Appearance

Courtney A. Kurinec*† and Charles A. Weaver III

Department of Psychology and Neuroscience, Baylor University, Waco, TX, United States

Black Americans who are perceived as more racially phenotypical—that is, who possess more physical traits that are closely associated with their race—are more often associated with racial stereotypes. These stereotypes, including assumptions about criminality, can influence how Black Americans are treated by the legal system. However, it is unclear whether other forms of racial stereotypicality, such as a person's way of speaking, also activate stereotypes about Black Americans. We investigated the links between speech stereotypicality and racial stereotypes (Experiment 1) and racial phenotype bias (Experiment 2). In Experiment 1, participants listened to audio recordings of Black speakers and rated how stereotypical they found the speaker, the likely race and nationality of the speaker, and indicated which adjectives the average person would likely associate with this speaker. In Experiment 2, participants listened to recordings of weakly or strongly stereotypical Black American speakers and indicated which of two faces (either weakly or strongly phenotypical) was more likely to be the speaker's. We found that speakers whose voices were rated as more highly stereotypical for Black Americans were more likely to be associated with stereotypes about Black Americans (Experiment 1) and with more stereotypically Black faces (Experiment 2). These findings indicate that speech stereotypicality activates racial stereotypes as well as expectations about the stereotypicality of an individual's appearance. As a result, the activation of stereotypes based on speech may lead to bias in suspect descriptions or eyewitness identifications.

Keywords: stereotypes, social categorization, race, Black Americans, phenotype, dialect, speech perception

INTRODUCTION

Every day, we interact with those we do not know in order to perform our jobs, run our errands, or engage in other, more leisurely activities. Making use of the available social information, we quickly form impressions about these unfamiliar people and use those impressions to guide our interactions. For instance, we may use the available cues to make assumptions about another

person's membership in certain social categories, such as their likely gender, ethnic group, occupation, or social status, and the stereotypes or beliefs we associate with those social categories influence the traits we expect this person to possess.

How strongly we link stereotypes for a social category to a specific individual often depends on the extent to which that individual is seen as a typical member or exemplar of that group. Individuals who possess more of the features related to their social group are often more closely associated with stereotypes about that group (Blair et al., 2004b; Walker and Wänke, 2017). Unfortunately for Black Americans, these stereotypes include expectations about criminality (Eberhardt et al., 2004) and may influence how more stereotypical Black Americans are perceived and treated by the legal system. More stereotypically Black individuals are more likely to be associated with crime or a criminal label both by members of the general public and police officers (Eberhardt et al., 2004; Kleider et al., 2012) and are more likely to be misidentified as a suspect by eyewitnesses (Knuycky et al., 2014; Kleider-Offutt et al., 2017). The influence of stereotypicality extends to sentencing, such that when the victim is White, Black defendants with more stereotypical features are more likely to be given the death penalty than less stereotypical looking Black defendants (Eberhardt et al., 2006).

Studies investigating stereotypicality often focus on physical characteristics; for example, those with fuller lips or a wider nose are more likely to be viewed as phenotypically Black (Blair et al., 2002, 2004b; Hagiwara et al., 2012). These Afrocentric facial features can be used separately or in combination with skin tone to influence judgments about race typicality (Stepanova and Strube, 2009, 2012b; Dunham et al., 2015). However, judgments about individuals are influenced by more than their static visual appearance; indeed, some of our interactions do not include visual information, e.g., telephone calls or online voice chat. It is unclear how other aspects of an individual—specifically how they speak—contribute to judgments of whether an individual is "stereotypically Black."

Language as a Marker of Social Category

Language is an important tool in social categorization (Rakiæ et al., 2011b; Dragojevic et al., 2018). How a person speaks can act as an index or sign of one's background (Bucholtz and Hall, 2005), and this indexical information about a person's social category information is quickly and automatically accessed. For instance, information about a person's likely gender category can be accessed around 150 ms after voice presentation (Munson and Solum, 2010). In some cases, the way a person talks can be an even more important indicator of the social category to which an individual belongs than facial features (Rakiæ et al., 2011a).

Social category information obtained from a person's speech can activate stereotypes or other assumptions not only about the speaker's linguistic background, but also the social groups to which they likely belong (Giles and Rakiæ, 2014), providing a gateway for individuals to make judgments about the speaker (Giles, 1970; Mulac and Rudd, 1977). Importantly, the stereotypes and attitudes activated in relation to a given speaker depend

heavily on the listener. The stereotypes a listener associates with a specific social group are dictated by that listener's social environment and cultural context, to include both implicitly and explicitly held beliefs, as well as the listener's ability to notice and classify certain linguistic features, the listener's expectations about the conversation, and their own communication goals (Cargile and Bradac, 2001; Preston, 2018).

One cultural factor that shapes how speakers are perceived is the level of standardization of the dialect they employ. Dialects, or the way of speaking associated with a certain regional, cultural, or ethnic group, are often described as either "standard" or "non-standard," with standard variants of a language being ones that are supported by the state and/or other influential institutions (Milroy, 2001; Milroy and Milroy, 2012). Accordingly, the identities and dialects of those in power influence which dialects are considered standard (Lippi-Green, 1997). Standard dialects, such as General American English in the United States (U.S.) or Received Pronunciation in the United Kingdom (UK), are often viewed more favorably and seen as more prestigious than non-standard dialects (Dent, 2004; Morales et al., 2012), even by speakers of non-standard dialects (Anisfeld et al., 1962; Carter and Callesano, 2018). How those who use non-standard dialects are viewed often depends on how the people most closely associated with that dialect are perceived. For instance, people from the Southern U.S. are stereotyped as being uneducated, poor, and lazy (Slade and Narro, 2012). Unsurprisingly, speakers using a Southern U.S. dialect, a non-standard dialect of American English associated with this region, are seen as less wealthy, less intelligent, less healthy, and less attractive than speakers using a more standard American dialect (Dent, 2004; Phillips, 2010; Shamina, 2016). Further, linguistic features associated with Southern U.S. English are implicitly associated with blue collar jobs and lower intelligence (Campbell-Kibler, 2012; Loudermilk, 2015). In this way, judgments about speakers of a given dialect can reflect the stereotypes about members of that social group.

Sounding Black

Given the interconnectivity between language, social categorization, and stereotypes, it is likely that individuals who "sound Black" are more likely to be identified as Black Americans and therefore more likely to be associated with stereotypes about the group. One way individuals may be thought to "sound Black" is through their use of African American Vernacular English (AAVE). AAVE is a non-standard dialect of American English closely associated with and spoken predominantly (but not only) by Black Americans (Cutler, 2003; Rickford, 1999). Often denigrated as slang or improper English, AAVE is in fact a valid language system, with regular phonological and grammatical features such as -ing dropping (e.g., "goin'" vs. "going"), r-lessness (e.g., "fo'" vs. "four"), negative concord (e.g., "He ain't seen nothin'"), and the use of habitual be (e.g., "She be workin'" indicates "She's often working") (Pullum, 1999; Thomas, 2007; see Jones, 2015 for more on regional variations in AAVE). Like speakers of other non-standard dialects, speakers of AAVE are seen less favorably than speakers of the more standard General American English in most contexts (Payne et al., 2000; Koch et al., 2001; Dent,

2004; Rodriguez et al., 2004; Billings, 2005). Speakers of AAVE are seen as less competent, less sociable, less professional, less educated, and of poorer character than speakers of more standard American English (Payne et al., 2000; Koch et al., 2001; Dent, 2004; Billings, 2005). As with the Southern U.S. dialect, many of the traits associated with AAVE are also associated with its dominant speakers: Black Americans (Devine and Elliot, 1995; Maddox and Gray, 2002). For instance, individuals show a greater implicit association between weapons and AAVE speakers than more standard speakers (Rosen, 2017), suggesting that stereotypes about criminality and violence, often associated with Black Americans, are also linked to AAVE speakers. Further, AAVE's close association with Black Americans has also led to linguistic profiling, or discrimination against those who speak a certain way due to their assumed membership in a social group. Discriminating against someone for their way of speaking can allow for anti-Black bias to circumvent legal protections, leading to worse outcomes and fewer opportunities in areas such as housing for those who use AAVE and are assumed to be Black (Purnell et al., 1999; Massey and Lundy, 2001).

However, not all Black Americans speak AAVE, and those who do speak AAVE do not use it all the time or in all contexts (Rickford, 1999; McCluney et al., 2021). Yet listeners can reliably identify the race of Black speakers regardless of dialect with over 85% accuracy in longer (10 second) clips (Kushins, 2014) and approximately 60–70% accuracy after listening to a one-second clip or a single word (Walton and Orlikoff, 1994; Purnell et al., 1999). This ability to quickly identify the racial identity of a speaker is likely due to the presence of certain phonological features, e.g., final consonant dropping or vowel quality (Walton and Orlikoff, 1994; Thomas and Reaser, 2004; Perrachione et al., 2010), but there is no consensus on which specific linguistic cues trigger the perception of a speaker as Black. Despite this lack of scientific consensus, it is likely that listeners have learned, through their social or cultural environment, to associate certain linguistic features with Black Americans and use those cues to identify speaker race (Perrachione et al., 2010). As a result, Black speakers who do not employ the expected linguistic features can be miscategorized as members of other races or ethnic groups (Thomas and Reaser, 2004; Perrachione et al., 2010).

Regardless of whether listeners are picking up on AAVE or other linguistic features associated with Black Americans, the strength with which speakers employ these features likely predicts whether listeners will categorize speakers as Black and the stereotypes assigned to them. Rodriguez et al. (2004) found that speakers who had a stronger AAVE dialect (i.e., used more AAVE features) were rated less favorably than those with a more moderate AAVE dialect. Thus, one would expect that speakers who have stronger dialects and sound "more Black" to listeners will be not only more likely to be identified as Black, but also will be more associated with stereotypes about Black Americans—including expectations about criminality and violence. However, this has yet to be directly investigated.

If sounding "more Black" does lead to an increase in Black stereotypes, it could also lead to the assumption that the speaker has a more stereotypically Black appearance as well. Previous work has found that stereotypes can influence expectations about appearance. Hughes and Miller (2016) found that, in line with the "what sounds beautiful is good" stereotype (Zuckerman and Driver, 1989), individuals with more attractive voices are expected to have more attractive faces. Separately, Osborne and Davies (2013) found that participants who watched a video of a crime stereotypically associated with Black people remembered the perpetrator as appearing more phenotypically Black than those who watched the perpetrator commit a crime stereotypically associated with White people, even when the crimes were matched on severity and violence. Being perceived as sounding more stereotypically Black could similarly activate listeners' stereotypes about Blackness and influence the expectations a listener has for their appearance—a supposition with critical implications for the legal system, e.g., in ensuring reliable suspect identifications.

Given that linguistic profiling and discrimination based on how a person speaks are not explicitly prohibited under U.S. law (Wiehl, 2002; MacNeal et al., 2019), understanding how speech stereotypicality influences assumptions about a speaker is needed before any countermeasures to minimize linguistic bias can be developed. To address this issue, we conducted two experiments to investigate the relationship between sounding more stereotypically Black and the assignment of stereotypical traits associated with Black Americans (Experiment 1) and decisions about likely appearance (Experiment 2). In both experiments, participants listened to audio recordings from Black speakers before making their judgments. We expected that speakers whose speech is perceived as more stereotypically Black would be associated with more stereotypical traits about Black Americans and with a more phenotypical Black appearance.

EXPERIMENT 1

Those who are perceived as looking more phenotypically Black are also more likely to be associated with stereotypes about Black Americans (Blair et al., 2004b). In Experiment 1 we explored whether this pattern extended to those who are perceived as *sounding* more Black. Participants listened to and evaluated audio recordings taken from the internet of American and British Black male speakers. We had two hypotheses for this experiment. First, we expected that participants would assign more stereotypical traits to speakers they rated as sounding more stereotypically Black. Second, since listeners rely on learned linguistic cues to identify Black speakers, we anticipated that our U.S.-based listeners' concept of "stereotypically Black" would be informed by their cultural context. Thus, they would assign more stereotypes to more stereotypical-sounding Black speakers who were also perceived as American.

Materials and Methods
Participants
We recruited 75 participants from Amazon's Mechanical Turk (MTurk) using the TurkPrime interface (Litman et al., 2017). Only U.S.-based workers who had completed at least 100 Human Intelligence Tasks (HIT) and who had HIT approval rates of 98% or greater were allowed to participate in this study. Workers were

paid $0.50 for approximately 10 min of work. The sample was predominantly female (62.7% women; 34.7% men; 1.3% gender neutral; 1.3% prefer not to answer) and White (69.3% White; 5.3% Black; 8.0% Asian; 5.3% Hispanic or Latino/a; and 12.0% multiracial), and the mean age of the sample was 39.77 years (SD = 12.72; Range = 20–71). Participants were overall well-educated; 45.3% of the sample reported they had a bachelor's degree, and 26.7% reported they had at least some college credit. Only 2.7% reported they did not possess at least a high school diploma or its equivalent.

Participants used a variety of terms when freely describing their own dialects, with the most common labels being some derivative of "American English" (13.3%), "Midwestern" (13.3%), or "Standard American" (10.7%). Other expected regional (e.g., "American East Coast," "Bostonian," "Texan," or "Southern English") and racial/ethnic terms (e.g., "African American," "Chinese English," or "Italian American") also appeared. Interestingly, some participants labeled their dialects as "White" or "Caucasian" (6.7%).

In order to minimize low-effort or bot responses, we removed four participants' data for providing nonsensical or off-topic responses (e.g., "NICE") to our free response dialect question. Other responses that were related to a way of speaking but did not describe a dialect *per se* (e.g., "slang" or "soft spoken") were retained in analyses, leaving us with data from 71 participants (675 observations). According to G*Power (Faul et al., 2009), *a priori* power analysis for a two-tailed multiple regression with our seven predictors would require a sample size of at least 55 to detect an interaction of medium effect size (f^2 = 0.15), with α = 0.05 and 1-β = 0.80. Using these same parameters for α and 1-β, our sample of 71 would allow us to detect at least an effect size of f^2 = 0.11.

Materials

Audio Recordings

We created 20 audio recordings featuring 14 Black North American male and 6 Black British male speakers. To have more ecologically valid recordings, the audio was taken from YouTube videos by searching terms such as "Black British," "Black English," and "Black American." The audio was then shortened to a sample of the speaker's speech. These clips ranged between 18 and 35 s due to variations in the length of speakers' utterances. In the clips, speakers discussed a variety of topics, i.e., travel experiences (n = 3), restaurants/food/diet (n = 7), domestic/foreign culture (n = 5), working abroad (n = 2), creating a better life for one's family (n = 1), comedians (n = 1), and sport (n = 1)[1]. The audio clips were screened for any explicit mentions of the speaker's race or national origin. After data collection, one of the British speakers (who discussed sports) was revealed to be an American actor; as a result, this speaker was removed from analyses, leaving us with data on 19 speakers. Although we could reasonably

assume the rest of our British speakers were from the UK due to the content of either the full video or their profiles, we could not find information about the nationality of two of the American speakers. Due to their use of American English dialects, we assume these speakers are from the U.S.; however, we recognize that some of these speakers could be Canadian given the overlap between Canadian and American English (see Labov et al., 2008).

Stereotypical Traits

To rate the speakers from the voice clips, participants were shown a list of 30 adjectives taken from Devine and Elliot (1995) and Maddox and Gray (2002). The adjectives included those associated with Black American stereotypes (athletic, criminal, lazy, poor, rhythmic, uneducated, unintelligent, hostile, loud, dirty, inferior, ostentatious, sexually aggressive, and aggressive), as well as counter-stereotypic (intelligent, kind, educated, motivated, and wealthy), and neutral adjectives (attractive, bad attitude, self-assured, unattractive, superstitious, naïve, unreliable, talkative, materialistic, arrogant, and ambitious). To create our dependent variable, we calculated the proportion of stereotypical adjectives out of all the adjectives a participant assigned to a given voice.

Speaker Perceived Demographics

Participants rated the speakers' perceived race, nationality, age, voice attractiveness, speech stereotypicality, and dialect. Participants indicated the speaker's perceived race from a list of five races/ethnicities: White/Caucasian, Black/African American, Hispanic or Latino/a, American Indian/Alaska Native, or Native Hawaiian/Other Pacific Islander; participants could also indicate "Other" and enter their own label. For perceived nationality, participants were shown five major English-speaking countries: the United Kingdom, the United States, Canada, Australia, or New Zealand. Once again, participants could also write-in an alternate answer. The speaker's perceived age (in years) was indicated by entering a number. Perceived voice attractiveness and stereotypicality were both rated on 7-point Likert-type scales (1 = Not at all attractive/stereotypical; 4 = Neutral; 7 = Extremely attractive/stereotypical). Speech stereotypicality was rated in terms of the perceived race of the speaker; for example, if the participant believed the speaker's race was White/Caucasian, the speech stereotypicality question asked them how stereotypically White/Caucasian the speaker's voice was. Finally, participants were shown four options for perceived dialect, as well as the option to provide another response: African American English ("Ebonics"), Standard American English ("Midwestern"), Black British English, or Standard British English ("Received Pronunciation" or "Queen's English"). The labels of African American English, Standard American English, Black British English, and Standard British English are equivalent to AAVE, General American English, Multicultural London English, and Received Pronunciation, respectively. These labels were used in lieu of the more appropriate naming conventions in order to make the dialects more easily understood by participants. Further, our use of the Black British English label was used so our U.S.-based participants, who are likely unfamiliar with Multicultural London English, would have an equivalent British

---

[1] After collecting data it was noted that in one audio clip, a speaker mentioned eating chicken (Speaker 14), which, given the stereotype about Black Americans and chicken (Demby, 2013), could have potentially primed participants to view this speaker as more Black (see Gaither et al., 2015). However, removing this speaker from analyses did not meaningfully change our pattern of results. As a result, this speaker was left in our final dataset.

racial dialect to AAVE. Participants were also asked to indicate how familiar they were with the speaker's dialect on a 5-point Likert-type scale (1 = Not at all; 3 = A moderate amount; 5 = A great deal).

Procedure

After indicating their informed consent to the study procedures, all participants were assigned a subject number to safeguard their identity. At the beginning of the study, participants supplied demographic information and completed a short audio test to ensure that they were able to hear the recordings. Participants then listened to 10 of the 20 audio clips; the selection of audio clips and order of presentation was randomly determined. After listening to a clip, participants selected the adjectives that they believed the average person would use to describe the speaker. The instructions emphasized that the participant did not have to personally agree with the description. Next, participants chose the speaker's likely race, nationality, and age and rated how attractive the speaker's voice was. Finally, participants indicated how stereotypical the speaker's voice was for their perceived race (as chosen by the participant) before choosing a likely dialect for the speaker and rating how familiar they were with that dialect. The audio clip remained on the screen while participants provided their ratings so participants could refer back to it. After rating all of their assigned voice clips, participants were asked to provide a term or label to describe their own dialect or way of speaking. Upon completing all study procedures, participants were thanked for their work and debriefed on the purpose of the study.

Results

Data Analysis

To explore how perceived speech stereotypicality influences the traits people assign to a speaker, we ran a mixed effects model predicting the proportion of stereotypical traits assigned. Due to the aforementioned overlap between U.S. and Canadian speakers, we coded responses of the U.S. and Canada to the speaker's perceived country question under the umbrella of North American (N. Am.). Therefore, the model included fixed effects for perceived race (Non-Black, Black), perceived country of origin (Non-N. Am., N. Am.), speech stereotypicality (mean-centered by participant), and their interactions, and random intercepts for participants and the individual speakers.

We also ran an additional model with perceived speaker age and voice attractiveness (mean-centered by participant) included as covariates. Speaker age was controlled for as age may moderate how stereotype content may vary not only by race, but also by the age of the individual (e.g., Andreoletti et al., 2015), which may affect both how stereotypical listeners rated speakers' voices as well as what traits they associated with the speaker. Stereotypes about Blackness in particular may be more salient for younger rather than older Black men, as an analysis of Pennsylvania sentencing data from the late 1980s to early 1990s revealed that, controlling for crime severity and other court-related factors, young Black men received harsher sentences than older Black men and White men and White and Black women of any age

(Steffensmeier et al., 1998). Voice attractiveness was included as a covariate as more attractive voices are often associated with more positive traits (Zuckerman and Driver, 1989), which may influence the adjectives listeners associate with speakers. How stereotypical a speaker sounds as a member of their perceived race and how attractive their voice is rated are likely related, given that conceptions of attractiveness broadly favor Eurocentric traits (Maddox, 2004). However, previous work investigating facial features and attractiveness found racial typicality and attractiveness had small to moderate correlations (Stepanova and Strube, 2018).

In both models, predictors were sum coded, with Non-Black and Non-N. Am. serving as the reference groups. Significance tests were run by conducting likelihood ratio tests comparing the full model to the model without the predictor of interest for all predictors. Follow-up pairwise comparisons were Bonferroni-adjusted for four tests. The data were imported into R Studio (RStudio Team, 2019) using the haven package (Wickham and Miller, 2019). Data were analyzed using the psych (Revelle, 2017), gmodels (Warnes et al., 2018), afex (Singmann et al., 2019), lme4 (Bates et al., 2015), and emmeans (Lenth, 2019) packages.

Speaker Perceived Demographics

Perceived Country

We compared speakers' actual country of origin (N. Am. or UK) to participants' perceived country choices (N. Am. or UK). Participants were generally able to correctly identify the country each speaker originated from, with greater accuracy for N. Am speakers than UK speakers (UK speakers 89.4% correct; N. Am. speakers 94.4% correct).

Speaker Perceived Race

Although all speakers were Black, participants identified speakers as Black only around two-thirds of the time (61.3%). Out of all possible racial options, participants identified our UK speakers nearly equally often as White (49.2%) or Black (44.7%), whereas our N. Am. speakers were identified primarily as Black (67.3%), with White as the second most frequent option (26.4%). A logistic mixed effects model on perceived race (Non-Black, Black) with fixed effects for speaker's actual country (UK, N. Am.) and their perceived country (Non-N. Am., N. Am.) and random effects for participants and speakers found that neither actual nor perceived country of origin predicted perceptions of the speaker's race ($ps \geq 0.240$).

Speaker Perceived Dialect

Although our speakers used different dialects at different strengths, participants categorized UK and N. Am. speakers by region-appropriate labels. UK speakers were mostly labeled as using Standard British English (63.1%) or Black British English (29.6%), and N. Am. speakers were more likely labeled as using Standard American English (56.9%) or AAVE (35.7%).

Speaker Dialect Familiarity

Using the 5-point scale, participants indicated they were fairly familiar with General American English ($M = 3.78$, $SD = 1.04$, $n = 289$) and AAVE ($M = 3.61$, $SD = 0.91$, $n = 179$). As

expected, participants were less familiar with Standard British English ($M = 2.74$, $SD = 0.98$, $n = 125$) and Black British English ($M = 2.36$, $SD = 1.06$, $n = 66$).

Proportion of Stereotypical Traits

On average, participants assigned each speaker 4.21 traits ($SEM = 0.09$). All participants assigned speakers at least one trait (Range = 1–17). Around a third of all assigned adjectives were stereotypical ($M = 0.31$, $SEM = 0.01$). The percent of each stereotypical trait assigned to speakers based on perceived country and perceived race are presented in **Table 1**.

We observed a significant two-way interaction of perceived race and speech stereotypicality on proportion of stereotypical traits assigned, $B = -0.03$, $SE = 0.01$, $p = 0.005$. A test of simple slopes indicated that the slope of stereotypicality on assigned stereotypical traits was greater for perceived Black speakers than non-Black speakers, $t = 2.83$, $p = 0.005$. For perceived Black speakers, those rated 1 SD above the mean on speech stereotypicality were assigned more traits than those rated 1 SD below the mean, $t = 2.94$, $p = 0.014$. The proportion of assigned traits did not differ by rating for non-Black speakers, $t = -1.14$, $p > 0.999$, nor did it differ by race for those rated 1 SD above or below the mean ($ps > 0.185$).

The two-way interaction was qualified by a significant three-way interaction of perceived race, perceived country, and speech stereotypicality, $B = 0.02$, $SE = 0.01$, $p = 0.027$. To follow-up the three-way interaction, we conducted tests of simple slopes for the perceived race by speech stereotypicality interaction separately for perceived N. Am. and Non-N. Am. speakers. For perceived N. Am. speakers, the slope of stereotypicality on assigned traits was greater for perceived Black speakers than perceived non-Black speakers, $t = 4.84$, $p < 0.001$. Pairwise comparisons indicated

that among perceived Black speakers, those rated 1 SD above the mean on speech stereotypicality were assigned more stereotypical traits than those rated 1 SD below the mean, $t = 4.65$, $p < 0.001$. For perceived non-Black speakers, this pattern was reversed; speakers rated 1 SD above the mean on speech stereotypicality were assigned fewer stereotypical traits than those rated 1 SD below the mean, $t = -2.61$, $p = 0.037$. Additionally, at 1 SD above the mean on stereotypicality ratings, perceived Black speakers were assigned more stereotypical traits than their non-Black counterparts, $t = 3.53$, $p = 0.002$, but at 1 SD below the mean perceived Black speakers were assigned fewer traits than their counterparts, $t = -2.88$, $p = 0.017$. In other words, as perceived Black American speakers were rated as more stereotypical-sounding, they were also assigned more stereotypical traits, but as perceived non-Black American speakers were rated more stereotypical-sounding, they were assigned fewer stereotypical traits. There was no such difference in slopes of stereotypicality for those perceived as Non-N. Am. speakers, $t = -0.38$, $p = 0.704$. The mean proportion of stereotypical traits assigned to speakers by perceived country, race, and stereotypicality are presented in **Table 2**.

There were no significant main effects of perceived race, $B = -0.009$, $SE = 0.02$, $p = 0.547$, perceived country, $B = -0.08$, $SE = 0.02$, $p = 0.663$, nor speech stereotypicality, $B = 0.01$, $SE = 0.01$, $p = 0.202$. No other interactions were significant ($ps > 0.691$).

Perceived age was significantly correlated to stereotypicality (uncentered), although the relationship was small, $r = 0.13$, $p = 0.001$. Voice attractiveness (uncentered), on the other hand, had a small negative correlation with stereotypicality, $r = -0.15$, $p < 0.001$. When adding perceived age and voice attractiveness as covariates to our model, the pattern of results remained the same. The two- and three-way interactions remained significant, $B = -0.02$, $SE = 0.01$, $p = 0.045$; and $B = 0.02$, $SE = 0.01$, $p = 0.007$, respectively. Only voice attractiveness ratings predicted proportion of stereotypical traits, $B = -0.07$, $SE = 0.01$, $p < 0.001$. As perceptions of voice attractiveness increased, the proportion of stereotypical traits decreased. Age did not have a significant effect, $B = 0.0005$, $SE = 0.002$, $p = 0.793$.

Discussion

The findings from Experiment 1 provide support to our two hypotheses. First, speakers who were perceived as sounding more

TABLE 1 | Percent of stereotypic adjectives assigned to speakers in Experiment 1, by perceived country and perceived race.

Adjective	Non-North American		North American	
	Black	Non-Black	Black	Non-Black
Athletic	**16.5**	3.9	10.6	**11.4**
Criminal	1.2	0.0	5.8	0.0
Dirty	1.2	1.0	5.2	1.9
Inferior	1.2	4.9	6.4	5.1
Lazy	3.5	3.9	9.4	6.3
Ostentatious	7.1	**12.6**	4.0	2.5
Poor	2.4	4.9	**19.5**	2.5
Rhythmic	**15.3**	9.7	15.5	6.3
Sexually aggressive	2.4	4.9	3.3	3.2
Aggressive	10.6	**12.6**	12.2	**11.4**
Uneducated	9.4	11.7	**24.9**	10.1
Unintelligent	5.9	10.7	**20.1**	**14.6**
Hostile	7.1	6.8	4.9	5.1
Loud	**22.4**	**25.2**	15.8	**20.3**

Percentages are out of all ratings for each category. Bold indicates the top three adjectives assigned to speakers by category; four values are in bold for North American Non-Black speakers due to a tie.

TABLE 2 | Mean proportion of stereotypical traits and standard errors for speakers in Experiment 1, by perceived country, perceived race, and stereotypicality rating.

Perceived country	Stereotypicality rating	Black	Non-Black
Non-North American	−1 SD	0.26 (0.05)	0.26 (0.05)
	+1 SD	0.33 (0.08)	0.29 (0.06)
North American	−1 SD	0.22 (0.04)	0.35 (0.04)
	+1 SD	0.39 (0.04)	0.23 (0.04)

Standard errors of the mean (SEM) are in parentheses. Mean proportion of stereotypes is calculated for stereotypicality ratings ±1 standard deviation (SD) from the mean. Higher stereotypicality ratings indicate that speakers are more stereotypical sounding as a member of their perceived race.

stereotypically Black were assigned more stereotypes associated with Black Americans compared to those who were perceived as less stereotypically Black sounding. Second, the relationship between sounding stereotypically Black and the assignment of Black stereotypes was true only for those seen as Black Americans. For those perceived as North Americans, more stereotypically sounding Black speakers were assigned a larger proportion of stereotypical traits than less stereotypical-sounding Black speakers. Interestingly, this pattern was the opposite for those perceived as non-Black speakers: more stereotypically sounding non-Black speakers were assigned fewer traits than less stereotypical-sounding non-Black speakers. We did not observe any differences in how speech stereotypicality affected the assignment of traits for those not perceived as North Americans. Notably, the observed pattern results remained when perceived age and voice attractiveness were added to the model. Although voice attractiveness did significantly predict the proportion of stereotypic traits assigned, it was not strongly correlated with perceived stereotypicality, echoing previous findings investigating race-related facial features (Stepanova and Strube, 2018).

It is unclear from this study why non-Black American speakers would be assigned more Black stereotypes as they sounded less stereotypically non-Black. This may represent a sort of black sheep effect (Marques et al., 1988), where our mostly White listeners are more biased against perceived non-Black (predominantly identified as White) speakers who do not conform to their expected speech. However, future research is needed to determine whether this pattern replicates when potential confounds, such as the option to identify speakers as belonging to races other than White or Black, are minimized. Regardless, our findings regarding Black American speakers are consistent with previous work on racial phenotypes. Like those seen as phenotypically Black, those who are rated as more stereotypically Black (American) sounding are associated with more stereotypes about Black Americans. Additionally, as these stereotypical traits are specific to Black Americans, listeners use vocal cues about nationality to distinguish between those from the U.S. and Canada and those from other nations.

EXPERIMENT 2

Because we found that speech stereotypicality elicited stereotypes in a similar manner as racial phenotypes, we next investigated in Experiment 2 whether speakers who sounded more stereotypical were also expected to appear more phenotypical. After piloting audio from male Black American speakers and images of Black male faces to find high and low stereotypicality exemplars, we had participants listen to a speaker and choose which of two faces was most likely to be the speaker. We expected that participants would be more likely to choose a face that matched the stereotypicality level of the speaker's voice; in other words, that speakers who sounded less stereotypically Black would be associated with less phenotypically Black faces and those who sounded more stereotypically Black would be associated with more phenotypically Black faces. Further, we expected that

participants' social attitudes, such as their feelings about Black Americans or their willingness to respond in a desirable manner, would be significant covariates for participants' choice of face.

Materials and Methods
Participants

Participants ($N = 155$) were recruited from Amazon's MTurk and paid \$1.00 for successful completion of the 15-min study. Only U.S.-based workers who had completed at least 100 HIT and who had HIT approval rates of 90% or greater were allowed to participate in this study, and those who participated in the pilot were not allowed to participate in this study. Five participants were removed from analyses due to failure to follow instructions, resulting in a final sample size of 150 ($M_{Age} = 34.40$, $SD = 11.17$; Range = 20–70). Participants were relatively equally divided by gender (56.7% men; 43.3% women). Participants were predominantly White (64.0%), with Black (16.7%), Hispanic and Latino/a (8.0%), Native American (5.3%), and Asian (4.7%) individuals also represented. Most participants reported having a Bachelor's degree (46.0%), followed by those with some college (19.3%), a high school degree or equivalent (14.0%), and Master's degree (12.7%). No participants reported having an education level lower than that of a high school degree.

To calculate an *a priori* power analysis for a two-tailed logistic regression, we used the accuracy index of 72% found by Purnell et al. (1999) as the probability that participants would choose a face that matched the stereotypicality level of the speaker's voice. Thus, the probability of choosing a face that did not match the stereotypicality level of the speaker's voice (p_1) was 0.28. According to G*Power (Faul et al., 2009), a sample size of at least 40 would be needed to detect an odds ratio of 6.61, with α = 0.05 and 1-β = 0.80. Using these same parameters for α, 1-β, and p_1, a sample size of 150 would allow us to detect an odds ratio of at least 2.59.

Materials
Stimuli

Twenty-four Black male faces and 12 Black male voices were selected based on data from a pilot study rating Black male faces and voices (see **Supplementary Material**). The faces were from the Chicago Face Database (Ma et al., 2015), and the voices were from the (International Dialects of English Archive [IDEA], 2011). In the audio recordings from the International Dialects of English Archive (2011), the speakers all read from one of two standard passages: one about a veterinary nurse and one about rainbows. The faces were organized into pairs of strongly and weakly phenotypical faces with similar age and attractiveness based on the normed ratings from the Chicago Face Database and confirmed by the results from our pilot. We matched the 12 face pairs with the 12 voices based on perceived age to ensure that the pairs would be believable. The face pairs were randomly assigned to be viewed while participants listened to only a predetermined portion (22–40 s) of a strongly or weakly stereotypical voice, such that participants saw six of the face pairs with a strongly stereotypical voice and the remaining six with a weakly stereotypical voice. The order of the voices and the side

on which the high phenotypicality face appeared (left or right) were counterbalanced to account for any order effects.

Manipulation Checks

To ensure participants were paying attention, we asked them to indicate the race of the faces they viewed from a list of six races/ethnicities: White/Caucasian, Black/African American, Hispanic, Latino/a, or Spanish origin, Native American, Asian American, or Native Hawaiian, Samoan, or other Pacific Islander. Participants could also indicate "Other" and enter their own label. Participants were also asked to select the task they completed from a list of four possible tasks: "Rated the attractiveness of faces," "Identified a speaker's face from an audio clip," "Listened to audio clips and chose what race the speaker is," or "Wrote about stereotyping I have experienced."

Face Choice

Participants indicated which face was likely the speaker's using radio buttons under the faces. We recoded these values for analyses (0 = Low phenotypicality, 1 = High phenotypicality).

Confidence

Participants indicated their confidence in their choice using a slider bar (0% = No confidence, 100% = Complete confidence).

Social Attitude Scales

Participants completed a series of measures designed to assess their beliefs about the criminal justice system, racial bias against Black Americans, their own racial identity, and likelihood of engaging in desirable responding.

Pretrial Juror Attitude Questionnaire. The Pretrial Juror Attitude Questionnaire (PJAQ; Lecci and Myers, 2008) is a 29-item scale evaluating pretrial juror attitudes and beliefs about the criminal justice system, with six subscales evaluating conviction proneness, system confidence, cynicism toward the defense, social justice, racial bias, and innate criminality. Given the stereotypes associating Black Americans and criminality, only the racial bias subscale (four items) was included in analyses. Due to experimenter error, each item was measured on a 6-point Likert-type scale (1 = strongly disagree, 6 = strongly agree) rather than the intended 5-point scale. Higher scores denote more bias against minorities related to the criminal justice system. Although the total measure showed good reliability, Cronbach's α = 0.93, the subscale showed poorer reliability (see **Table 3**).

TABLE 3 | Means, standard errors, and reliability for the social attitude scales used in Experiment 2.

Measure	M	SEM	Cronbach's α
Pretrial juror attitude questionnaire—racial bias subscale	12.7	0.31	0.513
Symbolic racism scale	17.19	0.40	0.816
Collective self-esteem scale—race specific version	67.31	1.42	0.808
Balanced inventory of desirable responding (short form)			
Self-deceptive enhancement	33.12	0.85	0.820
Impression management	32.70	0.83	0.800

Symbolic Racism Scale. The Symbolic Racism Scale (SRS; Henry and Sears, 2002) is an 8-item measure used to assess more symbolic or subtle racism against African Americans. Items are measured on a 4-point scale with the exception of item 3 ("Some say that black leaders have been trying to push too fast. Others feel that they haven't pushed fast enough. What do you think?"), which is measured on a 3-point scale. Total scores can range from 8 to 31, with higher scores indicating higher bias against African Americans.

Symbolic Racism Scale Collective Self-Esteem Scale—Race Specific Version. The Collective Self-Esteem Scale—Race Specific Version (CSE-R; Luhtanen and Crocker, 1992) is a 16-item scale assessing a person's own racial or ethnic identity, with four subscales assessing their racial/ethnic membership, their personal view of their racial/ethnic memberships, how others view their racial/ethnic memberships, and how important group membership is to their identity. Each item is measured on a 7-point scale (1 = strongly disagree, 7 = strongly agree). Total scores on the CSE-R can range from 16 to 112, with higher scores showing more collective self-esteem.

Balanced Inventory of Desirable Responding. The Balanced Inventory of Desirable Responding short-form (BIDR-16; Hart et al., 2015) is a 16-item scale that assesses socially desirable responding, with two subscales: self-deceptive enhancement and impression management, which reflect unintentional and intentional socially desirable responding, respectively. Traditionally, each item is measured dichotomously; however, Stöber et al. (2002) recommend using continuous scoring to improve reliability and validity. Thus, we assessed each item using a 7-point scale (1 = not true, 7 = very true). For both subscales, total scores can range from 8 to 56, with higher scores denoting greater self-deceptive enhancement and impression management, respectively.

Procedure

After indicating their informed consent, participants filled out demographic information and completed an audio check. Next, participants were told a cover story that the research team was interested in understanding how well people can identify a speaker's face from an audio recording. Participants were told that they could see either White or Black faces in the study, although only Black faces were used. Participants listened to either a strongly or weakly stereotypical voice before choosing which of a pair of faces taken from the pilot was more likely to be the speaker and indicating their confidence in their decision. Participants made 12 two-alternative forced choice decisions and 12 confidence ratings in total. Finally, all participants answered the manipulation checks and completed the social attitude scales before being debriefed and thanked for their work.

Results

Data Analysis

To explore how perceived speech stereotypicality influences face selections, we first ran a mixed effects logistic regression on participants' chosen faces (Low or High Phenotypicality). The initial model included voices (Low or High Stereotypicality) as a

fixed effect and participants and the individual face pairs entered as random intercepts. We also ran a mixed effects regression on choice confidence with the same fixed and random effects to see if speech stereotypicality had any undue influence on participants' confidence in their face selections.

We followed up both of these analyses by adding our social attitude scales to control for social attitudes about race as well as desirable responding. The racial bias subscale of the PJAQ and the SRS were included, as well as their interaction terms with voices, to investigate whether racism against minorities related to crime or racism against Black Americans affected which face was selected or moderated the effect of voice stereotypicality on face choices. Work by Stepanova and Strube (2012a) found that implicit bias against Blacks moderated the effect of skin color on ratings of racial typicality, indicating that an individual's racial bias may affect how individuals are categorized. The CSE-R and two BIDR subscales were added as covariates as previous work has found that higher collective self-esteem can affect how individuals categorize faces (Elliott et al., 2017), and higher socially desirable responding, whether intentional or unintentional, may have influenced participants' face choices, e.g., in an effort to appear less biased. In all models, the voices variable was sum coded, with Low Stereotypicality serving as the reference group. Each social attitude scale was mean-centered before being added to the models. Significance tests were run by conducting likelihood ratio tests comparing the full model to the model without the predictor of interest for all predictors. Follow-up pairwise comparisons were Bonferroni-adjusted for two tests. The data were imported into R Studio using the haven package (Wickham and Miller, 2019). Data were analyzed using the afex (Singmann et al., 2019), lme4 (Bates et al., 2015), and emmeans (Lenth, 2019) packages.
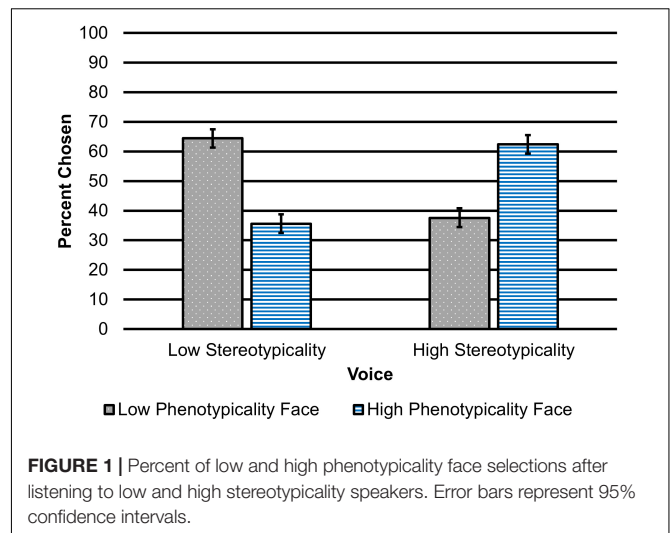
Manipulation Checks

The majority of participants correctly answered our manipulation check questions. For the race of the faces, 88.67% of participants noted the faces were Black/African American. For our question on the task, 85.33% of participants correctly noted they were asked to identify a speaker's face from an audio recording. Despite 34 participants failing at least one of our manipulation check questions, our pattern of results did not differ when these participants were included in the dataset (see **Supplementary Material**). Thus, our reported results include the full sample of 150 ($N = 1,800$ choices).

Face Selection

We began by examining the results from our model on face choice. As expected, speech stereotypicality influenced participants' decisions about which face was more likely to belong to the speaker, $B = -0.60$, $SE = 0.15$, $p < 0.001$. Participants who heard the low stereotypicality voice were less likely to choose the high phenotypicality face. After hearing the low stereotypicality voice, participants chose the low phenotypicality face 64.44% of the time and chose the high phenotypicality face 35.56% of the time (**Figure 1**).

We then looked at the model on choice confidence. Participants' confidence ratings spanned the range of possible scores ($M = 66.68$, $SEM = 0.47$, Range = 0–100); however, their

FIGURE 1 | Percent of low and high phenotypicality face selections after listening to low and high stereotypicality speakers. Error bars represent 95% confidence intervals.

level of confidence in their decision did not significantly differ by which voice they heard, $B = -0.47$, $SE = 0.82$, $p = 0.576$[2].

Finally, we looked at the models with the scores from the racial bias subscale from the PJAQ race bias subscale, the SRS, CSE-R, and the two BIDR subscales added (means and SEMs are reported in **Table 3**). For choice, speech stereotypicality remained a significant predictor, $B = -0.61$, $SE = 0.16$, $p < 0.001$. The interactions between the racial bias subscale and voice and the SRS and voice were also significant, $B = 0.11$, $SE = 0.02$, $p < 0.001$, and $B = -0.06$, $SE = 0.01$, $p < 0.001$, respectively. The slope for racial bias on face choice was greater for low stereotypicality voices than high stereotypicality voices, $z = 5.71$, $p < 0.001$. At 1 SD above the mean on the racial bias subscale, participants were not significantly more likely to choose the high phenotypicality face after hearing the high stereotypicality voice, $z = 1.22$, $p = 0.442$. However, at 1 SD below the mean, participants were over seven times more likely to choose the high phenotypicality face after hearing the high rather than low stereotypicality voice, $z = 5.85$, $p < 0.001$, $OR = 7.58$.

Separately, the slope for racism against Black Americans on face choice was greater for high stereotypicality voices than low stereotypicality voices, $z = 4.05$, $p < 0.001$. At 1 SD above the mean on the SRS, participants were around six times more likely to choose the high phenotypicality face after hearing the high vs. low stereotypicality voice, $z = 5.19$, $p < 0.001$, $OR = 6.01$. At 1 SD below the mean, participants were not significantly more likely to choose the high phenotypicality face after hearing the high stereotypicality voice, $z = 1.90$, $p = 0.114$. No other predictors were significant, $ps \geq 0.197$.

For confidence, speech stereotypicality remained non-significant, $B = -0.47$, $SE = 0.81$, $p = 0.575$. However, the racial bias subscale of the PJAQ, $B = 1.87$, $SE = 0.41$, $p < 0.001$, and the BIDR self-deceptive enhancement subscale were significant predictors of confidence, $B = 0.32$, $SE = 0.14$, $p = 0.026$. As

---

[2]Adding the counterbalance condition to the base models for choice and confidence did not alter our pattern of results, and counterbalance condition was not a significant predictor in either model, $ps \geq 0.235$. Therefore, we did not include counterbalance condition in our covariate models.

participants showed greater racial bias and as their own tendency to unintentionally engage in socially desirable responding increased, so too did people's confidence in their choices. The other predictors were not significant, $ps \geq 0.053$.

Discussion

Speech stereotypicality influenced people's judgments about physical appearance in Experiment 2. Participants were more likely to choose faces that matched the level of stereotypicality (low or high) of the voice they heard. Critically, even when controlling for participants' social attitudes, the link between speech and face stereotypicality remained. However, participants' explicit and subtle racism moderated the effect of speech stereotypicality on their face choices. Specifically, as participants expressed more racial bias on the PJAQ, they were less likely to choose the face that matched the level of stereotypicality with the voice they heard. Separately, as participants endorsed more subtle racist beliefs against Black Americans, they were more likely to choose the face that had the same level of stereotypicality as the voice. In other words, participants who endorsed more racist beliefs about minorities and crime appeared less sensitive to the linguistic features when making their face selections, and those who showed more subtle racism against Black Americans appeared more sensitive to those features. The latter finding is consistent with prior work that found those with higher implicit racism against Blacks showed a stronger relationship between skin tone and judgments of race typicality (Stepanova and Strube, 2012a) and suggests that implicit or less direct racial bias influences how individuals categorize others. Why those with more racist beliefs related to crime were less likely rather than more likely to discriminate between the faces remains to be determined, but the low reliability of this scale may have been a factor.

The stereotypicality of the voice did not appear to affect their confidence in their judgments, and this pattern remained when the covariates were added. However, the racial bias subscale of the PJAQ and the self-deceptive enhancement subscale of the BIDR-16 did influence confidence ratings. Those who indicated more racial bias were more confident in their decisions. As the other racism measure was not a significant predictor, it is again unclear what aspect of this racial bias subscale is related to confidence. Those who indicated a greater tendency to engage in unintentional socially desirable responding were also more confident in their decisions, likely reflecting their desire to respond in an overly positive manner (Hart et al., 2015).

From these findings, it appears that how a speaker sounds activates certain expectations about what that speaker should look like, and that these expectations persist regardless of a listeners' self-reported attitudes about Black Americans, their own racial or ethnic identity, or their propensity to respond in a socially desirable manner.

GENERAL DISCUSSION

Individuals are not evaluated in a vacuum. Rather, people use multiple cues to classify individuals into social categories, which can then activate assumptions about that person's abilities, traits, and social status. Individuals seen as more typical members of their given social group are more strongly associated with stereotypes about that group (Blair et al., 2002; Walker and Wänke, 2017). Most studies exploring the relationship between how stereotypical an individual is perceived to be and the activation of stereotypes focus on stereotypicality in terms of physical features. We demonstrated that language is another important source of social category information that informs stereotypicality judgments. In two experiments, we investigated whether individuals who sound more stereotypically Black are also more likely to be seen as more stereotypically Black in terms of character traits (Experiment 1) and in terms of their appearance (Experiment 2). Individuals perceived by listeners as sounding more stereotypically Black were associated with more stereotypical traits about Black Americans and with more phenotypically Black faces. Importantly, these findings suggest that speech stereotypicality may operate similarly to phenotypicality, as the linguistic features associated with Black Americans appear to activate stereotypes about members of that group without the listener's knowledge of the actual race of the speaker.

In Experiment 1, we extended work on perceived stereotypicality and the activation of stereotypical traits from the domain of phenotypicality to speech stereotypicality. Speakers who were rated as sounding more stereotypically Black were assigned a greater proportion of stereotypical traits associated with Black Americans than those who were rated as sounding less stereotypically Black. These results are in line with previous work that greater use of AAVE features, which may be comparable to sounding more stereotypically Black, led to lower ratings of attractiveness and social status (Rodriguez et al., 2004). Our finding also suggests that poorer ratings of speakers who use more AAVE features may be related to a greater activation of stereotypes about Blackness, many of which are negative (e.g., criminal, aggressive, uneducated, lazy).

We also found evidence that this relationship between race, speech stereotypicality, and assigned stereotypic traits only applied to those who were perceived as both Black and North American. This finding, along with the fact that our U.S.-based listeners were less likely to identify UK speakers as Black and did not categorize our Black speakers with 100% accuracy, is consistent with the expectations of the Dialectal-Race Hypothesis; that is, to categorize speakers, listeners rely on their cultural knowledge of how a group of speakers talk (Perrachione et al., 2010). If our listeners were relying only on acoustic features caused by anatomical differences that may exist between groups, we would expect that listeners would have no problems correctly identifying the race of our Black speakers and that perceptions of "sounding stereotypically Black" alone, regardless of country of origin, would be sufficient to activate stereotypes about Black Americans. Instead, our results suggest that listeners' knowledge of dialects influenced how they attributed traits to stereotypical-sounding Black speakers from different regions.

In Experiment 2, speech stereotypicality activated expectations about phenotypicality, such that strongly stereotypical Black voices were associated with more

phenotypically Black faces, and weakly stereotypical Black voices were associated with less phenotypically Black faces. While some physical characteristics do influence how speakers sound, such as height and weight (Krauss et al., 2002) or age-related changes to the structure of the vocal tract (Caruso et al., 1995), many of the associations between voice and appearance are informed by the surrounding cultural context. Given that there is no inherent reason why more stereotypical-sounding Black speakers should also look more stereotypically Black, the results from Experiment 2 suggest that the linguistic cues that listeners have associated with being more stereotypically Black activate expectations about the speaker's appearance.

Notably, while anti-Black bias can influence face identifications (e.g., when choosing between faces with lighter or darker skin; Alter et al., 2016), the relationship between voice and face stereotypicality persisted regardless of participants' social attitudes. As is true with physical Afrocentric features, listeners appear unaware of the influence of linguistic features on their judgments. Blair and colleagues (Blair et al., 2002, 2004b) found that individuals relied on Afrocentric features independently from race to assign stereotypic traits, and that although they could suppress the influence from race, they could not suppress the influence of Afrocentric features, even when they were instructed to. Further, their inability to suppress the influence of Afrocentric features was not due to ignorance, as individuals were able to identify Afrocentric features when asked. Linguistic cues associated with "sounding Black" could be another example of feature-based stereotyping. Participants in our studies may have been similarly unaware that they were relying on these cues to assign stereotypic traits or to make face judgments because discrimination against others for their way of speaking is not particularly taboo in American society. It remains to be determined if participants could suppress the influence of linguistic features when made aware of them. More conclusive research on this topic is needed to determine whether linguistic features are used for feature-based stereotyping, particularly to establish whether these cues operate independently from race, as all of the speakers in our study were from the same racial group.

These results have critical implications about how speakers who sound more stereotypically Black may be perceived and treated by others. Previous work has found that those who use AAVE are less likely to have access to housing compared to those using General American English (Purnell et al., 1999; Massey and Lundy, 2001), and those who can be identified as Black from speech earn 12% less than their White counterparts, even when controlling for skill, family background, and schooling (Grogger, 2011). How speech patterns influence such biases, particularly wage discrimination, is unknown, but stereotypicality is likely a contributing factor. Concerningly, given the association between Blackness and criminality, speakers who sound more stereotypically Black may also face similar biases within the legal system. Mock courtroom studies have previously found that in some circumstances defendants who are Black and use a non-standard dialect are judged more harshly than those who use a more standard dialect (Dixon et al., 2002; Cantone et al., 2019; cf. Kurinec and Weaver, 2019). However, whether the perceived stereotypicality of the speaker moderates this effect and whether it influences other judgments, such as those made by law enforcement, judges, and eyewitnesses, remains an area for investigation.

That speech stereotypicality could influence decisions in the legal system despite people's awareness of anti-Black discrimination in the field is not entirely unfounded, as stereotypical features have previously been shown to affect these types of judgments. More phenotypically Black individuals are more likely to be misidentified by eyewitnesses (Knuycky et al., 2014; Kleider-Offutt et al., 2017), particularly when other Black stereotypes (e.g., drug dealer) are activated (Kleider et al., 2012; Osborne and Davies, 2013). Additionally, regardless of race, those with more Afrocentric features are more likely to receive longer criminal sentences (Blair et al., 2004a). Speech stereotypicality may operate in a similar manner. For instance, individuals who sound more stereotypically Black may be expected to appear more phenotypically Black, biasing eyewitness's memory for the individual and leading to misidentifications. Separately, those who sound more Black may directly activate negative stereotypes about Black Americans, impacting identifications and legal judgments. Future research is needed to determine how speech stereotypicality may influence eyewitness identifications and other forms of legal decision-making.

Although this work provides initial evidence that speech stereotypicality activates racial stereotypes, several topics warrant further investigation. First, both experiments in this study used predominantly non-Black samples, with the majority of participants identifying as White. While it is possible that using such a sample may have exacerbated out-group biases, strengthening the relationships between speech stereotypicality and stereotypical traits or more phenotypical faces, previous work has found that Black Americans often report similar perceptions of AAVE as White Americans, rating the use of AAVE more negatively than standard American English (Doss and Gross, 1992; Payne et al., 2000; McCluney et al., 2021). Further, previous work looking at the stereotypes associated with light- and dark-skinned Blacks found that both White and Black participants associated more stereotypical Black traits with darker-skinned Blacks (Maddox and Gray, 2002). Given that darker skin is seen as more phenotypically Black, it is likely that a sample of only Black Americans would similarly associate more Black stereotypes—to include stereotypes about appearance—with those perceived as sounding more stereotypically Black. However, it is important to note that there are instances when Black Americans express more favorable opinions of AAVE, such as in less formal community settings or when the individual has greater commitment to their Black identity (White et al., 1998; Koch et al., 2001; Rahman, 2008). Thus, whether these effects would replicate in an all-Black sample may depend on how formal the participants perceive the setting and the strength of their identification as members of the Black community.

Second, the experiments in this study intentionally included multiple speakers using different dialects at differing degrees, both to capture more natural speech utterances and, in Experiment 1, to provide listeners with a range of speakers to evaluate. Had we utilized a handful of speakers switching between a more or less stereotypical-sounding style of speech, we could

have risked speakers sounding unauthentic or forced, impacting our ratings (Garrett, 2010; Guy and Cutler, 2011). However, if at all possible, future work should try to find speakers who are sufficiently proficient in shifting their degree of AAVE or similar dialect use. This would allow researchers to implement the matched-guise technique, which utilizes the same speaker for two or more dialects or languages (Lambert, 1967), minimizing the influence of any linguistic variables that are not of interest (e.g., intonation, pace) while maintaining the desired range in dialect strength in an authentic way.

Third, we relied on listeners' ratings to determine how stereotypical speakers sounded. Listeners have been shown to have similar ideas of the standardness of a given dialect as that of researchers (Schüppert et al., 2015; Nejjari et al., 2019), but it is unclear whether this reliability in ratings extends to perceived stereotypicality. However, what constitutes stereotypically Black speech is likely subjective, depending on an individual listener's own familiarity with Black speech and their social or cultural knowledge of what does or doesn't sound stereotypical. It would be beneficial to have a more objective approach to stereotypicality research. For instance, the audio recordings used in stereotypicality research could be analyzed for the linguistic features present to determine whether an increase in the quantity of linguistic features associated with Black Americans leads to an increase in perceived stereotypicality, or if there are specific features that trigger such judgments.

Finally, the speakers in this study varied in the content of their utterances. The speakers in Experiment 1 talked about a variety of topics and spoke in either more casual (e.g., video blog) or more formal (e.g., panel talk) settings. Alternatively, the speakers in Experiment 2 read one of two standard texts. On one hand, the audio used in Experiment 1 likely provided more natural utterances for listeners to evaluate. On the other hand, the audio in Experiment 2 more carefully controlled for the content of the utterances, and previous research suggests that judgments about speakers who sound Black may be moderated by their message content (Johnson and Buttny, 1982). Thus, it would be worthwhile for future research to find a balance between natural sounding utterances and ensuring that the content of said utterances is suitably limited to avoid influencing judgments of stereotypicality.

In sum, this work provides an initial look into the relationship between how stereotypically Black an individual speaker sounds and the activation of stereotypes about character traits and phenotypicality. These findings extend previous work on stereotypicality and suggest that other aspects of an individual beyond physical appearance may serve as cues that inform how stereotypical an individual is believed to be. Since discrimination based on how a person speaks is not explicitly prohibited under U.S. law, understanding how speech stereotypicality contributes to judgments about others is needed to guide the development of regulations and other protections to ensure more equitable treatment for such speakers in housing, employment, and the legal system. Those currently in those fields should be mindful of how speech stereotypicality may be influencing decision-making.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://osf.io/ungrs/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Baylor University Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CK and CW contributed to the conception and design of the study. CK oversaw the data collection and analysis and wrote the initial draft of the manuscript. Both authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.785283/full#supplementary-material

REFERENCES

Alter, A. L., Stern, C., Granot, Y., and Balcetis, E. (2016). The "bad is black" effect: why people believe evildoers have darker skin than do-gooders. *Pers. Soc. Psychol. Bull.* 42, 1653–1665. doi: 10.1177/0146167216669123

Andreoletti, C., Leszczynski, J. P., and Disch, W. B. (2015). Gender, race, and age: the content of compound stereotypes across the life span. *Int. J. Aging Hum. Dev.* 81, 27–53. doi: 10.1177/0091415015616395

Anisfeld, M., Bogo, N., and Lambert, W. E. (1962). Evaluational reactions to accented English speech. *J. Abnormal Soc. Psychol.* 65, 223–231. doi: 10.1037/h0045060

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Statist. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Billings, A. C. (2005). Beyond the ebonics debate: attitudes about Black and Standard American English. *J. Black Stud.* 36, 68–81. doi: 10.1177/0021934704271448

Blair, I. V., Judd, C. M., and Chapleau, K. M. (2004a). The influence of afrocentric facial features in criminal sentencing. *Psychol. Sci.* 15, 674–679. doi: 10.1111/j.0956-7976.2004.00739.x

Blair, I. V., Judd, C. M., and Fallman, J. L. (2004b). The automaticity of race and afrocentric facial features in social judgments. *J. Pers. Soc. Psychol.* 87, 763–778. doi: 10.1037/0022-3514.87.6.763

Blair, I. V., Judd, C. M., Sadler, M. S., and Jenkins, C. (2002). The role of Afrocentric features in person perception: judging by features and categories. *J. Pers. Soc. Psychol.* 83, 5–25. doi: 10.1037/0022-3514.83.1.5

Bucholtz, M., and Hall, K. (2005). Identity and interaction: a sociocultural linguistic approach. *Discourse Stud.* 7, 585–614. doi: 10.1177/1461445605054407

Campbell-Kibler, K. (2012). The Implicit Association Test and sociolinguistic meaning. *Lingua* 122, 753–763. doi: 10.1016/j.lingua.2012.01.002

Cantone, J. A., Martinez, L. N., Willis-Esqueda, C., and Miller, T. (2019). Sounding guilty: how accent bias affects juror judgments of culpability. *J. Ethn. Crim. Justice* 17, 228–253. doi: 10.1080/15377938.2019.1623963

Cargile, A. C., and Bradac, J. J. (2001). Attitudes toward language: a review of speaker-evaluation research and a general process model. *Ann. Int. Commun. Assoc.* 25, 347–382. doi: 10.1080/23808985.2001.11679008

Carter, P. M., and Callesano, S. (2018). The social meaning of Spanish in Miami: dialect perceptions and implications for socioeconomic class, income, and employment. *Latino Stud.* 16, 65–90. doi: 10.1057/s41276-017-0105-8

Caruso, A. J., Mueller, P. B., and Shadden, B. B. (1995). Effects of aging on speech and voice. *Phys. Occup. Ther. Geriatrics* 13, 63–79. doi: 10.1080/J148v13n01_04

Cutler, C. (2003). "Keepin' It Real": white hip-hoppers' discourses of language, race, and authenticity. *J. Linguist. Anthropol.* 13, 211–233. doi: 10.1525/jlin.2003.13.2.211

Demby, G. (2013). *Where Did That Fried Chicken Stereotype Come From?* United States: NPR. Available online at: https://www.npr.org/sections/codeswitch/2013/05/22/186087397/where-did-that-fried-chicken-stereotype-come-from (accessed August 10, 2021).

Dent, S. (2004). *Attitudes of Native and Nonnative Speakers of English Toward Various Regional and Social U.S. English Accents.* United States: Iowa State University. doi: 10.31274/rtd-180813-5645

Devine, P. G., and Elliot, A. J. (1995). Are racial stereotypes really fading? The Princeton trilogy revisited. *Pers. Soc. Psychol. Bull.* 21, 1139–1150. doi: 10.1177/01461672952111002

Dixon, J. A., Mahoney, B., and Cocks, R. (2002). Accents of guilt? effects of regional accent, race, and crime type on attributions of guilt. *J. Lang. Soc. Psychol.* 21, 162–168. doi: 10.1177/02627X02021002004

Doss, R. C., and Gross, A. M. (1992). The effects of Black English on stereotyping in intraracial perceptions. *J. Black Psychol.* 18, 47–58. doi: 10.1177/00957984920182005

Dragojevic, M., Berglund, C., and Blauvelt, T. K. (2018). Figuring out who's who: the role of social categorization in the language attitudes process. *J. Lang. Soc. Psychol.* 37, 28–50. doi: 10.1177/0261927X17706942

Dunham, Y., Stepanova, E. V., Dotsch, R., and Todorov, A. (2015). The development of race-based perceptual categorization: skin color dominates early category judgments. *Dev. Sci.* 18, 469–483. doi: 10.1111/desc.12228

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., and Johnson, S. L. (2006). Looking deathworthy: perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychol. Sci.* 17, 383–386. doi: 10.1111/j.1467-9280.2006.01716.x

Eberhardt, J. L., Goff, P. A., Purdie, V. J., and Davies, P. G. (2004). Seeing Black: race, crime, and visual processing. *J. Pers. Soc. Psychol.* 87, 876–893. doi: 10.1037/0022-3514.87.6.876

Elliott, S. L., Chu, K., and Coleman, J. (2017). Attitudes about race predict individual differences in face adaptation aftereffects. *Vis. Res.* 141, 237–246. doi: 10.1016/j.visres.2016.12.011

Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160. doi: 10.3758/BRM.41.4.1149

Gaither, S. E., Cohen-Goldberg, A. M., Gidney, C. L., and Maddox, K. B. (2015). Sounding Black or White: priming identity and biracial speech. *Front. Psychol.* 6:457. doi: 10.3389/fpsyg.2015.00457

Garrett, P. (2010). *Attitudes to Language.* Cambridge: Cambridge University Press.

Giles, H. (1970). Evaluative reactions to accents. *Educ. Rev.* 22, 211–227. doi: 10.1080/0013191700220301

Giles, H., and Rakić, T. (2014). *Language Attitudes: Social Determinants and Consequences of Language Variation.* Oxford: Oxford University Press. 11–26.

Grogger, J. (2011). Speech patterns and racial wage inequality. *J. Hum. Resour.* 46, 1–25. doi: 10.3368/jhr.46.1.1

Guy, G. R., and Cutler, C. (2011). Speech style and authenticity: quantitative evidence for the performance of identity. *Lang. Variat. Change* 23, 139–162. doi: 10.1017/S0954394510000232

Hagiwara, N., Kashy, D. A., and Cesario, J. (2012). The independent effects of skin tone and facial features on Whites' affective reactions to Blacks. *J. Exp. Soc. Psychol.* 48, 892–898. doi: 10.1016/j.jesp.2012.02.001

Hart, C. M., Ritchie, T. D., Hepper, E. G., and Gebauer, J. E. (2015). The Balanced Inventory of Desirable Responding Short Form (BIDR-16). *SAGE Open* 5, 1–9. doi: 10.1177/2158244015621113

Henry, P. J., and Sears, D. O. (2002). The Symbolic Racism 2000 Scale. *Polit. Psychol.* 23, 253–283. doi: 10.1111/0162-895X.00281

Hughes, S. M., and Miller, N. E. (2016). What sounds beautiful looks beautiful stereotype: the matching of attractiveness of voices and faces. *J. Soc. Pers. Relat.* 33, 984–996. doi: 10.1177/0265407515612445

International Dialects of English Archive [IDEA] (2011). *International Dialects of English Archive.* Available online at: https://www.dialectsarchive.com (accessed August 13, 2021).

Johnson, F. L., and Buttny, R. (1982). White listeners' responses to "sounding Black" and "sounding White": the effects of message content on judgments about language. *Commun. Monogr.* 49, 33–49. doi: 10.1080/03637758209376069

Jones, T. (2015). Toward a description of African American Vernacular English Dialect regions using "Black Twitter". *Am. Speech* 90, 403–440. doi: 10.1215/00031283-3442117

Kleider, H. M., Cavrak, S. E., and Knuycky, L. R. (2012). Looking like a criminal: stereotypical black facial features promote face source memory error. *Mem. Cogn.* 40, 1200–1213. doi: 10.3758/s13421-012-0229-x

Kleider-Offutt, H. M., Knuycky, L. R., Clevinger, A. M., and Capodanno, M. M. (2017). Wrongful convictions and prototypical black features: can a face-type facilitate misidentifications? *Legal Criminol. Psychol.* 22, 350–358. doi: 10.1111/lcrp.12105

Knuycky, L. R., Kleider, H. M., and Cavrak, S. E. (2014). Line-up misidentifications: when being 'prototypically Black' is perceived as criminal. *Appl. Cogn. Psychol.* 28, 39–46. doi: 10.1002/acp.2954

Koch, L. M., Gross, A. M., and Kolts, R. (2001). Attitudes toward Black English and code switching. *J. Black Psychol.* 27, 29–42. doi: 10.1177/0095798401027001002

Krauss, R. M., Freyberg, R., and Morsella, E. (2002). Inferring speakers' physical attributes from their voices. *J. Exp. Soc. Psychol.* 38, 618–625. doi: 10.1016/S0022-1031(02)00510-3

Kurinec, C. A., and Weaver, C. A. (2019). Dialect on trial: use of African American Vernacular English influences juror appraisals. *Psychol. Crime Law* 25, 803–828. doi: 10.1080/1068316X.2019.1597086

Kushins, E. R. (2014). Sounding like your race in the employment process: an experiment on speaker voice, race identification, and stereotyping. *Race Soc. Problems* 6, 237–248. doi: 10.1007/s12552-014-9123-4

Labov, W., Ash, S., and Boberg, C. (2008). *The Atlas of North American English: Phonetics, Phonology and Sound Change.* Berlin: Walter de Gruyter.

Lambert, W. E. (1967). A social psychology of bilingualism. *J. Soc. Issues* 23, 91–109. doi: 10.1111/j.1540-4560.1967.tb00578.x

Lecci, L., and Myers, B. (2008). Individual differences in attitudes relevant to juror decision making: development and validation of the Pretrial Juror Attitude Questionnaire (PJAQ). *J. Appl. Soc. Psychol.* 38, 2010–2038. doi: 10.1111/j.1559-1816.2008.00378.x

Lenth, R. (2019). *emmeans: Estimated Marginal Means, aka Least-Squares Means.* Available online at: https://CRAN.R-project.org/package=emmeans (accessed December 8, 2020).

Lippi-Green, R. (1997). *The Standard Language Myth English With an Accent: Language, Ideology, and Discrimination in the United States.* London: Routledge. 53–62.

Litman, L., Robinson, J., and Abberbock, T. (2017). TurkPrime.com: a versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behav. Res. Methods* 49, 433–442. doi: 10.3758/s13428-016-0727-z

Loudermilk, B. C. (2015). "Implicit attitudes and the perception of sociolinguistic variation" in *Responses to Language Varieties: Variability, Processes and Outcomes*. eds A. Prikhodkine and D. R. Preston (Netherlands: John Benjamins Publishing Company), 137–156. doi: 10.1075/impact.39.06lou

Luhtanen, R., and Crocker, J. (1992). A collective self-esteem scale: self-evaluation of one's social identity. *Pers. Soc. Psychol. Bull.* 18, 302–318. doi: 10.1177/0146167292183006

Ma, D. S., Correll, J., and Wittenbrink, B. (2015). The Chicago face database: a free stimulus set of faces and norming data. *Behav. Res. Methods* 47, 1122–1135. doi: 10.3758/s13428-014-0532-5

MacNeal, A., Fiallo, K., Jones, A., Jones, S., Laureano, S., Monjarrez, M., et al. (2019). "Sounding Black": the legal implications of linguistic profiling" in *Northeastern University Working Papers in Linguistics* (Boston, MA, Northeastern University), 4.

Maddox, K. B. (2004). Perspectives on racial phenotypicality bias. *Pers. Soc. Psychol. Rev.* 8, 383–401. doi: 10.1207/s15327957pspr0804_4

Maddox, K. B., and Gray, S. A. (2002). Cognitive representations of Black Americans: reexploring the role of skin tone. *Pers. Soc. Psychol. Bull.* 28, 250–259. doi: 10.1177/0146167202282010

Marques, J. M., Yzerbyt, V. Y., and Leyens, J.-P. (1988). The "Black Sheep Effect": extremity of judgments towards ingroup members as a function of group identification. *Eur. J. Soc. Psychol.* 18, 1–16. doi: 10.1002/ejsp.2420180102

Massey, D. S., and Lundy, G. (2001). Use of Black English and racial discrimination in urban housing markets: new methods and findings. *Urban Affairs Rev.* 36, 452–469. doi: 10.1177/10780870122184957

McCluney, C. L., Durkee, M. I., Smith, R., Robotham, K. J., and Lee, S. S.-L. (2021). To be, or not to be…Black: the effects of racial codeswitching on perceived professionalism in the workplace. *J. Exp. Soc. Psychol.* 97:104199. doi: 10.1016/j.jesp.2021.104199

Milroy, J. (2001). Language ideologies and the consequences of standardization. *J. Socioling.* 5, 530–555. doi: 10.1111/1467-9481.00163

Milroy, J., and Milroy, L. (2012). *Authority in Language: Investigating Standard English*. London: Routledge. doi: 10.4324/9780203124666

Morales, A. C., Scott, M. L., and Yorkston, E. A. (2012). The role of accent standardness in message preference and recall. *J. Advert.* 41, 33–46. doi: 10.2753/JOA0091-3367410103

Mulac, A., and Rudd, M. J. (1977). Effects of selected American regional dialects upon regional audience members. *Commun. Monogr.* 44, 185–195. doi: 10.1161/STR.0b013e3182213e24

Munson, B., and Solum, R. (2010). "When is indexical information about speech activated? evidence from a cross-modal priming experiment" in *INTERSPEECH 2010 (Makuhari, Chiba, Japan)*, 1521–1524. Available online at: https://www.isca-speech.org/archive/interspeech_2010/munson10_interspeech.html (accessed August 13, 2021).

Nejjari, W., Gerritsen, M., van Hout, R., and Planken, B. (2019). Refinement of the matched-guise technique for the study of the effect of non-native accents compared to native accents. *Lingua* 219, 90–105. doi: 10.1016/j.lingua.2018.12.001

Osborne, D., and Davies, P. G. (2013). Eyewitness identifications are affected by stereotypes about a suspect's level of perceived stereotypicality. *Group Proces. Intergroup Rel.* 16, 488–504. doi: 10.1177/1368430212454927

Payne, K., Downing, J., and Fleming, J. C. (2000). Speaking Ebonics in a professional context: the role of ethos/source credibility and perceived sociability of the speaker. *J. Techn. Writ. Commun.* 30, 367–383. doi: 10.2190/93U1-0859-0VC3-F5LK

Perrachione, T. K., Chiao, J. Y., and Wong, P. C. M. (2010). Asymmetric cultural effects on perceptual expertise underlie an own-race bias for voices. *Cognition* 114, 42–55. doi: 10.1016/j.cognition.2009.08.012

Phillips, T. (2010). Put your money where your mouth is: the effects of southern vs. standard accent on perceptions of speakers. *Soc. Sci.* 53–57. Available online at: http://www.stanford.edu/group/journal/cgi-bin/wordpress/wpcontent/uploads/2012/09/Phillips_SocSci_2010.pdf

Preston, D. R. (2018). Changing research on the changing perceptions of Southern U.S. English. *Am. Speech* 93, 471–496. doi: 10.1215/00031283-7271283

Pullum, G. K. (1999). "African American vernacular English is not standard English with mistakes," in *The Workings of Language: From Prescriptions to Perspectives*, ed. R. S. Wheeler (Westport, CT: Praeger Publishers), 39–58.

Purnell, T., Idsardi, W., and Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *J. Lang. Soc. Psychol.* 18, 10–30. doi: 10.1177/0261927X99018001002

Rahman, J. (2008). Middle-class African Americans: reactions and attitudes toward African American English. *Am. Speech* 83, 141–176. doi: 10.1215/00031283-2008-009

Rakiç, T., Steffens, M. C., and Mummendey, A. (2011a). Blinded by the accent! The minor role of looks in ethnic categorization. *J. Pers. Soc. Psychol.* 100, 16–29. doi: 10.1037/a0021522

Rakiç, T., Steffens, M. C., and Mummendey, A. (2011b). When it matters how you pronounce it: the influence of regional accents on job interview outcome. *Br. J. Psychol.* 102, 868–883. doi: 10.1111/j.2044-8295.2011.02051.x

Revelle, W. (2017). *psych: Procedures for Personality and Psychological Research*. Evanston: Northwestern University.

Rickford, J. (1999). *African American Vernacular English: Features, Evolution, Educational Implications*. Malden: Blackwell, 3–14.

Rodriguez, J. I., Cargile, A. C., and Rich, M. D. (2004). Reactions to African-American Vernacular English: do more phonological features matter? *Western J. Black Stud.* 28, 407–414.

Rosen, R. K. (2017). *What's in a Voice? Effects of Dialect Perception on Activation of Crime Stereotypes*. Chicago, IL: University of Chicago.

RStudio Team. (2019). *RStudio: Integrated Development for R*. Boston: RStudio, Inc.

Schüppert, A., Hilton, N. H., and Gooskens, C. (2015). Swedish is beautiful, Danish is ugly? Investigating the link between language attitudes and spoken word recognition. *Linguistics* 53, 375–403. doi: 10.1515/ling-2015-0003

Shamina, E. (2016). "An experimental study of English accent perception" in *Proceedings of 7th Tutorial and Research Workshop on Experimental Linguistics*, ed. A. Botinis (St. Petersburg: International Speech Communication Association). doi: 10.36505/ExLing-2016/07/0036/000295

Singmann, H., Bolker, B., Westfall, J., Aust, F., and Ben-Shachar, M. S. (2019). *afex: analysis of Factorial Experiments version 0.28-0 from CRAN*. Available online at: https://CRAN.R-project.org/package=afex (accessed December 8, 2020).

Slade, A., and Narro, A. J. (2012). "An acceptable stereotype: the Southern image in television programming" in *Mediated images of the South: the Portrayal of Dixie in Popular Culture*. eds A. Slade, A. J. Narro, and D. Givens-Carroll (Lanham: Lexington Books), 5–19.

Steffensmeier, D., Ulmer, J., and Kramer, J. (1998). The interaction of race, gender, and age in criminal sentencing: the punishment cost of being young. Black, and male. *Criminology* 36, 763–798. doi: 10.1111/j.1745-9125.1998.tb01265.x

Stepanova, E. V., and Strube, M. J. (2009). Making of a face: role of facial physiognomy, skin tone, and color presentation mode in evaluations of racial typicality. *J. Soc. Psychol.* 149, 66–81. doi: 10.3200/SOCP.149.1.66-81

Stepanova, E. V., and Strube, M. J. (2012b). What's in a face? The role of skin tone, facial physiognomy, and color presentation mode of racial primes in affective priming effects. *J. Soc. Psychol.* 152, 212–227. doi: 10.1080/00224545.2011.597797

Stepanova, E. V., and Strube, M. J. (2012a). The role of skin color and facial physiognomy in racial categorization: moderation by implicit racial attitudes. *J. Exp. Soc. Psychol.* 48, 867–878. doi: 10.1016/j.jesp.2012.02.019

Stepanova, E. V., and Strube, M. J. (2018). Attractiveness as a function of skin tone and facial features: evidence from categorization studies. *J. General Psychol.* 145, 1–20. doi: 10.1080/00221309.2017.1394811

Stöber, J., Dette, D. E., and Musch, J. (2002). Comparing continuous and dichotomous scoring of the Balanced Inventory of Desirable Responding. *J. Pers. Assess.* 78, 370–389. doi: 10.1207/S15327752JPA7802_10

Thomas, E. R. (2007). Phonological and phonetic characteristics of African American Vernacular English. *Lang. Linguist. Compass* 1, 450–475. doi: 10.1111/j.1749-818X.2007.00029.x

Thomas, E. R., and Reaser, J. (2004). Delimiting perceptual cues used for the ethnic labeling of African American and European American voices. *J. Sociolinguist.* 8, 54–87. doi: 10.1111/j.1467-9841.2004.00251.x

Walker, M., and Wänke, M. (2017). Caring or daring? Exploring the impact of facial masculinity/femininity and gender category information on first impressions. *PLoS One* 12:e0181306. doi: 10.1371/journal.pone.0181306

Walton, J. H., and Orlikoff, R. F. (1994). Speaker race identification from acoustic cues in the vocal signal. *J. Speech Lang. Hear. Res.* 37, 738–745. doi: 10.1044/jshr.3704.738

Warnes, G. R., Bolker, B., Lumley, T., and Johnson, R. C. (2018). *gmodels: various R Programming Tools for Model Fitting*. United States: SAIC-Frederick, Inc.

White, M. J., Vandiver, B. J., Becker, M. L., Overstreet, B. G., Temple, L. E., Hagan, K. L., et al. (1998). African American evaluations of Black English and standard American English. *J. Black Psychol.* 24, 60–75. doi: 10.1177/0095798498024 01005

Wickham, H., and Miller, E. (2019). *haven: import and Export "SPSS", "Stata" and "SAS" Files*. Available online at: https: //CRAN.R-project.org/package=haven (accessed December 8, 2020).

Wiehl, L. (2002). Sounding Black in the courtroom: court-sanctioned racial stereotyping. *Harv. Blackletter L. J.* 18:185.

Zuckerman, M., and Driver, R. E. (1989). What sounds beautiful is good: the vocal attractiveness stereotype. *J. Nonverb. Behav.* 13, 67–82. doi: 10.1007/ BF00990791

Value Alignment and Public Perceived Legitimacy of the European Union and the Court of Justice

Eva Grosfeld[1]*, Daan Scheepers[2,3] and Armin Cuyvers[1]

[1]*Europa Institute, Leiden Law School, Leiden University, Leiden, Netherlands, [2]Department of Psychology, Faculty of Social and Behavioural Sciences, Leiden University, Leiden, Netherlands, [3]Department of Psychology, Faculty of Social and Behavioural Sciences, Utrecht University, Utrecht, Netherlands*

The present study aims to extend research on the role of values for the perceived legitimacy of legal authorities by focusing on (1) supranational legal authorities and (2) a broad range of values. We examine how (alignment between) people's personal values and their perception of the values of the European Union (EU) are related to perceived legitimacy of the Court of Justice of the EU (CJEU) and the EU more broadly. Inspired by moral foundations theory, we distinguish between individualizing (i.e., "democracy", "liberty", and "fairness") and binding values (i.e., "rule of law", "respect for national authority", and "respect for tradition"). An online survey was conducted in six EU member states (*N* = 1,136). A factor analysis confirmed a two-factor model (individualizing vs. binding values) for both personal values and perceived EU values. Four regression models were run for each of the value factors, including personal values, perceived EU values, and their interaction, on each of the outcomes (i.e., perceived CJEU and EU legitimacy). Perceived endorsement by the EU of both individualizing and binding values predicted higher legitimacy perceptions of the CJEU and EU. Furthermore, personal binding values had a negative effect on perceived EU legitimacy when participants perceived the EU to weakly support binding values, but a positive effect when the EU was perceived to strongly support binding values. The results suggest that value alignment plays an important role in perceived legitimacy of the CJEU and EU, and that better representing binding values might be a strategy to improve perceived EU legitimacy.

Keywords: perceived legitimacy, court of justice, European Union, value alignment, moral foundations

INTRODUCTION

Although disputed, the perceived legitimacy of the Court of Justice of the European Union (CJEU) is according to some studies declining (Pollack, 2018). Perceived legitimacy can be defined as the belief that an institution exercises its authority appropriately (Tyler, 2006). Analyses of Eurobarometer data suggest that since 2010, following the trend of trust in the European Union (EU) more generally, public trust in the CJEU has declined while distrust has increased (Pollack, 2018). The same may be true for perceptions of CJEU legitimacy among national

authorities, who show resistance in terms of non-compliance with CJEU rulings and efforts to limit the effectiveness of CJEU decisions (Hofmann, 2018). Regardless of whether the CJEU's and EU's legitimacy levels are actually declining or not, it is generally agreed upon that legal authorities require a widespread basis of perceived legitimacy to maintain social order, settle disputes, and solve societal issues (Tyler and Lind, 1992; Tyler, 2006; Trinkner and Tyler, 2016). Moreover, low legitimacy may have far-reaching consequences for the EU as a whole. For example, taking back control over British law was one of the red lines of the Leave-campaign in the Brexit referendum. It is therefore important to understand when and why people perceive the CJEU and EU as legitimate.

In this brief research report, we examine how people's personal values, their perceived values of the EU, and alignment between these values relate to public perceived legitimacy of the CJEU and EU. In doing so, we will look beyond "individualizing" values and also consider "binding" values. In what follows, we first discuss theories on how moral judgments may influence the perceived legitimacy of legal authorities, and then elaborate on how individual differences in moral intuitions may explain why some people perceive the CJEU and EU to be legitimate and others do not.

Through interactions with the legal system throughout their lives, individuals develop a relationship with legal authorities. When this relationship is based on the subjective belief that power is exercised appropriately, rather than on fear for punishments, people are more likely to accept the law, even when it goes against their own self-interest (Trinkner and Tyler, 2016). Such legitimacy attributions develop in an ongoing dialogue between power-holders, which claim that their authority and exercise of power are rightful, and members of the "audience", which process and respond to these claims (Bottoms and Tankebe, 2012).

Legal authorities draw a large part of their legitimacy from "value alignment", that is, the extent to which the values they endorse align with people's personal values (Jackson et al., 2012). There are two routes through which value alignment is thought to promote legitimacy. First, shared values communicate to people that they are valued members of the group, which provides them with status and a positive social identity (Tyler and Lind, 1992; Tyler, 1997). As a consequence, personal concerns become less relevant, and people are more likely to internalize the conviction that it is right to obey the rules which are imposed upon them (Tyler and Jackson, 2013). Second, the belief that an authority is acting morally appropriate and in line with one's own sense of right and wrong normatively validates its power and forms a source of trust (Suchman, 1995; Jackson et al., 2015).

A central way in which *national* legal authorities create value alignment with their audience is through procedural justice. Perceived fair procedures contribute to positive group identification and express the moral appropriateness of authority (Tyler and Blader, 2003; Jackson et al., 2012, 2015). Over the past decades, research has convincingly shown that perceived procedural justice is positively related to the perceived legitimacy of the police, judges, and other court officials (e.g., Sunshine

and Tyler, 2003; van den Bos et al., 2014; Tyler et al., 2015; Grootelaar and van den Bos, 2018; see for a meta-analysis Walters and Bolger, 2019). In sum, legal authorities can generate legitimacy by demonstrating value alignment through procedural justice.

However, this is not to say that procedural justice is the only foundation of value alignment. The expression of other values, such as effectiveness or distributive justice, may also justify the exercise of legal power (Bottoms and Tankebe, 2012; Jackson et al., 2015). Where *supranational* authorities are concerned, values such as democracy and transparency, have, for example, been found relevant (Dellmuth et al., 2019). For the legitimacy of the CJEU's supranational authority, procedural justice may also play a less important role. Although EU law is an integral part of national legal systems, and the CJEU plays a central role in upholding EU law and safeguarding its uniform interpretation and application, lay people seldom interact with the CJEU, and many are even not very aware of its existence (Caldeira and Gibson, 1995). To begin with, the standing of non-privileged parties, such as private individuals, for direct actions to the CJEU is very limited. In addition, the chances of an individual ending up in front of the CJEU via a preliminary reference procedure—in which national courts ask the CJEU for a judgment on the interpretation or validity of EU law within the context of a national dispute—are extremely limited as well (see judicial activity in the annual report of the CJEU, 2020). For these reasons, legitimacy of the CJEU as perceived by the public may not solely rely on procedural justice.

Moreover, when people have no information about the trustworthiness or objective legitimacy of a supranational organization, they are inclined to resort to their feelings about more well-known and visible related authorities, which has been termed the "vertical legitimacy spillover effect" (Haack et al., 2014). Prior research has, for example, shown that the perceived legitimacy of the CJEU is strongly related to the perceived legitimacy of national legal systems (Voeten, 2013) and of the EU in general (Caldeira and Gibson, 1995; Voeten, 2013; Pollack, 2018). People may thus partly base their legitimacy judgments about the CJEU on value alignment with the EU.

In sum, what matters is that people experience a sense of shared values. This in turn depends on which values people *personally* endorse. According to moral foundations theory (MFT; Haidt, 2007), the range of human moral values can be classified into two main categories. On the one hand, there are "individualizing" moral foundations, which are focused on protecting the *individual*. These values are "care" and "fairness." On the other hand, there are "binding" moral foundations, such as "ingroup loyalty", "respect for authority", and "purity", which are focused on protecting the *group*. While some people are predominantly drawn to individualizing foundations, others are equally or even more drawn to binding foundations (Haidt, 2007). These dispositions in turn have shown to underlie political values and opinions, as political liberals typically only rely on individualizing moral foundations, while political conservatives endorse all five foundations equally (Haidt and Graham, 2007; Graham et al., 2009). For example, the Brexit campaign showed

to appeal to all of the public's moral foundations, which may have influenced votes to leave the EU (Smith, 2019).

In western societies, there is a narrow focus on individualizing moral foundations (Haidt, 2007). Values that resonate strongly with these foundations, such as freedom, equality, and respect for human rights, form the very foundation of the EU and EU law (Article 2 of the Treaty on the European Union). Yet, in these same societies, a large number of people also endorse binding moral foundations (Haidt, 2007). Considering that value alignment constitutes an important element of perceived legitimacy (Jackson et al., 2012) and that individuals have different moral intuitions, it is necessary to look beyond individualizing values when trying to understand perceived legitimacy. For example, people who appreciate tradition and loyalty to their nation may not perceive their values to be particularly represented in EU law, which claims supremacy over even the national constitution.

The present study therefore examines how personal values and perceived values of the EU, as well as alignment between personal and perceived EU values, are related to the perceived legitimacy of the CJEU and EU. These associations are tested for a range of values, which differ in their individualizing (i.e., "democracy", "liberty", and "fairness") versus binding orientation (i.e., "rule of law", "respect for national authority", and "respect for tradition").

MATERIALS AND METHODS

Procedure

We collected data among 1,180 individuals from Finland, France, Germany, Italy, Netherlands, and Poland via the online participant platform Prolific. This data collection comprised multiple measures, of which different parts may appear in future publications. After providing informed consent, participants filled in an online questionnaire, which was designed with Qualtrics software and took approximately 10 min. After that, participants were debriefed and reimbursed with £1. The data and material can be accessed at https://osf.io/6hcw4/?view_only=dfd482abc82548d2afcfd08ae5aaef07. Data collection was ethically approved by the Psychology Research Ethics Committee of the Faculty of Social Sciences at Leiden University (2020-10-22-D.T. Scheepers-V1-2710).

Participants

After removing data of participants who failed to correctly answer both of two attention checks ($n = 36$), who finished the study in less than 5 minutes ($n = 6$), and whose data were missing ($n = 2$), the total sample consisted of 1,136 participants ($n_{Finland} = 164$, $n_{France} = 197$, $n_{Germany} = 195$, $n_{Italy} = 200$, $n_{Netherlands} = 196$, $n_{Poland} = 184$). The mean age was 27.60 ($SD = 8.98$). Of the participants, 456 identified as female (40.1%), 663 as male (58.4%), 15 as other (1.3%), and 2 did not indicate their gender (0.2%). The sample was on average highly educated, with 29.8, 26.1, and 5.5% having earned, respectively, a university's bachelor, master, and doctoral degree. Only 3.3% received primary education, whereas 19.5% received secondary education and 15.8%

vocational or professional education. The sample was leaning toward the left on the political spectrum ($M = 28.11$, $SD = 24.46$, on a 100-point scale ranging from left/progressive to right/conservative). Most participants were either not very aware (32.9%) or somewhat aware (46.6%) of the CJEU, and only few participants had never heard of it (9.9%) or were very aware of it (10.7%).

Materials
Political Ideology

Political ideology was measured to test whether the values we included in the study followed the individualizing-binding pattern predicted by MFT. We asked participants: "In political matters people talk of 'the left' and 'the right'. How would you place your views on this scale with regard to the economic and social dimension?." Only responses on the social dimension mattered for the current study. Participants could indicate their social political orientation on a scale from 1 ("Left/progressive") to 100 ("Right/conservative").

Awareness CJEU

Awareness of the CJEU was measured with one item: "The Court of Justice of the European Union sits in Luxembourg and is the highest court of the European Union as a whole. How aware would you say you are of this court?," with four answer options: *never heard of it before now*, *not very aware*, *somewhat aware*, and *very aware*.

Values and Value Alignment

The perceived values of the EU were measured with the question: "To what extent do you consider each of these values to be endorsed by the European Union?," for each of the values: "Democracy", "Liberty", "Fairness", "Rule of law (e.g., respect for independence of the judiciary, the integrity and impartiality of the electoral system)", "Respect for national authority", and "Respect for tradition." Answers were provided on 5-point Likert scales ranging from *not at all endorsed* to *extremely endorsed*. After measuring EU values, we also measured personal values by asking participants to indicate for the same values: "How important are each of these values to yourself?," using 5-point Likert scales ranging from *not at all important* to *extremely important*. Value alignment was operationalized as overlap between personal values and perceived EU values.

Perceived Legitimacy

The perceived legitimacy of the CJEU and the EU was operationalized as institutional trust and felt duty to obey. We adapted items from previous work that measured the perceived legitimacy of the police (Sunshine and Tyler, 2003; Jackson et al., 2012), resulting in a total scale of nine items ($\alpha_{CJEU} = 0.95$, $\alpha_{EU} = 0.93$; e.g., "I have confidence in the CJEU [EU]," "People should obey decisions from the CJEU [laws made by the EU] even if they will not be caught for breaking them"), which were answered on 7-point Likert scales ranging from *strongly disagree* to *strongly agree*.

Data Analysis

The data were analyzed in RStudio (Version 1.3.959). After descriptive analyses, we performed a maximum-likelihood Confirmatory Factor Analysis (CFA) for the set of personal values and the set of perceived EU values, where individualizing values and binding values were specified as two separate factors. Then, two times two ordinary least squares (OLS) regression models with fixed effects were run for each of the value factors (i.e., individualizing values and binding values) on each of the outcome variables (i.e., perceived legitimacy of the CJEU and perceived legitimacy of the EU). The models controlled for country, age, and education. Predictor variables were personal endorsement of [individualizing-binding] values, perceived endorsement by the EU of [individualizing-binding] values, and the interaction term between [individualizing-binding] values, which was used as an indicator of value alignment. All continuous variables were mean centered.

RESULTS

Figure 1 presents the distribution of personal values and perceived EU values for each value separately and in each of the country subsamples. With regard to both the personal and perceived EU values democracy, liberty, fairness, and rule of law, the data were left-skewed, indicating that most participants highly supported these values and also perceived

the EU to support these values. Respect for national authority and respect for tradition were more evenly distributed among all samples, both regarding personal endorsement and perceived endorsement by the EU. Means, standard deviations, and skewness scores are reported in **Supplementary Table S1**.

Figure 2 shows the relation between each of the personal values and political ideology. Pearson correlations showed that more right-oriented political ideology was related to lower scores on individualizing values ($r = -0.23$, $p < 0.001$) and higher scores on binding values ($r = 0.33$, $p < 0.001$). An OLS regression model in which personal values were regressed on political ideology, controlling for demographic variables (age, gender, and education), confirmed that political ideology was predicted by personal values. More specifically, a more right-oriented political ideology was positively predicted by respect for national authority ($b = 4.71$, $SE = 0.74$, $p < 0.001$) and respect for tradition ($b = 4.46$, $SE = 0.62$, $p < 0.001$), and negatively predicted by democracy ($b = -4.98$, $SE = 0.87$, $p < 0.001$), liberty ($b = -2.74$, $SE = 0.97$, $p = 0.005$), and fairness ($b = -2.92$, $SE = 1.02$, $p = 0.004$). Personal endorsement of the rule of law did not predict political ideology ($b = 0.31$, $SE = 0.83$, $p = 0.705$).

Both the CFA on personal values and the CFA on perceived EU values revealed a good fit for the two specified factors, where democracy, liberty, and fairness loaded on "individualizing values" and rule of law, respect for national authority, and respect for tradition loaded on "binding

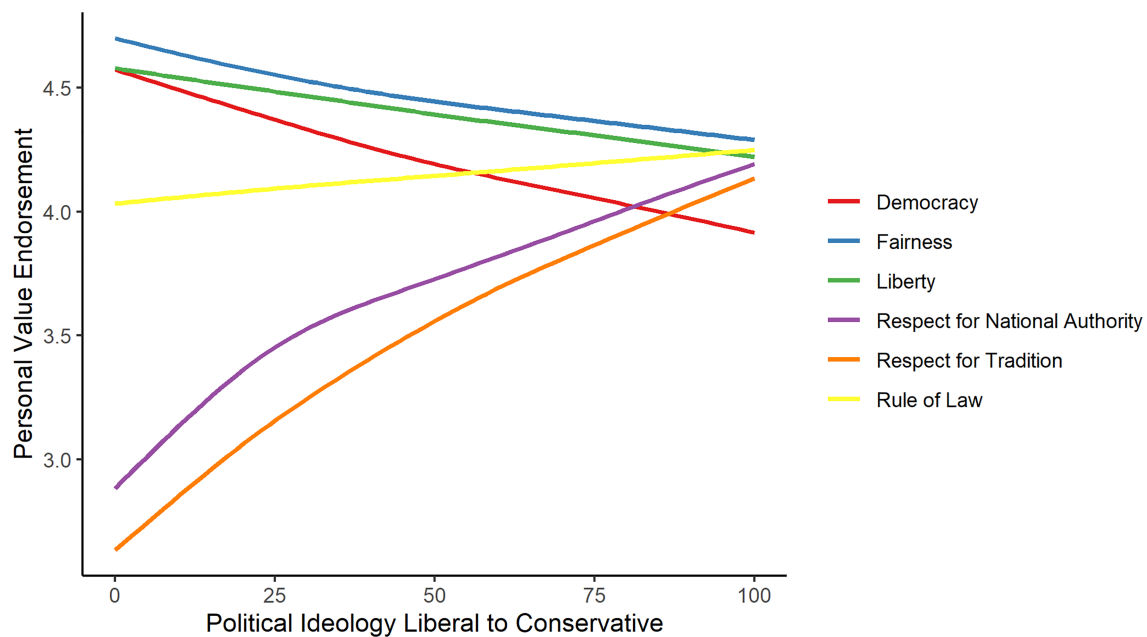**FIGURE 1 |** The distribution of personal values and perceived values of the EU in all country subsamples.
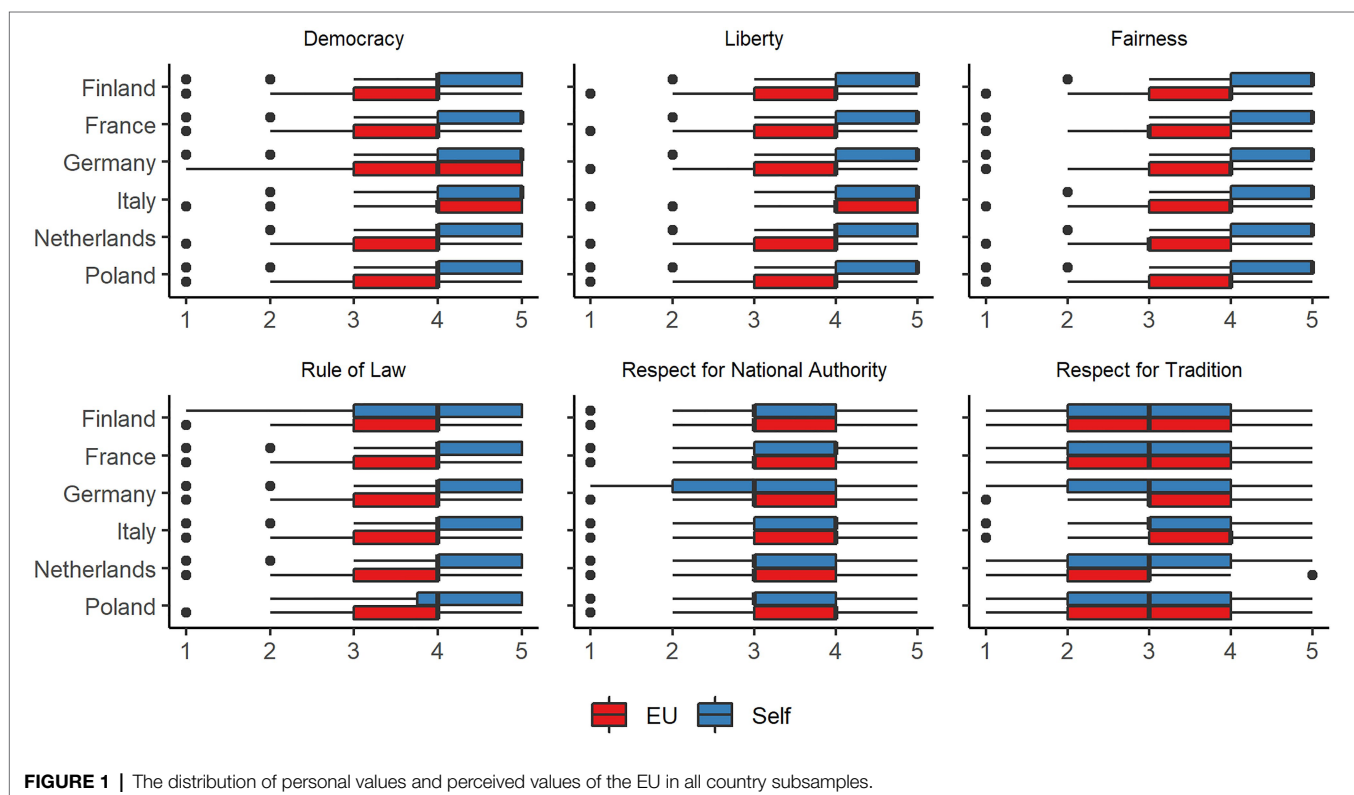
FIGURE 2 | The relation between personal values and political ideology.

TABLE 1 | Descriptive statistics and Pearson correlations between study variables.

	1	2	3	4	5	6	7	8	9	10
1. Age										
2. Education	0.20***									
3. Political ideology	0.11***	−0.09**								
4. Awareness CJEU	0.12***	0.15***	−0.03							
5. Pers. ind. Values	0.10***	0.16***	−0.23***	0.13***						
6. Pers. bind. Values	0.09**	0.02	0.33***	0.12***	0.19***					
7. Perc. ind. Values EU	−0.07*	0.09**	−0.22***	0.12***	0.30***	0.11***				
8. Perc. bind. Values EU	−0.07*	0.03	−0.21***	0.09**	0.21***	0.15***	0.67***			
9. Perc. legitimacy CJEU	0.12***	0.24***	−0.18***	0.23***	0.22***	0.12***	0.51***	0.46***		
10. Perc. legitimacy EU	<0.01	0.17***	−0.24***	0.13***	0.18***	0.09**	0.61***	0.56***	0.80***	
M	27.60	–	28.11	2.58	4.46	3.55	3.65	3.36	5.02	4.62
SD	8.98	–	24.46	0.81	0.55	0.82	0.87	0.85	1.15	1.15

Pers. = personal; perc. = perceived; ind. = individualizing; and bind. = binding. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

values" (see **Supplementary Tables S2 and S3**).[1] The analyses were therefore continued with the factors.

Table 1 shows descriptive statistics and Pearson correlations between the variables. The average perceived legitimacy of the CJEU and EU were both relatively high, with the former being still somewhat higher than the latter, $\Delta_{means} = 0.39$, 95% CI [0.30, 0.49], $t(2270) = 8.18$, $p < 0.001$. Legitimacy perceptions of both institutions were highly correlated. Moreover, for both

---

[1]As also shown in **Supplementary Tables S3 and S4**, the rule of law loaded on both factors, and regarding perceived EU values, the loading on individualizing values was even higher than on binding values. This is not surprising, since the rule of law could be reasonably considered a value which not only relates to binding values, but also to individualizing values—breaches of the rule of law have even become illustrative for conservative governments in Poland and Hungary, which may enhance the significance of this value for people with individualizing values who are often more liberal.

individualizing and binding values, the CJEU and EU were perceived as more legitimate when participants personally endorsed these values and when they perceived the EU to endorse them.

The regression model with individualizing values on perceived legitimacy of the CJEU showed that perceived individualizing values of the EU were a significant positive predictor of perceived CJEU legitimacy, $b = 0.49$, $SE = 0.03$, 95% CI [0.43, 0.54], $p < 0.001$, indicating that the more participants perceived the EU to endorse individualizing values, the higher the perceived legitimacy of the CJEU. The regression model with binding values showed a similar effect of perceived binding values of the EU on perceived legitimacy of the CJEU, $b = 0.44$, $SE = 0.03$, 95% CI [0.38, 0.49], $p < 0.001$. Perceived legitimacy of the CJEU was not predicted by personal individualizing-binding values nor

by the interaction between personal values and perceived EU values (see **Supplementary Tables S4** and **S5** for the results of these models).
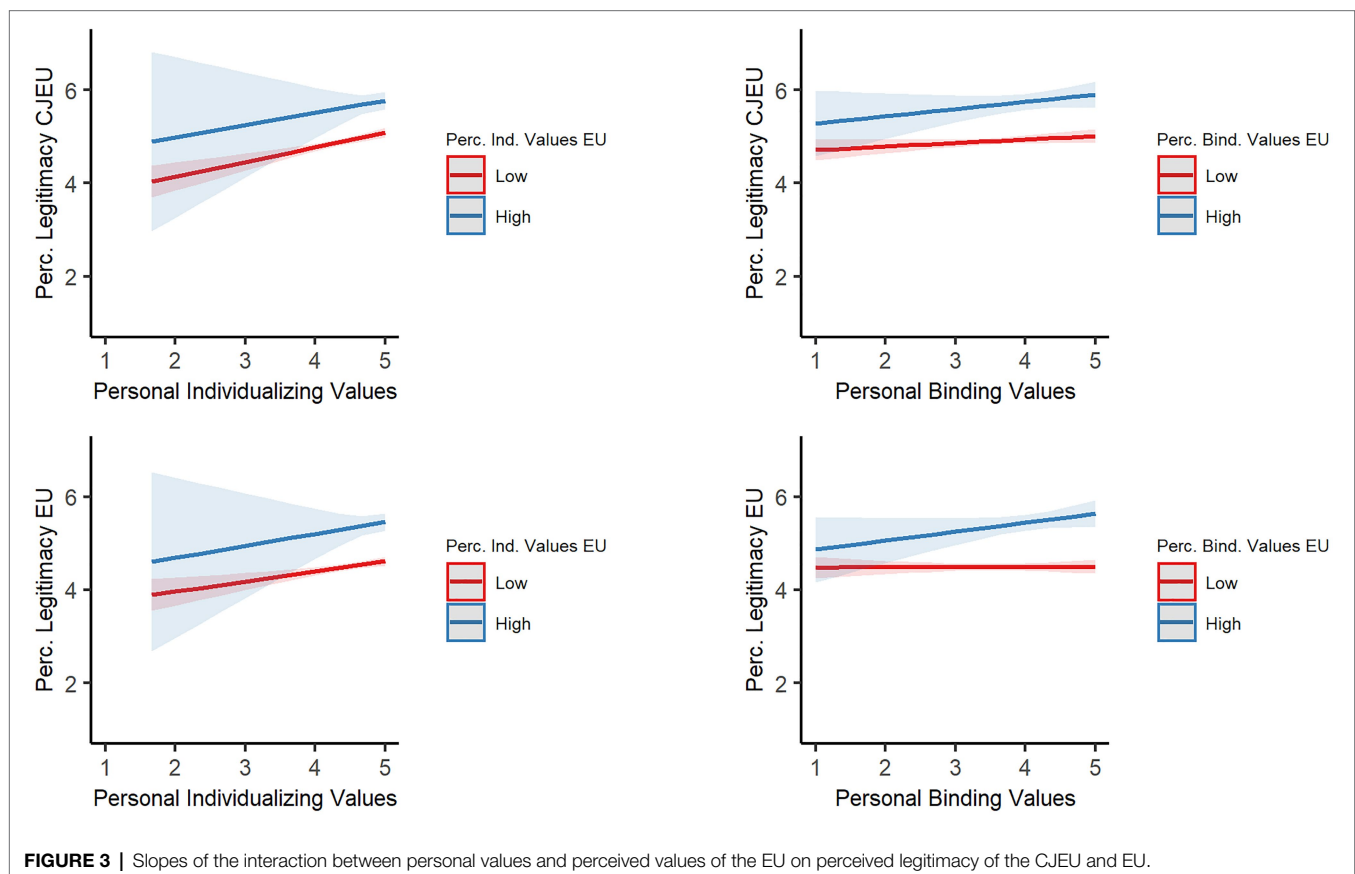
The model with individualizing values on perceived legitimacy of the EU showed that perceived individualizing values of the EU were a significant positive predictor, $b = 0.60$, $SE = 0.03$, 95% CI [0.55, 0.65], $p < 0.001$. This suggests that perceived legitimacy of the EU is higher when people perceive the EU to endorse individualizing values. The model with binding values also showed a significantly positive effect of perceived binding values of the EU on perceived EU legitimacy, $b = 0.52$, $SE = 0.03$, 95% CI [0.47, 0.58], $p < 0.001$. This model in addition revealed a positive, significant interaction between personal binding values and perceived binding values of the EU, $b = 0.07$, $SE = 0.02$, 95% CI [0.02, 0.11], $p = 0.002$ (see **Supplementary Tables S6** and **S7** for the results of these models). Simple slope analyses were conducted to better understand this interaction. These showed that personal binding values had a significantly negative effect on perceived EU legitimacy when participants perceived that the EU weakly represents binding values [1 SD below mean; $\beta = -0.06$, $SE = 0.03$, 95% CI (−0.12, 0), $p = 0.070$], but that this effect was positive when participants perceived that the EU strongly represents binding values [$\beta = 0.08$, $SE = 0.04$, 95% CI (0.01, 0.15), $p = 0.030$].

In sum, as illustrated in **Figure 3**, the CJEU and EU were perceived as more legitimate when the EU's endorsement of individualizing and binding values was high (1 SD above mean)

versus low (1 SD below mean). The bottom-right panel of **Figure 3** shows that the positive effect of EU binding values on perceived EU legitimacy was qualified by an interaction with personal binding values. This interaction entails that personal binding values were unrelated to perceived EU legitimacy when the EU was perceived to weakly endorse binding values (red line); however, personal binding values predicted perceived EU legitimacy when the EU was perceived to strongly support them (blue line).

DISCUSSION

How do people come to perceive the CJEU and EU as (il) legitimate? Understanding these processes is important for the effectiveness and viability of these legal institutions, especially now that the rule of law crisis in *inter alia* Poland and Hungary openly challenges the CJEU's power to uphold EU law in the face of national opposition. Prior research has proposed that alignment between the values of a legal authority and its audience is a key source of perceived legitimacy, but this work has predominantly focused on national legal authorities and procedural justice (Jackson et al., 2012, 2015; Tyler and Jackson, 2013). In the present study, we extended these findings to the supranational level and explored which other values are relevant to the perceived legitimacy of the CJEU and EU. As the EU is a diverse society, comprising multiple countries and cultures,

FIGURE 3 | Slopes of the interaction between personal values and perceived values of the EU on perceived legitimacy of the CJEU and EU.

we investigated how perceived legitimacy of the CJEU and EU is related to both individualizing (i.e., democracy, liberty, and fairness) and binding values (i.e., rule of law, respect for national authority, and respect for tradition).

The findings showed that legitimacy perceptions of the CJEU and EU are, after controlling for demographic variables, higher when people perceive the EU to support both individualizing and binding values. Except for one model, we found no interactions between personal values and perceived EU values. However, although we conceptualized value alignment as the interaction between personal values and perceived EU values, we would not argue that these results should be interpreted as meaning that value alignment is not a source of perceived legitimacy. After all, there was little variation on personal individualizing values scores, showing that all participants strongly supported democracy, liberty, and fairness (i.e., a ceiling effect). Considering the positive effect of perceived endorsement of these values by the EU on perceived legitimacy of the CJEU and EU, it could be argued that value alignment with regard to these values is a source of legitimacy, but that people with individualizing moral foundations find their values already represented by the EU.

For the model where we did find an interaction, i.e., the model where binding values were regressed on perceived legitimacy of the EU, scores on personal values were more evenly distributed across participants. Here, the results revealed that legitimacy decreased as personal support for binding values increased and that this effect was neutralized and actually reversed into a positive effect when participants believed that the EU also supports binding values. These findings are indicative of a value alignment effect for binding values, implying that when the EU fails to serve people with binding moral foundations, the EU may be perceived as less legitimate by these people. Practically, these findings may imply that when the EU better represents binding values, in addition to individualizing values, there will be a positive effect on perceived legitimacy of the EU.

Furthermore, the findings indicated that the CJEU is still not widely known among the public. Awareness increased with higher education but was unrelated to political ideology. As suggested by the high correlation between perceived legitimacy of the CJEU and EU, people's conferral of legitimacy to the CJEU is partly derived from their feelings toward the EU. This is consistent with the vertical legitimacy spill-over effect, which holds that people use affect heuristics to judge the legitimacy of a transnational authority (Haack et al., 2014). Although we cannot say with certainty that perceived values of the EU causally spill over to the perceived legitimacy of the CJEU, the findings provide correlational evidence to suggest that this effect also applies to international courts.

The findings should be interpreted while noting the study's limitations. First, our sample included people from only six member states, who may have represented a certain social class as they were required to be fluent in English, limiting generalizability of the findings. In addition, due to the study's cross-sectional nature, the results cannot give insight into causality. Furthermore, we based our expectations about the differences between individualizing and binding values on MFT but did not measure the "traditional" moral foundations. Although the values that were included in our study were closely related to the values of MFT and followed a similar pattern with regard to political ideology, it would nevertheless be interesting to see whether the results remain when using the traditional MFT items. The value of freedom, moreover, may represent a sixth moral foundation, which has been identified as not belonging to the individualizing or binding moral foundations: liberty, which is characterized by strong endorsement of individual liberty and resentment of any sign of domination or repression (Iyer et al., 2012). Follow-up studies could look into when and how freedom/liberty is relevant for perceived legitimacy of the EU. For example, by studying how political parties' framing of freedom affects perceived legitimacy, as it could be framed as "the freedom of minority groups to make individual choices without oppression from majority elites" but also as "the freedom to decide for 'ourselves' without interference from 'Brussels.'"

Finally, no scholarly consensus exists about the meaning of perceived legitimacy and the best way to measure it. Here, we operationalized perceived legitimacy as institutional trust and felt duty to obey (Sunshine and Tyler, 2003). However, legitimacy, trust, and duty to obey may overlap and differ, and can differently affect law-related behavior (Jackson and Gau, 2016). It is therefore important that future research finds novel ways to measure perceived legitimacy, for example, with behavioral measures, which would also improve our conceptual understanding of perceived legitimacy (cf., Dellmuth and Schlipphak, 2020).

As for other future directions, future research should test the causal directions between value alignment, identification with the EU, and perceived legitimacy of the EU in more controlled lab experiments. Another direction is to examine whether it matters how *effective* the EU is in the eyes of the public at safeguarding their values, since effectiveness in achieving policy objectives has been defined as an institutional source of legitimacy (Dellmuth et al., 2019). Finally, it would be interesting to take into account and better understand the different discourses that may lead to perceived (il)legitimacy of the CJEU and EU, since Euroscepticism can be rooted in different concerns and narratives (Baldassari et al., 2020).

Notwithstanding these limitations, our study sheds light on social psychological processes that lead to perceived legitimacy of the CJEU and EU, highlighting the role of value alignment. Although individualizing values are important to protect, our findings suggest that improving these values may not result in a net increase of perceived legitimacy. Instead, they suggest that some citizens find it equally important that the EU respects binding values, such as respect for tradition and national authority, and that these values are currently not perceived by the public as sufficiently safeguarded by the EU. Better serving people with binding values could therefore be a strategy to improve perceived legitimacy of the CJEU and EU. Of course, this perspective brings new legal and political difficulties, for it is harder for authorities to represent everyone's values in multicultural societies (cf.,

Tyler and Jackson, 2013). However, this only underlines the demand for a better understanding of the potential and pitfalls of value alignment for perceived legitimacy of the CJEU and EU.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://osf.io/6hcw4/?view_only=dfd482abc82548d2afcfd08ae5aaef07.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Psychology Research Ethics Committee, Leiden University (2020-10-22-D.T. Scheepers-V1-2710). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors discussed the objectives of the data generated for this study, which is part of a larger research project, and the structure of the paper. EG analyzed the data and wrote the manuscript. DS and AC reviewed and edited the manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.785892/full#supplementary-material

REFERENCES

Baldassari, M., Castelli, E., Truffelli, M., and Vezzani, G. (2020). *Anti-Europeanism: Critical Perspectives towards the European Union*. Cham: Springer International Publishing AG.

Bottoms, A., and Tankebe, J. (2012). Beyond procedural justice: A dialogic approach to legitimacy in criminal justice. *J. Crim. Law Criminol.* 102, 119–170.

Caldeira, G. A., and Gibson, J. L. (1995). The legitimacy of the court of justice in the European Union: models of institutional support. *Am. Polit. Sci. Rev.* 89, 356–376. doi: 10.2307/2082430

Dellmuth, L. M., and Schlipphak, B. (2020). Legitimacy beliefs towards global governance institutions: a research agenda. *J. Eur. Public Policy* 27, 931–943. doi: 10.1080/13501763.2019.1604788

Dellmuth, L. M., Scholte, J. A., and Tallberg, J. (2019). Institutional sources of legitimacy for international organisations: Beyond procedure versus performance. *Rev. Int. Stud.* 45, 627–646. doi: 10.1017/S026021051900007X

Graham, J., Haidt, J., and Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *J. Pers. Soc. Psychol.* 96, 1029–1046. doi: 10.1037/a0015141

Grootelaar, H. A. M., and van den Bos, K. (2018). How litigants in Dutch courtrooms come to trust judges: The role of perceived procedural justice, outcome favorability, and other sociolegal moderators. *Law Soc. Rev.* 52, 234–268. doi: 10.1111/lasr.12315

Haack, P., Pfarrer, M. D., and Scherer, A. G. (2014). Legitimacy-as-feeling: how affect leads to vertical legitimacy spillovers in transnational governance. *J. Manag. Stud.* 51, 634–666. doi: 10.1111/joms.12071

Haidt, J. (2007). The new synthesis in moral psychology. *Science* 316, 998–1002. doi: 10.1126/science.1137651

Haidt, J., and Graham, J. (2007). When morality opposes justice: conservatives have moral intuitions that liberals may not recognize. *Soc. Justice Res* 20, 98–116. doi: 10.1007/s11211-007-0034-z

Hofmann, A. (2018). Resistance against the court of justice of the European Union. *Int. J. Law Context* 14, 258–274. doi: 10.1017/S174455231800006X

Iyer, R., Koleva, S., Graham, J., Ditto, P., and Haidt, J. (2012). Understanding libertarian morality: The psychological dispositions of self-identified libertarians. *PLoS One* 7:e42366. doi: 10.1371/journal.pone.0042366

Jackson, J., Bradford, B., Hough, M., Myhill, A., Quinton, P., and Tyler, T. R. (2012). Why do people comply with the law? *Br. J. Criminol.* 52, 1051–1071. doi: 10.1093/bjc/azs032

Jackson, J., and Gau, J. M. (2016). "Carving up concepts? Differentiating between trust and legitimacy in public attitudes towards legal authority,"

in *Interdisciplinary Perspectives on Trust: Towards Theoretical and Methodological Integration*. eds. E. Shockley, T. M. S. Neal, L. M. PytlikZillig and B. H. Bornstein (Cham: Springer International Publishing), 49–69.

Jackson, J., Hough, M., Bradford, B., and Kuha, J. (2015). "Empirical legitimacy as two connected psychological states," in *Trust and Legitimacy in Criminal Justice: European Perspectives*. eds. G. Meško and J. Tankebe (Cham: Springer International Publishing), 137–160.

Pollack, M. A. (2018). "The legitimacy of the court of justice of the European Union," in *Legitimacy and International Courts*. eds. N. Grossman, H. G. Cohen, A. Follesdal and G. Ulfstein (Cambridge: Cambridge University Press).

Smith, D. S. (2019). Shaping the modern world with a stone-age brain: The brexit referendum and the moral foundations theory. *J. Soc. Polit. Psychol.* 7, 863–889. doi: 10.5964/jspp.v7i2.1032

Suchman, M. C. (1995). Managing legitimacy: strategic and institutional approaches. *Acad. Manag. Rev.* 20, 571–610. doi: 10.2307/258788

Sunshine, J., and Tyler, T. R. (2003). The role of procedural justice and legitimacy in shaping public support for policing. *Law Soc. Rev.* 37, 513–548. doi: 10.1111/1540-5893.3703002

Trinkner, R., and Tyler, T. R. (2016). Legal socialization: coercion versus consent in an era of mistrust. *Annu. Rev. Law Soc. Sci.* 12, 417–439. doi: 10.1146/annurev-lawsocsci-110615-085141

Tyler, T. R. (1997). The psychology of legitimacy: A relational perspective on voluntary deference to authorities. *Personal. Soc. Psychol. Rev.* 1, 323–345. doi: 10.1207/s15327957pspr0104_4

Tyler, T. R. (2006). Psychological perspectives on legitimacy and legitimation. *Annu. Rev. Psychol.* 57, 375–400. doi: 10.1146/annurev.psych.57.102904.190038

Tyler, T. R., and Blader, S. L. (2003). The group engagement model: procedural justice, social identity, and cooperative behavior. *Personal. Soc. Psychol. Rev.* 7, 349–361. doi: 10.1207/S15327957PSPR0704_07

Tyler, T. R., Goff, P. A., and MacCoun, R. J. (2015). The impact of psychological science on policing in the United States: procedural justice, legitimacy, and effective law enforcement. *Psychol. Sci. Public Interes.* 16, 75–109. doi: 10.1177/1529100615617791

Tyler, T. R., and Jackson, J. (2013). "Future challenges in the study of legitimacy and criminal justice," in *Legitimacy and Criminal Justice: An International Exploration*. eds. J. Tankebe and A. Liebling (Oxford: University Press), 83–104.

Tyler, T. R., and Lind, E. A. (1992). A relational model of authority in groups. *Adv. Exp. Soc. Psychol.* 25, 115–191. doi: 10.1016/S0065-2601(08)60283-X

van den Bos, K., van der Velden, L., and Lind, E. A. (2014). On the role of perceived procedural justice in citizens' reactions to government decisions and the handling of conflicts. *Utr. Law Rev.* 10, 1–26. doi: 10.18352/ulr.287

Voeten, E. (2013). Public opinion and the legitimacy of international courts. *Theor. Inq. Law* 14, 411–436. doi: 10.1515/til-2013-021

Walters, G. D., and Bolger, P. C. (2019). Procedural justice perceptions, legitimacy beliefs, and compliance with the law: a meta-analysis. *J. Exp. Criminol.* 15, 341–372. doi: 10.1007/s11292-018-9338-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Psychometric Challenges in the Measurement of Constructs Underlying Criminal Responsibility in Children and Young Adults: A Cross-Sectional Study

Yuxi Shang[1,3], Yumiao Fu[2], Beibei Ma[1], Li Wang[3]* and Dexin Wang[1]*

[1] School of Law, Shandong Normal University, Jinan, China, [2] Master of Law Education Centre, East China University of Political Science and Law, Shanghai, China, [3] School of Psychology, Shandong Normal University, Jinan, China

At present, many countries have lowered the minimum age of criminal responsibility to deal with the trend of juvenile crime. In practical terms, whether countries advocate for lowering the age of criminal responsibility along with early puberty, or regulating the minimum age of juvenile criminal responsibility through their policies, their deep-rooted hypothesis is that age is tied to adolescents' psychological growth, and, with the rise in age, the capacity for dialectical thinking, self-control, and empathy gradually improves. With this study, we aimed to test whether this hypothesis is valid. The participants were 3,208 students from junior high school, senior high school, and freshman in the S province of the People's Republic of China (PRC). We subjected the gathered materials to independent-samples t-tests, one-way analysis of variance (ANOVA), linear regression analysis, and Bonferroni *post hoc* test. The influence of the age variable upon dialectical thinking, self-control, and empathy was significant ($p = 0.002$, $p = 0.000$, $p = 0.072$), but only empathy was positively correlated with age variable ($B = 0.032$); dialectical thinking ability ($B = -0.057$), and self-control ability ($B = -0.212$) were negatively correlated with the age variable. Bonferroni *post hoc* test confirmed these findings. Therefore, we concluded the following: (1) Juvenile criminal responsibility, based on the capacity for dialectical thinking, self-control, and empathy, is not positively correlated with age. (2) Age is not the only basis on which to judge a juvenile's criminal responsibility. (3) More research that directly links age differences in brain structure and function to age differences in legally relevant capacities and capabilities(e.g., dialectical thinking, self-control, and empathy) is needed. (4) Political countries should appropriately raise the minimum age of criminal responsibility and adopt the doli incapax principle in the judicial process.

Keywords: juvenile, minimum age of criminal responsibility, dialectical thinking, self-control, empathy

INTRODUCTION

On October 24, 2019, a 13-year-old boy in Dalian in the People's Republic of China (PRC) killed a 10-year-old girl and dumped her body in his home in a brutal manner (Xue et al., 2019). Similarly, on the other side of the world, in November of 2015, news that an 8-year-old child had been held in custody for viciously attacking and killing a 1-year-old child in the United States state of Alabama circulated in the media (Crofts, 2016). With the rapid spread of media reports and communications, an increasing number of malignant incidents committed by young minors have come into the public view.

Faced with the exposure of many younger malignant criminal cases, some countries and regions have chosen to implement strict laws for juvenile offenders; that is, to lower the age of criminal responsibility, so as to try to achieve the goal of social defense by cracking down on juvenile delinquency. Like America in the 1980s, with rising juvenile crime rates and media attention (Butts and Mitchell, 2000; Cook and Laub, 2002), some nations have lowered the age threshold for sending a teenager into the criminal justice system (some have lowered it to 12 years old), ushering in a "Hard Age" of juvenile justice in United States (Fowler and Kurlychek, 2018). Recently, facing the rising trend of juvenile delinquency, the South Korean government is trying to lower the minimum age of criminal responsibility from 14 to 13 (Hong, 2020). On December 22, 2020, at the 24th session of the 13th Standing Committee of the National People's Congress of the PRC, an amendment (XI) was made to the criminal law to lower the minimum age of criminal responsibility for the murder and aggravated assault that causes death from 14 to 12 years old (NPC, 2020).

Is there a scientific basis for lowering the minimum age of criminal responsibility in adolescents? In addition to curbing the rise of juvenile crime rates and eliminating the public's fear and risk of victimization by juvenile delinquents (Blumstein, 1995; Baker et al., 2016; Sickmund et al., 2017), proponents point to the "early onset of [the] concept of right and wrong and understanding of the meaning of one's behavior as socioeconomic development and online life spread" (Rosenfield et al., 2000; O'Brien and Fitz-Gibbon, 2017). However, there are many objections, such as exposing young people to the criminal justice system at an early age does not produce a good preventive effect, but instead leads to an increase in the rate of juvenile recidivism (Doreleijers and Fokkens, 2010; Casey, 2014; O'Brien and Fitz-Gibbon, 2017). Unfortunately, these arguments rest on theoretical discussions or indirect proof (Newton and Bussey, 2012). It is well known that the basis of criminal responsibility in juveniles is the capacity for criminal responsibility; that is, appreciation and self-control (Zhang, 1994; Snyman, 2008; Elliott, 2011). Appreciation is the actor's capacity to distinguish the meaning, nature, function, and consequence of their behavior in criminal law (Snyman, 2008; Goldson, 2013). Self-control is the ability to moderate one's actions and to act in accordance with the law (Goldson, 2013). Whether countries advocate for lowering the age of criminal responsibility along with early puberty or regulating the minimum age of juvenile

criminal responsibility through their policies, one deep-rooted hypothesis is:

> H_1: Age is related to adolescents' psychological growth, and with increasing age, the capacity for appreciation and self-control gradually improves.

Our main objective was to test the validity of such a hypothesis using quantitative methods. Combining the basis of juvenile criminal responsibility (appreciation and self-control), we quantified juvenile criminal responsibility using three psychological indices:

(1) *The index of dialectical thinking.* Dialectical thinking skills enable adolescents to see the world objectively, observe events, and deal with problems in all aspects (Inhelder and Piaget, 1958; Nisbett et al., 2001; Cheng, 2009; Boucher, 2011). Moreover, the development of dialectical thinking ability can effectively reduce people's aggressive behavior (Zhang et al., 2011). In contrast, adolescents with inadequate development of dialectical thinking skills are prone to attribution bias and risky behaviors (Crick and Dodge, 1994).

(2) *The index of self-control.* Self-control is the capacity to suppress inappropriate emotions and behaviors, and to replace them with appropriate ones (Casey, 2014). Low self-control is often the root of problematic behaviors, like poor interpersonal relationships, job prospects, health, and especially of involvement in antisocial and criminal conduct (Walters, 2016). This is consistent with Gottfredson and Hirschi's (1990) assertion that low self-control is a major cause of crime.

Importantly, some countries have set a lower age of criminal responsibility for violent crimes. For example, the Russian Criminal Code (1996) stipulates that a person can be held criminally liable for any offense committed from the age of 16, and a child aged 14 or older can be held criminally liable for a number of serious violent crimes such as willful murder, rape, and robbery (Criminal Code, Article 20). In Ireland, children under 12 can generally not be prosecuted, but children over the age of 10 can be prosecuted for certain crimes such as murder, manslaughter, rape, or aggravated sexual assault (Crofts, 2016).

Psychological research has found that the development of empathy has an impact on juvenile delinquency (Narvey et al., 2021). For example, Miller and Eisenberg's (1988) study found a significant negative correlation between empathy and aggressive behavior, especially in adolescence. Empathy is the cognitive ability to experience and understand the emotions of others (Jolliffe and Farrington, 2004). Empathy is present in the early years of life, and infants have high levels of emotional empathy (Haviland and Lelwica, 1987). Brink et al. (2011) showed increased activation in the medial orbitofrontal cortex, left inferior frontal gyrus, and left dorsolateral prefrontal cortex in a story task that elicited emotional empathy. Low empathy is often associated with aggression and criminality. Research on the relationship between empathy and types of crimes found a significant

correlation between sexual and violent crimes and low empathy (Harpur et al., 1988; Jolliffe and Farrington, 2004, 2006a,b). Likewise, high empathy reduces violence and aggression (Broidy et al., 2003).

(3) Hence, in addition to the dialectical thinking and self-control indices, we also regard the index of *empathy* as a standard with which to measure the level of juvenile criminal responsibility. Our previous argument for H_1 could be modified to H_2:

H_2: Age is related to adolescents' psychological growth; the capacity for dialectical thinking, self-control, and empathy is positively correlated with age, and a developmental (or stable) level of empathy occurs earlier than the level of dialectical thinking and self-control.

THE CURRENT STUDY

Regarding the importance of age in the criminal responsibility system, it is the threshold that determines whether a juvenile will enter the criminal justice system (Crofts, 2016). Different countries have different age levels due to distinct historical traditions and cultures (Pillay, 2019), and there is no consensus on which age level is appropriate. In existing research, the area of the relationship between age and criminal responsibility is understudied. Therefore, we aimed to explore the connection between age and criminal responsibility; more specifically, whether juveniles' criminal responsibility ability is positively correlated with age, and whether they tend to have the capacity for adult criminal responsibility at a certain age. It is important to test the deep-rooted belief that age is the criterion for determining adolescents' criminal responsibility; with the development of society and the maturity of teenagers, the minimum age of criminal responsibility can be adjusted. We used quantitative analysis. First, through questionnaires, we measured adolescents' capacity for dialectical thinking, self-control, and empathy in order to establish a propensity for violent crime. Second, under the control of demographic variables such as academic achievement, parental occupation, and socioeconomic status, we analyzed the relationship between the three indices and adolescents' age. Finally, we attempted to address the following questions:

- Is adolescents' criminal responsibility (the capacity for dialectical thinking, self-control, and empathy) positively correlated with age?

  - If so, will adolescents' capacity for dialectical thinking, self-control, and empathy become more stable (mature) or more adult-like at some point in their lives?
  - If not, what does the developmental trend of young people's capacity for dialectical thinking, self-control, and empathy look like?

- Does the minimum age of the criminal responsibility system need to be reformed? If so, how?

MATERIALS AND METHODS

Data

The first sample consisted of students from grades 6 through 12(ages from 10 to 22) in S County, in S province of the PRC. S County is located in China's eastern coastal region, where young students have access to more advanced educational methods and technologies, but the level of economic growth is in the middle compared to the rest of the country (in 2020, S county's per capita disposable income was [PCDI] = [¥17,046, ¥30,933]; the PRC's per capita disposable income was [PCDI] = [¥15,204, ¥43,834]). To improve the representativeness of the sample, we used a whole group random sampling method to select 2,800 participants out of 27,031 students from urban and rural areas in S County. As we chose the high schools through a unified examination from the junior high schools in urban and rural areas, we did not distinguish between rural and urban schools. We divided the samples into a primary school group, a junior high school group, and a senior high school group. The primary school group only included pre-primary students (grade 6); we randomly selected 200 out of 2,659 pre-primary students from urban primary schools, and 200 out of 2,637 pre-primary students from rural primary schools, totaling 400 students. The junior high school group included students from grades 7 through 9. We randomly selected 200 urban school students and 200 rural secondary school students from each grade, totaling 1,200 students (the total number of students in each grade is 4,032, 3,940, 3,354 respectively). The senior high school group included students from grades 10 through 12. We randomly selected 400 students from each grade, totaling 1,200 students (the total number of students in each grade is 3,696, 3,477, 3,236, respectively). In all, we selected 2,800 samples, each of whom completed three questionnaires. We distributed a total of 8,400 paper-based questionnaires and collected 8,379, with a recovery rate of 99.75%. We found a small number of students aged 18 and over after the initial sample selection. We also conducted a second supplementary selection. From S University in S province, we chose 415 freshmen to fill out a questionnaire survey with a recovery rate of 100%; the final sample size was 3,208(ages from 11 to 19), the total recovery rate of 99.76%.

All procedures involving human participants in this study have been approved by the ethical standards of the Academic Board of Shandong Normal University. Participation was voluntary and anonymous, based on written informed consent and the right to withdraw participation at any time. We also obtained their guardians' consent for minors under age 18. To comply with the requirements of COVID-19 prevention and control, we could not personally enter the campus to hand out and administer the questionnaires, so they were handed out by school teachers who had received professional training. To a certain extent, this can ensure the legality and validity of the experimental data source.

We first preprocessed the data, using SPSS AU to screen out 35 invalid samples, and employing SPSS software to eliminate extreme questionnaire scores in each age group, excluding 29 samples. Due to the small number of samples aged 10 and over the age of 20, we excluded 58 samples from these two age

groups, leaving 3,086 samples and 9,258 valid questionnaires. The preliminary analysis showed that the final sample was 47.4% male and 52.6% female; 70.3% of the respondents came from rural families and 29.7% from urban families. Regarding parents' education level, 77.9% of students had fathers, and 82.8% had mothers, who had graduated from junior high school and below. Meanwhile, 22.1% of students had fathers, and 17.2% of students had mothers, who had completed senior high school or above. This is in line with the education levels of parents of middle school students in the PRC (Ji et al., 2018). The respondents ranged in age from 11 to 19, and the distribution proportion of respondents in different age stages is shown in **Table 1**.

Measures

Brief-Dialectical Self Scale

In 2016, Spencer-Rodgers et al. (2004) developed a self-report questionnaire called the Dialectical Self Scale (DSS). The scale has been translated into many languages. We adopted the brief Chinese version (B-DSS), α = 0.71, with 14 items. The scale has been shown to have good validity in previous studies (Hamamura et al., 2008; Spencer-Rodgers et al., 2008, 2010; Liu et al., 2013). The scale contains a 7-point scoring system from "very different" to "very much agree," and encompasses the three dimensions of conflict tolerance, cognitive change, and behavioral change, thereby reflecting people's dialectical thinking level. The higher the scale score, the higher the dialectical thinking level.

Self-Control Ability of Middle School Students Questionnaire

This questionnaire was developed by Wang and Lu (2004), scholars of the PRC. Adolescents' capacity for self-control is mainly reflected in three dimensions: emotional self-control, behavioral self-control, and thinking self-control. The split-half reliability is 0.856 (Wang and Lu, 2004). The scale has been shown to have good validity in previous studies (Wang and Lu, 2004; Feng et al., 2021; Tan, 2021). The questionnaire has a total of 36 items, including 10 forward-scoring questions and 26 reverse-scoring questions. Each item uses a 5-point scoring system, ranging from "totally disagree"

TABLE 1 | The age distribution of the respondents.

Age (M = 15.1; SD = 2.3)	Proportion (%)
11	5.2
12	11.8
13	12.8
14	12.9
15	12.9
16	12.0
17	11.8
18	15.1
19	5.5

N = 636. M, mean; SD, standard deviation.

to "totally agree." The higher the score, the stronger the self-control.

Basic Empathy Scale

There are many tools for measuring empathy, such as the widely used Interpersonal Response Indicator (IRI) in the PRC, but these scales have been questioned for confusing empathy with sympathy. Hence, for this study, we used the Basic Empathy Scale (BES; Darrick and David, 2006). The BES is divided into two dimensions: emotional and cognitive empathy. The scale contains 20 items, including 8 items for negative scoring and 12 items for positive scoring; the higher the score, the greater the respondent's empathy. Li et al. (2011) tested the structure of theoretical factors and the reliability and validity of the BES in the youth population of the PRC. They found that the BES met the relevant requirements of psychometrics (α = 0.777). The scale has been shown to have good validity in previous studies (Darrick and David, 2006; Li et al., 2011).

Plan of Analysis

We employed SPSS 19.0 to analyze the results. Before doing so, we calculated the scores of the B-DSS, Self-Control Ability of Middle School Students Questionnaire (SAMSSQ), and BES (this score is the average score of each item on the scale). The missing values in the scores are filled in by the mean of the scores in the sample's age group (Jin and Yu, 2015). After that, we employed the independent-samples *t*-tests and one-way ANOVA to gauge the influence of demographic variables on the capacity for dialectical thinking, self-control, and empathy. After controlling for the demographic variables, we observed the relationship between dialectical thinking, self-control, empathy, and age. Second, using two-variable correlation analysis, we explored whether it was necessary to carry out multivariate analysis of variance (MANOVA) on each dimension of the B-DSS, SAMSSQ, and BES. We used linear regression to derive the explanatory power of age for dialectical thinking, self-control, and empathy. Third, we employed the Bonferroni *post hoc* test, and further scrutinized the differences across ages in terms of dialectical thinking, self-control, and empathy.

RESULTS

Covariance Analysis

Table 2 shows the variance analysis of demographic information using the independent samples *t*-tests and one-way ANOVA. We employed independent samples *t*-tests for the two categorical variables (including gender, family location, and class member status) and we used one-way ANOVA for three or more variables (including grade, achievement ranking, father's level of education, mother's level of education, family income). The data in **Table 2** reveal that, in addition to age, other factors affected the B-DSS, SAMSSQ, and BES scores: The differences among the B-DSS, SAMSSQ, and BES scores across different grades were statistically significant. Gender only had an effect on the SAMSSQ and BES scores. Grade, whether the student's family was living in an urban or rural area, and whether the student was part

TABLE 2 | Variance analysis of the demographic variables and the scores for the three kinds of abilities.

Variables	B-DSS	SAMSSQ	BES
	T/F	T/F	T/F
Age	3.381*	66.349**	1.267
Gender	0.567	−2.985**	−11.611**
Achievement ranking	0.701	94.734**	8.843**
Family location	−0.151	5.848**	2.275*
Father's education level	0.635	8.845**	1.446
Mother's education level	0.424	10.832**	1.098
Family income	3.131*	55.858**	2.873
Class member status	0.409	34.407**	−5.100**

Gender coded as (1 = male, 2 = female).
*p < 0.05.
**p < 0.01.

of a class committee had an impact on the SAMSSQ and BES scores. However, parents' education and family economic level only affected the SAMSSQ scores.

Other factors besides age may influence judgments about the relationship between age and dialectical thinking, self-control, and empathy. **Table 3** controls for these relevant demographic variables, revealing the scores of the B-DSS, SAMSSQ, and BES of each age group. Covariance analysis indicated that the difference between age and the B-DSS scores was statistically significant ($F = 2.646$, $p = 0.007$). Likewise, the difference between age and the self-control scores was statistically significant ($F = 28.788$, $p = 0.000$). The difference between age and empathy was not statistically significant ($F = 1.086$, $p = 0.370$).

Linear Regression Analysis

Age and the scores of the three abilities are numerical variables. The normalized residuals of the dependent variables (the three scales' scores) followed a normal distribution, which confirmed that our study met the requirements of the linear regression analysis. We test a linear regression model that explored the effects of age on the capacity for dialectical thinking, self-control, and empathy after controlling for the demographic variables, and mainly tested the interpretation level and direction of age for the three abilities. The outcomes of linear regression showed that age was correlated with the capacity for dialectical thinking ($p = 0.002$; 95%CI = [−0.268, −0.061]) and self-control ($p = 0.000$; 95%CI = [−2.137, −1.540]), which is consistent with the results of the variance analysis of the questionnaire scores and the demographic variables in **Table 2**. The linear regression also indicated a correlation between age and the capacity for empathy ($p = 0.072$; 95%CI = [−0.011, 0.969], although $p > 0.050$, but p was within the range of acceptability), which was different from the outcomes of one-way ANOVA ($p = 0.370$). Besides, the linear regression data showed that the explanatory power and correlation direction of age to the three abilities were different. Age was explained by 0.4% of the variance in dialectical thinking (Nagelkerke's R^2), which pointed to a negative correlation (Beta = −0.057). Age accounted for 12.9% of the variance in self-control (Nagelkerke's R^2), with a

negative correlation (Beta = −0.212). Age accounted for 5.2% of the variance in empathy (Nagelkerke's R^2), with a positive correlation (Beta = 0.032).

After completing the above linear regression analysis, we needed to explain why we did not conduct a multivariate analysis of variance for dialectical thinking, self-control, and empathy, and why we did not analyze the dimensions of the three scales. The two-variable correlation analysis, shown by **Table 4**, suggests that dialectical thinking, self-control, and empathy are correlated, but their Pearson's product-moment correlation coefficients were all less than 0.3, which indicates that they were independent. After scoring the same type of questionnaire on different dimensions, data analysis demonstrated that the Pearson's product-moment correlation coefficients for conflict tolerance, cognitive change, and behavioral change (three dimensions) on the B-DSS were all less than 0.3. The Pearson's product-moment correlation coefficients for emotional self-control, behavioral self-control, and thinking self-control (three dimensions) on the SAMSSQ were all higher than 0.6, and the Pearson's product-moment correlation coefficients for emotional empathy and cognitive empathy (two dimensions) on the BES were lower than 0.3.

Bonferroni *post hoc* Test

The ANOVA showed that age was correlated with the capacity for dialectical thinking, self-control, and empathy. The linear regression analysis explains the degree and direction of the interpretation of age for the three abilities as a whole. The differences in these three abilities in each age group have not been fully revealed. The Bonferroni *post hoc* test (**Supplementary Appendix A**) was able to specifically compare the three abilities at different ages. **Figure 1** is based on the mean scores of

TABLE 3 | Descriptive analysis of age and the scores of the three kinds of abilities.

| Scale	B-DSS		SAMSSQ		BES	
Age	M	SD	M	SD	M	SD
11	4.376	0.038	3.859	0.041	3.653	0.036
12	4.397	0.025	3.747	0.027	3.586	0.024
13	4.355	0.024	3.558	0.025	3.600	0.023
14	4.369	0.024	3.340	0.025	3.613	0.022
15	4.447	0.024	3.249	0.025	3.651	0.022
16	4.357	0.025	3.260	0.026	3.637	0.023
17	4.349	0.025	3.279	0.027	3.667	0.024
18	4.284	0.022	3.396	0.024	3.646	0.021
19	4.323	0.037	3.484	0.039	3.615	0.035

N = 636. M, mean; SD, standard deviation.

TABLE 4 | Two-variable correlation analysis

Variables	Dialectical thinking	Self-control	Empathy
Dialectical thinking	–		
Self-control	−0.170**	–	
Empathy	0.123**	0.037*	–

*p < 0.05.
**p < 0.01.

FIGURE 1 | Trend chart of the scores for the three abilities along with age.
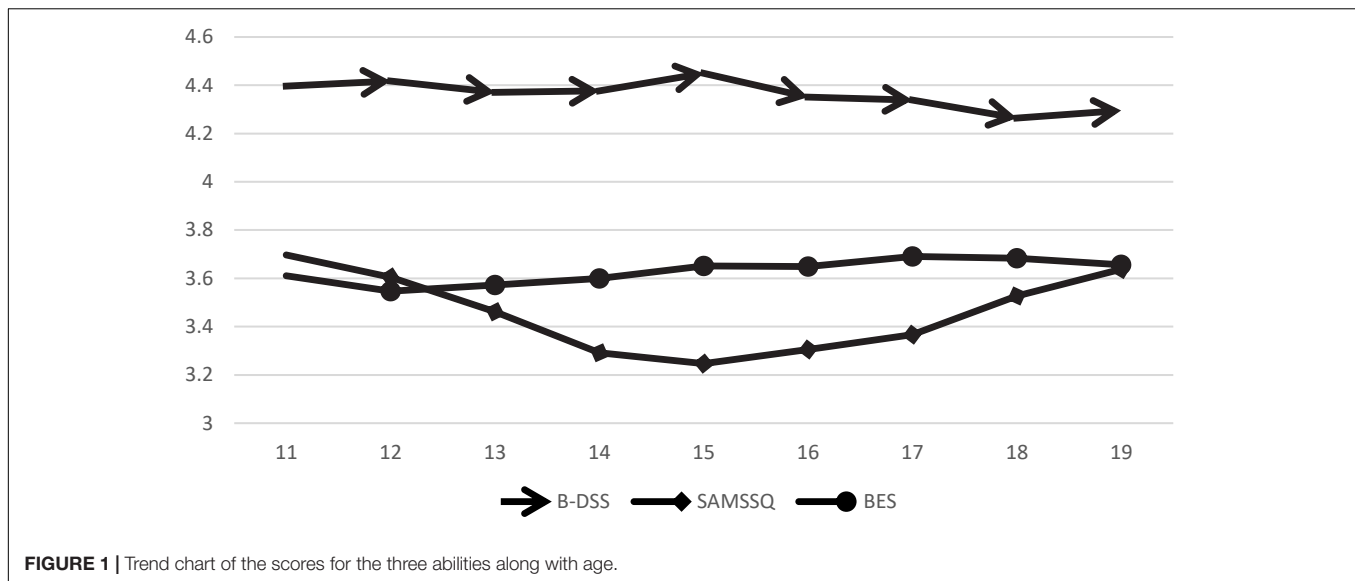
the capacity for dialectical thinking, self-control, and empathy at different ages. **Figure 1** and **Supplementary Appendix A** present the following: (1) The B-DSS scores were highest at age 15 (up: 11–15) and then fluctuated up and down (down: 15–18; up: 18–19). However, the dialectical thinking score of the 18-year-old group was lower than that of the 15-year-old group, and there was a significant difference ($p = 0.000$). (2) The self-control scores showed a more obvious, U-shaped trend with increasing age; the scores of students aged 11–15 decreased, those of 15–19 years old increased, and those of 14–16 were the lowest. The score of 11-year-olds was higher than that of 18- and 19-year-olds ($p = 0.000$), and the score of 12-year-olds was higher than that of 18- and 19-year-olds ($p = 0.000$). (3) The correlation between age and empathy was acceptable ($p = 0.072$). Overall, the BES scores indicate an increasing trend with age (12–18 years). Regarding the rising curve (BES), 16-year-old individuals had a slightly different score (less than 15-year-olds, but still more than 14-year-olds). However, the results of the Bonferroni *post hoc* test showed no significant difference in BES scores between different age groups (**Supplementary Appendix A**).

DISCUSSION

The age at which minors can be punished is controversial in different political countries (O'Brien and Fitz-Gibbon, 2017; Noroozi et al., 2018; Brown and Charles, 2019; Pillay, 2019; Schmidt et al., 2020). We employed a quantitative analysis of research methods, focusing on whether age can be used as a basis for measuring criminal responsibility, while also paying attention to the minimum age of criminal responsibility for violent crimes. The results of correlation tests showed that the influence of the age variable upon dialectical thinking, self-control, and empathy was significant, but only empathy was positively correlated with age variables

(but the results of the Bonferroni *post hoc* test showed no significant difference in BES scores between different age groups). Dialectical thinking ability and self-control ability were negatively correlated with the age variable. This basically disproves the underlying hypothesis that countries should set a minimum age of criminal responsibility for juveniles, and indicates that the capacity for appreciation and self-control is positively correlated with age (H_2). These results will be explained next.

Dialectical Thinking Ability

We found that dialectical thinking does not increase with age; adolescents' dialectical thinking is in a constant state of development until the age of 15, reaches a maximum then, and afterward declines. Next, it shows an upward trend after the age of 18. This finding is consistent with previous research on the current state of dialectical thinking development in middle school students (Lin and Qingan, 2005; Zhang, 2014). From age 11 (12) to age 17 (18) is the period when the mode of thinking transitions from the stage of formal operation to the stage of dialectical thinking (Inhelder and Piaget, 1958; Lin and Qingan, 2005). Generally speaking, with the increase of age, the dialectical thinking ability of minors is gradually increasing, which is not consistent with our conclusion (this trend does not begin until the age of 18). In other words, the development of adolescents' dialectical thinking is not only unbalanced, but also possibly delayed. This may be related to the changing environment in which we live. Currently, adolescents are mired in a changing and fast world, especially with the advancements of smartphones and online games, which makes the thinking of teenagers become more simple and flat, and they gradually lose their interest in deep thinking of things (Ye and Li, 2005; Zheng, 2018).

On the other hand, dialectical thinking arises in the post-formal operations stage of Piaget's cognitive developmental phases (Nisbett et al., 2001), in which individuals are able to

see things and deal with problems in a holistic, connected, and developmental manner; this stage is also the last and highest of Piaget's series of cognitive phases (Inhelder and Piaget, 1958). Lin and Qingan (2005) proposed that the dialectical thinking development of teenagers is the foundation laid by the knowledge learning in middle school. However, because it is the advanced stage of the development of cognition or thinking, the lag of development is inevitable.

In addition, emerging adults are in the process of identity exploration, during which they perceive themselves as neither teenagers nor adults and are unable to take responsibility and make decisions on their own, thus there is a lag in cognitive development in emerging adults (Arnett, 2000; Zheng, 2018; Kang, 2020). Of course, it should be noted that the development of dialectical thinking of adolescents after the age of 18 (early youth) needs to be further verified due to the limited data of subjects after the age of 18.

Self-Control

Self-control ability showed a U-shaped trend, reaching a minimum at approximately 15 years of age and rising again afterward. This is consistent with the findings of Wang and Lu, who created the SAMSSQ (Wang and Lu, 2004). The reason for these outcomes is that adolescents enter puberty at approximately 15 years old, a period of physical, psychological, and hormonal changes (Choudhury et al., 2008). With the increase of age, adolescents become more independent and want to get rid of the restrictions of adults, both dependent and rebellious to adults, and sometimes appear out of control (Wang and Lu, 2004). Emotionally, they sometimes appear unstable, and this imbalance in psychological development makes their self-control no longer as good as before.

Related brain imaging evidence suggests that the maturation and development of relevant tissues in the brain during adolescence do not always increase linearly, but also present a non-linear curve of development (Gogtay et al., 2004; Toga et al., 2006). For example, frontal cortex activity increases between childhood and adolescence, and decreases between adolescence and adulthood (Choudhury et al., 2008).

Further, the influence of the social environment is particularly evident during puberty (Steinberg et al., 2008; Somerville, 2013; Blakemore and Mills, 2014), and adolescents undergoing puberty are more susceptible to peer influences (Guyer et al., 2012). The presence of peers made teens more likely to engage in risky behavior. And teens exhibited relatively greater activation in the ventral striatum and orbitofrontal cortex when their peers were observing them than when they were alone (Chein et al., 2011). Steinberg and Monahan (2007) interpreted these findings to mean that peers elicit a higher motivational state, which then activates the individual's awareness, leading to a decrease in self-control.

Empathy

Regarding the development of empathy, there seems to be an upward trend from visual observation. However, the variability between age and the acquisition of empathy was not very significant (linear analysis: $p = 0.072$, one-way ANOVA:

$p = 0.370$), the results of the Bonferroni *post hoc* test also verified this result. The results are consistent with the dual processing model theory of empathic lifelong development (Huang and Su, 2010; Wang et al., 2021). This model suggests that the developmental trajectory of individual emotional empathy follows a U-shaped curve; its intensity remains relatively stable between adolescence and adulthood and then gradually increases (Liu and Cui, 2020). According to the results of this study, there is a linear trend in the development of adolescent empathy, which depends on a certain neuroscience basis (Wang et al., 2021). The maturation of the empathy response is closely related to the maturation of the prefrontal cortex (PFC), and adolescence is a critical period for individuals to reach the level of prefrontal cortex maturation (Yang et al., 2017). Decety et al. (2008) found that the ventromedial prefrontal cortex, which is strongly associated with cognitive empathy, becomes more active with age from childhood through functional magnetic resonance imaging.

Other researchers have found that the brain regions related to cognitive empathy such as the right temporo-parietal junction area and the left inferior frontal gyrus were significantly activated, while the brain regions related to emotional empathy did not show significant activation when someone else suffered a loss (Schwenck et al., 2017). This reflected the maturation of individual empathy, the stability of emotional empathy, and the development of cognitive empathy (Kunzmann et al., 2018). The development of empathy has an impact on juvenile delinquency (Narvey et al., 2021), which is also the reason why many political countries set a lower age of criminal responsibility for violent crimes. However, this kind of action needs further discussion, because the setting of criminal responsibility for minors is one that requires the simultaneous consideration of dialectical thinking, self-control, empathy.

LIMITATIONS AND FUTURE STUDIES

Although we reached significant conclusions regarding the relationship between age and criminal responsibility, our study also faced some limitations.

First, although supplementary analyses and the control of covariates enhanced the explanatory power, the current study is essentially just a cross-sectional study, meaning that it cannot serve to answer the question about the longitudinal association between age and the capacity for dialectical thinking, self-control, and empathy. Future longitudinal investigations (e.g., a follow-up survey can be conducted with a group of eight-year-olds to explore the trends of these three abilities from 8 to 25 years old) and cross-lagged analyses would help to address these limitations.

Second, because all of our data came from student self-assessment, although we emphasized the authenticity and confidentiality of questionnaire responses during the student response process, issues such as social desirability and student concerns may have influenced the data collected on students' dialectical thinking skills, self-control, and empathy, and future research could evaluate the above three skills in terms of peers, teachers and parents.

Third, there were limitations regarding the participants. As we found in the analysis of demographic variables, SES, including the role of juvenile parents and place of residence, is correlated with the three abilities of minors. However, we only sampled the population in province S, and although province S, as the second most populous province in the PRC, is highly representative, its representativeness to highly developed economic regions such as Beijing and Shanghai has yet to be verified due to the limitations of its economic growth. Therefore, subsequent studies should sample from populations nationwide to explore whether there are differences in dialectical thinking, self-control, and empathy among adolescents from different regions, ethnic groups and SES. The development of juveniles varies greatly from country to country due to differences in history, culture, level of economic development, and geography, so our conclusions cannot be universally applied to all nations (Yin and Du, 2014). Each state should choose a minimum age of criminal responsibility according to the developmental situation of its juveniles.

Last, in our study, age as the basis of criminal responsibility is challenged only from the aspect of psychology, which requires a more scientific basis, such as evidence from neuroscience and physiology (Carroll, 2015; Steinberg, 2017). For example, J.D.T., a 10-year-old boy sexually assaulted a 5-year-old boy, was controversially charged by the federal government, as J.D.T. had an undetectable level of testosterone in his bloodstream (Hamilton and Turner, 2015). As we mentioned in the discussion section, current explanations of adolescent development are based more on brain science. More research that directly links age differences in brain structure and function to age differences in legally relevant capacities and capabilitie (e.g., dialectical thinking, self-control, and empathy) is needed. In light of recent developments in neuroscience, researchers will need to focus on age differences in brainsystems and differences in brain regions or structures considered independently, and how brain development affects adolescent behavior.

CONCLUSION

Though, age as a criterion to determine the criminal responsibility of minors has economic benefits (the distinction is clear and simple, and normal circumstances do not require a lot of legal procedures to confirm), psychological science and neuroscience tend to challenge the public view, that the relationship between the age of adolescents and the index of criminal responsibility capacity (adolescents' dialectical thinking ability, self-control ability and empathy ability) is more complicated because of the non-linear development of the certain traits. In a word, age is not the only basis on which to judge a juvenile's criminal responsibility. In recent years, many countries have chosen to combat juvenile delinquency by lowering the age of criminal responsibility. Not only is this measure contrary to the intent of the UN Convention on the Rights of the Child and inconsistent with adolescents' developmental patterns; there is also clear,

overwhelming evidence that exposing adolescents to the justice system too early is not conducive to their rehabilitation. Therefore, it is urgent to reflect on how to set a minimum age of criminal responsibility and balance the relationship between the punishment of juvenile crimes and the protection of victims' rights.

Raise the Minimum Age of Criminal Responsibility

As we found that the slow development of dialectical thinking ability in adolescents (emerging adults), and their ability to control themselves sharply during adolescence (around age 15), a more desirable compromised solution or measure would be for political states to raise the age of criminal responsibility for adolescents, rather than the current orientation toward lowering it (Fowler and Kurlychek, 2018; Hong, 2020). The benefits of raising the minimum age of criminal responsibility would be to confirm age differences in legally relevant ones (the boundary is relatively clear, saving lots of judicial review resources), and further highlight the protection of the rights of young people to develop. The approach for minors will weaken the label effect caused by their crimes, which is not only conducive to the correction of minors' deviant behavior, but also conducive to the re-socialization of minors after education and guidance. on the contrary, the way the current political countries lower the minimum age of criminal responsibility to counter their deviant behavior is undoubtedly shirking the responsibility of the state and society, as minors are at the social stage, easily influenced by the social environment, and their deviant behavior needs more tolerance and positive guidance.

Accept the Rebuttable Presumption of Doli Incapax

The rebuttable presumption of doli incapax, derived from ancient Roman law, needs to be taken into account. This system measures the capacity to commit a crime not so much in terms of age as in terms of the understanding and judgment of the juvenile offender (Blackstone, 1966). Faced with criminal cases, juveniles below the minimum age of criminal responsibility (we encouraged legislators firstly to set a relatively higher age of criminal responsibility) can be pursued if the prosecution can provide the court with a "very clear and complete evidence" that the accused knew what they were doing was "seriously wrong" (or presumed to have mens rea)(Van Krieken, 2013; Lennings and Lennings, 2014). This necessitates a professionally qualified person to assess young children's capacity for criminal responsibility, including their cognitive, moral, emotional, psychological, and social growth [South African, Child Justice Act 75 of 2008, s11(2),(3)]. Doli incapax is consistent with the concept of criminal responsibility and the fact of juvenile development (Crofts, 2016); it is also in line with the basic principle of criminal law that "no penalty should be applied to a person unless he [or she] has had [the] capacity and a fair opportunity to adapt his [or her] conduct to the law" (Hart, 1968, p. 181).

In brief, we encouraged legislators to set a relatively higher age of criminal responsibility. Juveniles below this age can be pursued if the prosecution can prove they committed a crime with mens rea.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

YS: ideas, data collection, writing, and revisions. YF: data analysis, writing, and revisions. BM: data analysis and writing. LW: data

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.781669/full#supplementary-material

REFERENCES

Arnett, J. J. (2000). Emerging adulthood. a theory of development from the late teens through the twenties. *Am. Psychol.* 55, 469–680. doi: 10.1037/0003-066x.55.5.469

Baker, T., Cleary, H. M., Pickett, J. T., and Gertz, M. G. (2016). Crime salience and public willingness to pay for child saving and juvenile punishment. *Crime Delinq.* 62, 645–668. doi: 10.1177/0011128713505487

Blackstone, W. (1966). *Commentaries on the Laws of England: 1765-1769*. London: Dawsons.

Blakemore, S. J., and Mills, K. L. (2014). Is adolescence a sensitive period for sociocultural processing? *Annu. Rev. Psychol.* 65, 187–207. doi: 10.1146/annurev-psych-010213-115202

Blumstein, A. (1995). Violence by young people: why the deadly nexus. *Natl. Inst. Justice J.* 229, 2–9.

Boucher, H. C. (2011). The dialectical self-concept II: cross-role and within-role consistency, well-being, self-certainty, and authenticity. *J. Cross Cult. Psychol.* 42, 1251–1271. doi: 10.1177/0022022110383316

Brink, T. T., Urton, K., Held, D., Kirilina, E., Hofmann, M. J., Klann-Delius, G., et al. (2011). The role of orbitofrontal cortex in processing empathy stories in 4- to 8-year-old children. *Front. Psychol.* 2:80. doi: 10.3389/fpsyg.2011.00080

Broidy, L., Cauffman, E., Espelage, D. L., Mazerolle, P., and Piquero, A. (2003). Sex differences in empathy and its relation to juvenile offending. *Violence Vict.* 18, 503–516. doi: 10.1891/vivi.2003.18.5.503

Brown, A., and Charles, A. (2019). The minimum age of criminal responsibility: the need for a holistic approach. *Youth Justice* 21, 153–171. doi: 10.1177/1473225419893782

Butts, J. A., and Mitchell, O. (2000). Brick by brick: dismantling the border between juvenile and adult justice. *Crim. Justice* 2, 167–213.

Carroll, J. E. (2015). Brain science and the theory of juvenile mens rea. *NCL Rev.* 94, 539.

Casey, B. J. (2014). Beyond simple models of self-control to circuit-based accounts of adolescent behavior. *Annu. Rev. Psychol.* 66, 295–319. doi: 10.1146/annurev-psych-010814-015156

Chein, J., Albert, D., O'Brien, L., Uckert, K., and Steinberg, L. (2011). Peers increase adolescent risk taking by enhancing activity in the brain's reward circuitry. *Dev. Sci.* 14, F1–F10. doi: 10.1111/j.1467-7687.2010.01035.x

Cheng, C. (2009). Dialectical thinking and coping flexibility: a multimethod approach. *J. Pers.* 77, 471–494. doi: 10.1111/j.1467-6494.2008.00555.x

Choudhury, S., Charman, T., and Blakemore, S. J. (2008). Development of the teenage brain. *Mind Brain Educ.* 2, 142–147. doi: 10.1111/j.1751-228x.2008.00045.x

Cook, P. J., and Laub, J. H. (2002). After the epidemic: recent trends in youth violence in the United States. *Crime Justice* 29, 1–37. doi: 10.1086/652218

Crick, N. R., and Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychol. Bull.* 115:74. doi: 10.1093/deafed/enw030

Crofts, T. (2016). Reforming the age of criminal responsibility: state of the science. *S. Afr. J. Psychol.* 46, 436–448. doi: 10.1177/0081246316640116

Darrick, J., and David, P. F. (2006). Development and validation of the basic empathy scale. *J. Adolesc.* 29, 589–611. doi: 10.1016/j.adolescence.2005.08.010

Decety, J., Michalska, K. J., and Akitsuki, Y. (2008). Who caused the pain? An fMRI investigation of empathy and intentionality in children. *Neuropsychologia* 46, 2607–2614. doi: 10.1016/j.neuropsychologia.2008.05.026

Doreleijers, T. A. H., and Fokkens, J. W. (2010). Minderjarigen en jongvolwassenen: pleidooi voor een evidence based strafrecht. *Rechtstreeks* 2, 8–47.

Elliott, C. (2011). Criminal responsibility and children: a new defence required to acknowledge the absence of capacity and choice. *J. Crim. Law* 75, 289–308. doi: 10.1350/jcla.2011.75.4.717

Feng, Z., Mengjie, T., and Qiaoqiao, C. (2021). Investigation on self-control of high school students during epidemic period and its cultivation strategy. *Mod. Prim. Second. Educ.* 37, 6–12.

Fowler, E., and Kurlychek, M. C. (2018). Drawing the line: empirical recidivism results from a natural experiment raising the age of criminal responsibility. *Youth Violence Juv. Justice* 16, 263–278.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., et al. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proc. Natl. Acad. Sci. U.S.A.* 101, 8174–8179. doi: 10.1073/pnas.0402680101

Goldson, B. (2013). 'Unsafe, unjust and harmful to wider society': grounds for raising the minimum age of criminal responsibility in England and Wales. *Youth Justice* 13, 111–130.

Gottfredson, M. R., and Hirschi, T. (1990). *A General Theory of Crime*. Stanford, CA: Stanford University Press.

Guyer, A. E., Choate, V. R., Pine, D. S., and Nelson, E. E. (2012). Neural circuitry underlying affective response to peer feedback in adolescence. *Soc. Cogn. Affect. Neurosci.* 7, 81–92. doi: 10.1093/scan/nsr043

Hamamura, T., Heine, S., and Paulhus, D. (2008). Cultural differences in response styles: the role of dialectical thinking. *Pers. Individ. Differ.* 44, 932–942. doi: 10.1177/1948550612468774

Hamilton, L. D., and Turner, K. E. (2015). Juvenile delinquency and mens rea. *J. Offender Couns. Serv. Rehabil.* 5, 65–76.

Harpur, T. J., Hakstian, A. R., and Hare, R. D. (1988). Factor structure of the Psychopathy Checklist. *J. Consult. Clin. Psychol.* 56, 741.

Hart, H. L. A. (1968). *Punishment and Responsibility*. Oxford: Oxford University Press.

Haviland, J. M., and Lelwica, M. (1987). The induced affect response: 10-week-old infants' responses to three emotion expressions. *Dev. Psychol.* 23, 97–104. doi: 10.1037/0012-1649.23.1.97

Hong, S. (2020). A study on the dispute of abolition of the child law and alternatives. *Korean Assoc. Police Sci. Rev.* 22, 249–272. doi: 10.24055/kaps. 22.2.11

Huang, L., and Su, Y. (2010). Cognitive regulation and emotion sharing processes in empathy and their relationship. *J. Southw. Univ. Soc. Sci. Edn.* 06, 13–19. doi: 10.4324/9781315806556-8

Inhelder, B., and Piaget, J. (1958). *The Growth of Logical Thinking from Childhood to Adolescence: An Essay on the Construction of Formal Operational Structures*, Vol. 22. Hove: Psychology Press.

Ji, D., Hong, C., Miao, C., and Xinying, L. (2018). The influence of family atmosphere and parents' educational level on adolescent depression. *Chin. Public Health* 34, 38–41.

Jin, Y., and Yu, L. (2015). Responses to missing data in U.S. longitudinal surveys and implications for China. *Mod. Manag. Sci.* 09, 33–35.

Jolliffe, D., and Farrington, D. P. (2004). Empathy and offending: a systematic review and meta-analysis. *Aggress. Violent Behav.* 9, 441–476.

Jolliffe, D., and Farrington, D. P. (2006a). Examining the relationship between low empathy and bullying. *Aggress. Behav.* 32, 540–550. doi: 10.1016/j.adolescence. 2008.01.001

Jolliffe, D., and Farrington, D. P. (2006b). Development and validation of the basic empathy scale. *J. Adolesc.* 29, 589–611.

Kang, J. (2020). *The Relationship between Moderation thinking, Forgiveness and Aggression and Its Intervention in College Students.* Master's thesis. Nanchang: Jiangxi normal university.

Kunzmann, U., Wieck, C., and Dietzel, C. (2018). Empathic accuracy: age differences from adolescence into middle adulthood. *Cogn. Emot.* 32, 1611–1624. doi: 10.1080/02699931.2018.1433128

Lennings, N. J., and Lennings, C. J. (2014). Assessing serious harm under the doctrine of doli incapax: a case study. *Psychiatry Psychol. Law* 21, 791–800. doi: 10.1080/13218719.2014.894902

Li, C., Yue, L., Jie, L., and Jie, Z. (2011). Initial Revision of the basic empathy scale in a Chinese adolescent population. *Chin. J. Clin. Psychol.* 19, 163–166.

Lin, C., and Qingan, L. (2005). Characteristics of physical and mental development in adolescence. *J. Beijing Norm. Univ. Soc. Sci. Edn.* 48–56.

Liu, S., Kaiping, P., Fong, H., Yitian, M., and Zhuo, L. (2013). A probe into the causal relationship between dialectical thinking and dialectical emotion. *Psychol. Behav. Res.* 11, 457–462.

Liu, X., and Cui, L. (2020). Exploring the characteristics of empathic ability of left-behind children, factors influencing their development and their intervention. *J. Suzhou Univ. Educ. Sci. Edn.* 8, 65–75.

Miller, P. A., and Eisenberg, N. (1988). The relation of empathy to aggressive and externalizing/antisocial behavior. *Psychol. Bull.* 103, 324–344. doi: 10.1037/ 0033-2909.103.3.324

Narvey, C., Yang, J., Wolff, K. T., Baglivio, M., and Piquero, A. R. (2021). The interrelationship between empathy and adverse childhood experiences and their impact on juvenile recidivism. *Youth Violence Juv. Justice* 19, 45–67.

Newton, N. C., and Bussey, K. (2012). The age of reason: an examination of psychosocial factors involved in delinquent behaviour. *Legal Criminol. Psychol.* 17, 75–88. doi: 10.1080/15389580802335117

Nisbett, R. E., Peng, K., Choi, I., and Norenzayan, A. (2001). Culture and systems of thought: holistic versus analytic cognition. *Psychol. Rev.* 108:291.

Noroozi, M., Singh, I., and Fazel, M. (2018). Evaluation of the minimum age for consent to mental health treatment with the minimum age of criminal responsibility in children and adolescents: a global comparison. *Evid. Based Mental Health* 21, 82–86. doi: 10.1136/ebmental-2018-300032

NPC (2020). *Draft Amendments to the Criminal Law Again Adjust the Provisions of the Crime of Underage Minors.* Available online at: http://www.npc.gov. cn/npc/c35778/202012/2fadc51e4ec04346903567d2983c3a3d.shtml (accessed December 23, 2021).

O'Brien, W., and Fitz-Gibbon, K. (2017). The minimum age of criminal responsibility in Victoria (Australia): examining stakeholders' views and the need for principled reform. *Youth Justice* 17, 134–152. doi: 10.1177/ 1473225417700325

Pillay, A. L. (2019). The minimum age of criminal responsibility, international variation, and the Dual Systems Model in neurodevelopment. *J. Child Adolesc. Ment. Health* 31, 224–234. doi: 10.2989/17280583.2019.1692851

Rosenfield, R. L., Bachrach, L. K., Chernausek, S. D., Gertner, J. M., Gottschalk, M., Hardin, D. S., et al. (2000). Current age of onset of puberty. *Pediatrics* 106, 622–623. doi: 10.1542/peds.106.3.622

Schmidt, E. P., Rap, S. E., and Liefaard, T. (2020). Young adults in the justice system: the interplay between scientific insights, legal reform and implementation in practice in The Netherlands. *Youth Justice* 21, 172–191.

Schwenck, C., Ciaramidaro, A., Selivanova, M., Tournay, J., Freitag, C. M., and Siniatchkin, M. (2017). Neural correlates of affective empathy and reinforcement learning in boys with conduct problems: fMRI evidence from a gambling task. *Behav. Brain Res.* 320, 75–84. doi: 10.1016/j.bbr.2016.11.037

Sickmund, M., Sladky, T. J., Kang, W., and Puzzanchera, C. (2017). *Easy Access to the Census of Juveniles in Residential Placement 1997–2015.* Washington, DC: Office of Juvenile Justice and Delinquency Prevention.

Snyman, C. R. (2008). *Criminal Law.* Durban: LexisNexis.

Somerville, L. H. (2013). The teenage brain sensitivity to social evaluation. *Curr. Dir. Psychol. Sci.* 22, 121–127. doi: 10.1177/0963721413476512

Spencer-Rodgers, J., Peng, K., and Wang, L. (2010). Dialecticism and the co-occurrence of positive and negative emotions across cultures. *J. Cross Cult. Psychol.* 41, 109–115. doi: 10.1177/0022022109349508

Spencer-Rodgers, J., Peng, K., Wang, L., and Hou, Y. (2004). Dialectical self-esteem and east-west differences in psychological well-being. *Pers. Soc. Psychol. Bull.* 30, 1416–1432. doi: 10.1177/0146167204264243

Spencer-Rodgers, J., Srivastava, S., Boucher, H. C., English, T. L., Paletz, S. B., Hou, Y., et al. (2008). *The Dialectical Self Scale.* Unpublished Manuscript. Berkeley, CA: University of California, Berkeley.

Steinberg, L. (2017). Adolescent brain science and juvenile justice policymaking. *Psychol. Public Policy Law* 23:410. doi: 10.1037/law0000128

Steinberg, L., Albert, D., Cauffman, E., Banich, M., Graham, S., and Woolard, J. (2008). Age differences in sensation seeking and impulsivity as indexed by behavior and self-report: evidence for a dual systems model. *Dev. Psychol.* 44, 1764. doi: 10.1037/a0012955

Steinberg, L., and Monahan, K. C. (2007). Age differences in resistance to peer influence. *Dev. Psychol.* 43, 1531–1543. doi: 10.1037/0012-1649.43.6. 1531

Tan, J. (2021). *The Relationship Among Emotional Resilience, Self-Control and Aggression of Middle School Students and Intervention Study.* Hefei: Hefei Normal University.

Toga, A. W., Thompson, P. M., and Sowell, E. R. (2006). Mapping brain maturation. *Focus* 29, 148–390. doi: 10.1016/j.tins.2006.01.007

Van Krieken, R. (2013). *Doli Incapax and Its Vicissitudes: Childhood and Criminal Responsibility in England, Germany and Australia. Germany and Australia.* 185–262. Available online at: https://ssrn.com/abstract=2319773 (accessed September 3, 2013).

Walters, G. D. (2016). Are behavioral measures of self-control and the Grasmick self-control scale measuring the same construct? A meta-analysis. *Am. J. Crim. Justice* 41, 151–167. doi: 10.1007/s12103-015-9317-3

Wang, H., and Lu, J. (2004). Development of a self-control questionnaire for secondary school students and its investigation. *Psychol. Sci.* 198–203.

Wang, Q., Zan, L., and Yanjie, S. (2021). The lifelong development of empathy and its neural basis. *Sci. China Life Sci.* 51, 717–729. doi: 10.1360/ssv-2020-0250

Xue, S., Yan, Y., and Siwei, Z. (2019). 13-Year-Old Boy in Dalian Charged with Murder is Not Punished According to Law,Lawyer Expert: Punishment is Needed to Curb Evil. Available online at: https://www.thepaper.cn/newsDetail_ forward_4767268 (accessed December 23, 2021).

Yang, Y., Tang, Y., Peng, W. W., Lü, X., Li, H., and Chen, J. (2017). Empathy: the genetics-environment-endocrine-brain mechanism (in Chinese). *Chin. Sci. Bull.* 62, 3729–3742.

Ye, H., and Li, Y. (2005). The negative influence of internet on Teenagers' thinking and corresponding countermeasures. *Educ. Guide* 35–36.

Yin, X., and Du, A. (2014). "Chinese and American youth will cross-cultural studies," in *Proceedings of the Abstracts of the 17th National Conference of Psychology* (Beijing: Chinese Psychological Association), 1210–1212.

Zhang, M. (1994). Exploration of some issues of criminal responsibility capacity. *J. Zhongnan Polit. Sci. Law Univ.* 55–58. doi: 10.1080/16549716.2020.1726722

Zhang, P. (2014). *A Study on the Correlation between Thinking Styles and Coping Styles and Mental Health of Secondary School Students.* Master's thesis. Chongqing: Chongqing Normal University.

Zhang, X., Dingguo, G., and Hua, F. (2011). Dialectical thinking reduces aggression. *Acta Psychol. Sin.* 43, 42–51. doi: 10.3724/sp.j.1041.2011.00042

Zheng, Y. (2018). Research on the influence of mobile Internet on Teenagers and countermeasures. *New West China* 82–83.

Check for
updates

Justification of Sentencing Decisions: Development of a Ratio-Based Measure Tested on Child Neglect Cases

Eiichiro Watamura[1]*, Tomohiro Ioku[1] and Toshihiro Wakebe[2]

[1] Graduate School of Human Sciences, Osaka University, Suita, Japan, [2] Division of Psychology, Department of Human Sciences, Seinan Gakuin University, Fukuoka, Japan

Theoretically, people's justification of a sentencing decision involves a hybrid structure comprising retribution, incapacitation, general deterrence, and rehabilitation. In this study, a new ratio-type measure was developed to assess this structure and was tested to detect changes in the weighting of justification according to the content emphasized in a particular crime. Two child neglect scenarios were presented to participants, where they read either a severe-damage scenario (where a single mother's selfish neglect caused her son's death) or a moderate-damage scenario (where a single mother became apathetic due to economic deprivation and caused her child's debilitation). Participants then indicated the proportion of importance they placed on each justification in determining the defendant's punishment, with an overall proportion of 100%, along with responding to the sentence on an 11-point scale. This study involved a two-factor analysis of variance for justification ratios, a *t*-test for the sentence, and a multiple regression analysis with three demographic variables, the four justifications as independent variables, and the sentence as the dependent variable. The ratio of retribution to rehabilitation was reversed depending on the scenario: in the severe-damage scenario, retribution was weighted highest at 27.0% and rehabilitation was weighted at only 19.0%. By contrast, in the moderate-damage scenario, rehabilitation had the highest weighting of about 26.2%, while retribution was weighted at 21.5%. The sentence was more severe in the severe-damage scenario. Multiple regression analysis suggested that in the severe-damage scenario, most participants failed to deviate from choosing retribution by default and decided on heavier sentences, while some who considered rehabilitation and incapacitation opted for lighter sentences. The present measure succeeded in detecting changes in the weighting of justification, which can be difficult to detect with common Likert Scales. In addition, it was found that not only retribution but utilitarian justification was considered in the sentencing decisions of serious cases.

Keywords: sentencing decision, punishment, retribution, rehabilitation, justification, sentencing criteria, judicial sentence severity

INTRODUCTION

Sentencing justification is based on a hybrid of four categories: retribution, incapacitation, general deterrence, and rehabilitation (Robinson, 1987; Exum, 2017; Hoskins, 2020). Retribution can be considered as a past-oriented form of justification, whereby the offender is given a punishment that is commensurate with the severity of the crime, thereby correcting a moral imbalance (Carlsmith, 2006). Therefore, according to the retribution approach, punishment should be proportionate to the severity of the crime. Utilitarian justification, on the other hand, is a future-oriented, pragmatic perspective aiming to deter future crimes by influencing the criminals themselves or society in general (e.g., Vidmar and Miller, 1980; Weiner et al., 1997; Carlsmith, 2006). Utilitarian justifications can be categorized into incapacitation, general deterrence, and rehabilitation (or education) (McFatter, 1978; Robinson, 1987; McMunigal, 1998). Incapacitation is based on the assumption that the cause of crime is inherent in the offender and attempts to deter future crimes by isolating criminals from society for a certain period of time, such as by imposing long prison sentences for dangerous offenders (Goldman, 1982). General deterrence means that laypeople should be deterred from becoming potential criminals by showing them that they will be punished as severely as possible if they break the law (Nagin, 1998). When the incidence of crime is high, and the arrest rate is low, the need to make an example of the offender is high, and sentences for those who have been taken into custody become more severe. Rehabilitation seeks to deter future crime by working directly with offenders. This is more favorable to criminals than the other aforementioned approaches to justification; rehabilitation aims to reduce criminal intent and ultimately transform a person into a law-abiding citizen who can contribute to society (Robinson, 1987; Cotton, 2000). Although these four types are not the only sentencing goals possible for punishment, they are the most commonly endorsed by the public in the justice system (e.g., McFatter, 1978; Cotton, 2000) and have been adopted in several empirical studies examining sentencing decisions of the general population (e.g., Roberts and Gebotys, 1989; Templeton and Hartnagel, 2012; Niang et al., 2020).

Previous empirical studies have reported that the public adopts retribution as the predominant or nearly sole justification (McCorkle, 1993; Weiner et al., 1997; Carlsmith et al., 2002; Oswald et al., 2002; Orth, 2003; Rucker et al., 2004; Carlsmith, 2006, 2008; Alter et al., 2007; Carlsmith and Darley, 2008; Gromet and Darley, 2009; Okimoto et al., 2009; Keller et al., 2010; Watamura et al., 2011; Gerber and Jackson, 2013; Twardawski et al., 2020). With respect to sentencing decisions for serious crimes such as murder, retribution is the default justification "...their natural (default) approach to sentencing probably involved retribution" (Carlsmith, 2006, p. 447). However, the predominance of retribution does not imply that the other justifications are not considered. Moreover, retribution and other justifications are not necessarily conflicting (Crockett et al., 2014). For instance, long imprisonment may offset serious harm (retribution) and through punishment, restore justice that would otherwise be lost to crime. At the same time, it may support the transformation of offenders into citizens that disengage from crime (rehabilitation), withhold them from opportunities to reoffend (incapacitation), and intimidate potential offenders (general deterrence).

Prior studies have been limited in their ability to capture the relative weight people assign to the four justifications. With some exceptions (e.g., O'Toole and Fondacaro, 2017), previous studies have involved participants rating each of the four justifications for sentencing separately (e.g., Krosnick, 1999; McKee and Feather, 2008; Berryessa, 2018). Thus, the purpose of this study was to develop a measure that can assess mixed justifications of punishment. Furthermore, the new measure developed as part of this study was tested on the same type of offense to detect changes in the weighting of justification according to the emphasized content.

According to the hybrid theory, people do not determine punishment by retribution alone. As an individual characteristic, the tendency to blame and seek retribution against offenders is positively correlated (0.70) with permissive utilitarianism, which considers inflicting severe punishment as a means of deterrence (Yamamoto and Maeder, 2021). An experiment with college students found that as the length of incarceration increased, punishment appropriateness ratings increased, and participants were also more positive about the acceptance of offender rehabilitation (Brubacher, 2019). This suggests that participants perceived incarceration to be effective for both retribution and rehabilitation. In a scenario experiment conducted with members from the general population (Spiranovic et al., 2012), the most common justification chosen for sentencing serious crimes was a mixture of retribution and utilitarianism (burglary 36.0%, assault 34.4%); a single justification, including retribution, was less commonly selected. In the trust game paradigm (Cañadas et al., 2015), in which both parties maximize their mutual benefit by repeating the process of returning some or all of the money entrusted by the other party without monopolizing the money, the experimental manipulation of whether or not the punishment is accompanied by the message "I have punished you" can more clearly identify the justification of punishment. If the message is not conveyed to the punished party, the punishment is only self-satisfying and will not deter the next betrayal. In a study by Crockett et al. (2014), participants who played the role of the punisher were motivated to reduce the amount of money distributed to participants who acted as violators (i.e., punish them) by two types of justification: retribution, which without the message, seeks to punish violators based on mere moral revulsion (not related to the possibility of deterrence), and utilitarian justification, which with the message, seeks to deter violations of the distribution rule through punishment. In other words, in the present-message condition in this study, retribution and deterrence justifications are mixed. Thus, the general population determines the output punishment by changing the weighting of any of these justifications (i.e., by assigning more or less weight).

Unfortunately, assessment of the hybrid structure of sentencing justification is currently limited because of challenges in inferring the exact ratio. The most popular way to measure justification is through a unipolar Likert Scale. In this scale—for the items that ask, "how important is this justification?"—the scores are moderately or more highly aligned for almost all

responses (i.e., acquiescence-response bias; Krosnick, 1999). For example, in McKee and Feather's (2008) seven-point scale, the overall sentencing decisions for criminal offenders were retribution ($M = 4.66$, $SD = 1.16$), incapacitation ($M = 4.49$, $SD = 0.99$), general deterrence ($M = 4.74$, $SD = 1.06$), and rehabilitation ($M = 4.28$, $SD = 1.20$). In Berryessa's (2018) study, which asked participants to respond with a finer scale ranging from 1 to 100, the scores for homicide related to the four justifications were: retribution ($M = 68.83$, $SD = 30.03$), incapacitation ($M = 77.41$, $SD = 24.21$), general deterrence ($M = 77.85$, $SD = 26.02$), and rehabilitation ($M = 62.06$, $SD = 34.17$). Similar trends were observed for the scenario of rape on the scores of retribution ($M = 69.30$, $SD = 28.71$), incapacitation ($M = 75.24$, $SD = 27.21$), general deterrence ($M = 70.29$, $SD = 23.32$), and rehabilitation ($M = 63.56$, $SD = 31.70$). Even with the fine-tuning of scores, most offenses had medium to high scores for all justifications. Some studies, wary of this lopsided distribution of scores, have used the scale in a manner that forced a trade-off between retribution and another justification. O'Toole and Fondacaro (2017) used the item "Relative to giving a young offender what he deserves, how important is it to you that the juvenile justice system improve the young offender's psychological well-being?" to measure relative support for rehabilitation vs. retribution. However, although this scale shows the relative ratios of the two, it fails to convey the weighting of all four justifications.

In such cases, how can we assess that hybrid structure in which the sentencing justifications trade-off against each other? One solution is to implement the theoretical concept of four hybrids (Robinson, 1987; Exum, 2017; Hoskins, 2020) with the Summation Model (Hollands and Spence, 1998). According to the Summation Model, an anchor that represents the whole (such as 100%) makes it easier to perform the percentage judgment task intuitively. Therefore, a measure was devised in which the entire sentencing purpose was set to 100%, and the weighting of each justification was given a numerical input. For the description of the four justifications, Berryessa's (2018) items were used for clarity and brevity (p. 245).

Since the Japanese judicial system is similar to the jury system in Germany and other European countries, the four types of justification have been examined in previous studies (e.g., Gollwitzer and Bücklein, 2007). Consistent with European and American study findings, the general Japanese public demonstrates the strongest preference for retribution (Kita and Johnson, 2014). Japan is the only industrialized country, other than the United States, to have the death penalty; its support for the death penalty varies from survey to survey, but it is high, at over 60%, due to the high support for notions of retribution, such as "life should be paid for with life" (Jiang et al., 2010b; Andreescu and Hughes, 2020). Despite the shared dominance of retribution in Japan with Europe and the United States, as a collectivist culture (Kitayama et al., 2009), Japan possesses the unique feature of strong social norms that seek adherence to social values by punishing perpetrators. Thus, Japan also demonstrates high support for general deterrence. In fact, studies comparing Japan with the United States have consistently reported stronger support for general deterrence

in Japan (Gollwitzer and Bücklein, 2007; Jiang et al., 2010a). However, the degree of support toward general deterrence among Japanese people is not, in fact, clear; neither is its relative weightage in terms of other justifications, including retribution. Accordingly, the new measure developed in this study may help clarify the unique Japanese cultural characteristics. In the current study, we examined the hybrid structure that determines people's sentencing decisions in Japan, where both retribution and general deterrence are dominant.

MATERIALS AND METHODS

Overview

In this study, we examined whether the weighting of justification changed between a severe-damage scenario, where the damage was severe, and the offender's rehabilitation potential was low, and its opposite, a moderate-damage scenario, where the damage was moderate, and the offender's rehabilitation potential was high. We also examined the effect of the difference in weighting justification on the sentence. It was predicted that the ratio of retribution would be higher in the severe-damage scenario, resulting in a severer sentence, while the ratio of rehabilitation would be higher in the moderate-damage scenario, as the defendant would have a higher chance of being rehabilitated in contexts of less damage. Prior to the main study, a preliminary survey was conducted, and two scenarios were finalized for inclusion in the study.

Preliminary Survey

In this experiment, where the new measure was tested for the first time, child neglect, a type of child abuse, was selected as the offense type over common violent offenses, as the offense needed to be manipulated for the severity and rehabilitation potential to be more pronounced. While the effect of the defendant's experience of child abuse on justification has been studied (Berryessa, 2018, 2021), the effect of justification on sentencing decisions for child abuse also needs to be examined to break the negative cycle that leads to subsequent abuse. In Japan, where this study was conducted, public attention to child neglect deaths has increased immensely because of increased media coverage of recent fatal incidents and warnings from media experts (Takikawa, 2019). As a result, there are growing demands for harsher punishments for convicted parents. The fact that the crime was committed by the person who should protect the child heightens the sense of moral seriousness. It was postulated that if the crime was committed by a single mother with a compelling motive, such as poverty, the defendant would be seen as more likely to be rehabilitated. In the main study, it seemed important to compare scenarios with completely different sentences to determine the output of justifications. A preliminary survey was conducted online with participants from the general public ($N = 135$, female = 68, male = 67, $M_{age} = 50.68$, $SD = 12.43$), recruited from a Japanese internet research company. The participants read one of four scenarios combining two levels of damage (severe/moderate) and two levels of rehabilitation potential (low/high) (between-participants

design) and judged the sentence for the defendant on the same scale as the main study (11-point scale, see below). The results showed that only one scenario, namely, moderate-damage and high potential, had the lowest sentence score ($M = 6.63$, $SD = 2.67$, $p < 0.01$). The main study's purpose of comparing justification ratios cannot be achieved unless a comparison is made between scenarios with different sentences as the output. However, since the other three scenarios were almost identical ($M = 7.97$–8.46, $SD = 1.65$–2.82), suggesting that it is also difficult to separate the two factors, we decided to use the severe-damage and low potential scenario ($M = 8.41$, $SD = 2.30$) as representative of the three and compared the same with the only scenario with a different sentence. As explained earlier, these two scenarios are "the severe-damage scenario" and "the moderate-damage scenario." The study was conducted online after obtaining ethical review approval from Osaka University in accordance with the guidelines of the Japanese Psychological Association. All data have been published on the Open Science Framework platform.[1]

Power Analysis

As the required effect size of the multiple regression analysis was unknown, it was assumed to be 0.15 based on the f^2 index suggested by Cohen (1988). The α was set at 0.05, the power of the test (1-β) was set at 0.80, and a power analysis was conducted in G*Power 3.1.9.7 (The G*Power Team, Heinrich Heine Universität) for a multiple regression analysis with seven predictors (four justifications plus three demographic variables). This analysis revealed that the required sample size was $N = 103$; hence, data were collected with a target of 103 for each scenario. In power analysis for other statistical tests, a much smaller required sample size was calculated. However, that would not have been adequate for the multiple regression analysis. Therefore, the final number of participants was determined based on the results of the power analysis for the multiple regression analysis.

Participants

A total of 264 participants were recruited from a panel of individuals aged 20+ years who were registered with a Japanese internet research company. They resided in 42 prefectures in Japan and were representative of lay judges. They provided information on gender, age, parental status, marriage, prefecture, and their job as demographic variables at the time of participation. With the exception of age, these variables were coded to dummy variables such as 0, 1, 2. The reward for participation was points (equivalent to about 20 cents) that could be exchanged for an Amazon gift card. Those who could not provide informed consent (8) and those who did not respond to the questions about the defendant or provided unintelligible responses (discussed later in the procedure) (31) were excluded from the study. As a result, the sample size for the analysis was 112 (female = 56, male = 56) participants in the severe-damage scenario ($M_{age} = 44.09$, $SD = 14.39$) and 113 (female = 57, male = 56) participants in the moderate-damage

scenario ($M_{age} = 0.44.45$, $SD = 15.36$). When the two groups were compared to examine differences in demographic variables, no significant differences were observed [gender: $\chi^2(1) = 0.004$, $p = 0.947$, age: $t(223) = 0.182$, $p = 0.855$, parental status: $\chi^2(1) = 1.957$, $p = 0.162$, marriage: $\chi^2(1) = 0.213$, $p = 0.644$, prefecture: $\chi^2(41) = 47.610$, $p = 0.222$, job: $\chi^2(10) = 8.244$, $p = 0.605$], and it was concluded that they were comparable in the subsequent analysis. The three variables of gender, age, and parental status were used as independent variables in the multiple regression analysis, along with the four justifications.

Experimental Design

The current experiment comprised five blocks: Scenario reading, attention control task, in-house newly developed questionnaire, sentencing decision, and scenario manipulation check block.

First, participants read a child neglect case scenario of approximately 230 words (**Supplementary Data**). There were two types of scenarios, and participants were randomly assigned either one.

Severe-Damage Scenario

A single mother neglected and starved her 2-year-old child for more than 6 days to spend time with her boyfriend.

Moderate-Damage Scenario

A single mother neglected her 2-year-old child for more than 30 h due to economic deprivation and loss of energy, and the child wasted away.

The points emphasized by participants changed depending on the scenario. In the severe-damage scenario, the severity of the child's death and the low rehabilitation potential derived from selfish motives were emphasized. By contrast, in the moderate-damage scenario, the child was in a harmful but not life-threatening situation, and the motive of poverty suggests a higher probability of rehabilitation.

After reading the scenario, in the second block, participants were asked to imagine themselves in the courtroom and write one question to the defendant. This question was a device used to make them read the scenario carefully, and the description was not analyzed.

Those who did not respond, or wrote an invalid question such as meaningless strings or "nothing" were excluded from the study, as they may not have read the scenario carefully. In the third block, participants responded to the new measure developed in this research, in which they were asked to enter positive integers to indicate the percentages of importance they placed on each of the four justifications: retribution, incapacitation, deterrence, and rehabilitation, as defined by Berryessa (2018). Notably, the present study considered the addition of another device, a dummy "precedent," that is not theoretically related to the four justifications, so that the sum of the five items including the dummy would be 100%. Without the dummy, increasing the ratio of any of the justifications will decrease the ratios of the others. In other words, the independence of the observed values cannot be satisfied, and multicollinearity is strong when there is such a relationship between the measures in the analysis. However, when the

dummy is included, "100—dummy" becomes the sum of the four justifications, and such a relationship is not necessarily established. As a result, the violation of independence of observed values is eliminated to some extent, and multicollinearity is mitigated. The new measurement had the following question and items: In deciding the punishment for this mother, how important are the following five items to you? Please assign a percentage to each such that the total is 100%. Items were presented randomly.

Retribution
Retribution relies on the idea that for justice to be served, an offender deserves to be punished in a manner that is proportionate to the severity and moral heinousness of the committed crime.

Incapacitation
Incapacitation aims to remove offenders from society to protect the public from future unlawful behavior.

General Deterrence
Deterrence attempts to prevent the future committal of crimes through the threat of future punishments that outweigh an individual's motivation to commit future criminal acts.

Rehabilitation
Rehabilitation seeks ways to actively reform and address the underlying reasons for an offender's criminal behavior so that an individual will not reoffend.

Precedent (Dummy)
The sentence should be determined based on the sentencing decisions handed down in previous abuse cases and judges' opinions.

Next, in the fourth block, participants responded to the item "Please choose the one that is closest to your idea of punishment for this mother," with the following 11 possible punishments for the defendant (Robinson and Darley, 1995; Brubacher, 2019): (1) No punishment, (2) 1 day in prison, (3) 2 weeks in prison, (4) 2 months in prison, (5) 6 months in prison, (6) 1 year in prison, (7) 3 years in prison, (8) 7 years in prison, (9) 15 years in prison, (10) 30 years in prison, and (11) life in prison. This particular question was designed to examine the difference in the severity of the sentence between the scenarios and the effect of justification on sentencing.

Finally, the participants responded to items to check the manipulation of the scenario. If the participants in the severe-damage scenario estimated the damage as severe, they would weight retribution. On the other hand, if the participants in the moderate-damage scenario estimated the possibility of the defendant's rehabilitation highly, they would weight rehabilitation. Note that the purpose of this block was not to determine which justifications increased but to make sure that the factors in the scenario that increased justification (i.e., the child's suffering, the mother's potential for rehabilitation) were considered by the participants. For this purpose, the items needed to include the person in the scenario, such as the mother or the child; thus, the present study used a tentative modification of

items from Weiner et al. (1997), which includes descriptions of specific persons. They rated the options on a 6-point scale ranging from "1: Do not at all agree" to "6: Very much agree." The four items, one for each of the four justifications, were as follows:

1. Compared to other serious cases, the pain the child has suffered is much worse.
2. In order to prevent the mother from making the same mistake, it is important to keep her out of society.
3. I cannot help but wonder if this kind of child abuse is happening more often.
4. It is not entirely impossible that the mother can be rehabilitated.

Statistical Analysis
In this study, four statistical analyses were conducted. Based on the percentages obtained through the new measure, averaged ratios were calculated for each justification by summing the percentages assigned to each justification (e.g., 30% for retribution) and dividing it by the number of participants in the scenario. The averaged ratios were compared by a two-factor analysis of variance (ANOVA) to determine the differences in scores between groups on each justification variable. Precedent, the dummy justification, was not examined here and later. In the second analysis, the means of the 11 magnitude levels were compared by a t-test to examine if there was a difference in sentencing severity between the two scenarios. Then, in the third analysis, multiple regression analysis was implemented to determine predictive relations between the four justifications, age, sex, and parental status as the independent variables and the magnitude of sentencing scores as the dependent variable. Prior to the multiple regression analysis, these dependent variables were standardized to have a mean of 0 and standard deviation of 1 by z-score normalization, as per previous studies (Weiner et al., 1997; Okimoto et al., 2009), because they differed in units due to being coded and converted to ratios. Moreover, we checked the assumptions of regression analysis by examining normality according to whether the residuals followed a normal probability-probability plot, homogeneity of variance according to whether the residuals were dispersed, and independence of observed values from the correlation coefficients. For the four manipulation check items measured by the 6-point scale, the mean scores of the child's suffering, which strengthens retribution, and the mother's rehabilitative potential, which strengthens rehabilitation, were calculated for each scenario and compared by t-test (we also compared the factors that increased incapacitation and general deterrence to confirm that there was no difference). All analyses were performed using the HAD (Shimizu, 2016), a statistical software program that can analyze Excel format data with high accuracy.

RESULTS

Justification
The bivariate correlations for justifications are shown in **Table 1**, and the mean ratio for each justification is shown in

		Severe-damage scenario					Moderate-damage scenario				
		M	*SD*	1	2	3	*M*	*SD*	1	2	3
1	Retribution	26.973	18.298				21.504	12.154			
2	Incapacitation	15.482	13.064	−0.105			13.593	9.471	0.202*		
3	General deterrence	21.839	16.425	−0.316**	−0.258**		21.584	13.426	−0.280**	−0.066	
4	Rehabilitation	19.009	15.464	−0.495**	−0.352**	−0.185[+]	26.239	20.039	−0.519**	−0.541**	−0.372**

**p < 0.01, *p < 0.05, +p < 0.10.

Figure 1. Note that the ratio of retribution to rehabilitation is completely reversed in each scenario. In the severe damage scenario, retribution and incapacitation showed significant negative correlations with general deterrence and rehabilitation, respectively ($rs < -0.258$, $ps < 0.01$). A slightly weaker negative correlation between general deterrence and rehabilitation was also shown ($r = -0.185$, $p < 0.10$). In the moderate-damage scenario, retribution showed a significant positive correlation with incapacitation ($r = 0.202$, $p < 0.05$) and negative correlations with general deterrence and rehabilitation ($rs = -0.280$, $ps < 0.01$). Moreover, incapacitation and general deterrence showed significant negative correlations with rehabilitation ($rs < -0.372$, $ps < 0.01$). A two-factor ANOVA for the mixed design with scenario and justification as independent variables revealed significant differences in the interaction effect [$F(3, 669) = 5.448$, $p = 0.002$; partial $\eta_p^2 = 0.024$] and the main effect of justification [$F(3, 669) = 14.024$, $p = 0.000$; partial $\eta_p^2 = 0.059$]. Multiple comparisons using the Holm method confirmed the prediction, and the test was successful. In the severe-damage scenario, retribution was the highest, accounting for 27.0%, significantly higher ($p = 0.015$, $d = 0.474$) than rehabilitation (19.0%) and with a difference of 8%. By contrast, in the moderate-damage scenario, rehabilitation was the highest (26.2%) and differed from retribution (21.5%) by approximately 5%, which was not significant. Multiple comparisons using Tukey's HSD method confirmed this inverse relationship; in the severe damage scenario, retribution was significantly higher than rehabilitation ($p < 0.001$), and in the moderate-damage scenario, rehabilitation was significantly higher than retribution ($p = 0.014$). To summarize, as predicted, the weighting of justification changed depending on the scenario, and the new measure confirmed that the ratio of retribution to rehabilitation was reversed.

Although not the focus of this study, incapacitation was found to have the lowest ratio in both scenarios. In the severe-damage scenario, incapacitation was significantly lower for both retribution and general deterrence ($p = < 0.001$, $d = 1.125$; $p = 0.006$, $d = -0.476$, respectively). In the moderate-damage scenario, the difference between incapacitation and rehabilitation, general deterrence, and retribution were all significant ($p < 0.001$, $d = -0.840$; $p < 0.001$, $d = -0.598$; $p < 0.001$, $d = -0.579$). As suggested as a Japanese cultural characteristic, general deterrence was consistently rated second highest, regardless of the scenario (severe-damage 21.8% vs. moderate-damage 21.6%).

Severity of Sentencing

A *t*-test was conducted to determine whether there was a difference in the sentence severity, as measured by the 11-point scale (Robinson and Darley, 1995; Brubacher, 2019) after the justification ratio-based measure, between the scenarios: the severe-damage scenario had a mean of 8.50 ($SD = 1.95$), and the moderate-damage scenario had a mean of 6.92 ($SD = 2.31$), confirming that the severe-damage scenario resulted in significantly heavier sentences, [$t(223) = 5.533$, $p < 0.001$, $d = 0.735$].

Effect of Justification on Sentencing

Next, multiple regression analysis was conducted for each scenario with sentence severity as the dependent variable and gender, age, parental status, and the four justifications as independent variables. The results are shown in **Tables 2, 3**. In the moderate-damage scenario, a higher ratio of rehabilitation tended to result in a lighter sentence ($\beta = -0.257$); however, the standard partial regression coefficients for all variables were non-significant. By contrast, in the severe-damage scenario, a higher ratio of incapacitation and rehabilitation was associated with a lighter sentence ($\beta = -0.450$, $p = 0.002$ and $\beta = -0.314$, $p = 0.011$, respectively). In both scenarios, demographic variables did not predict the sentence.
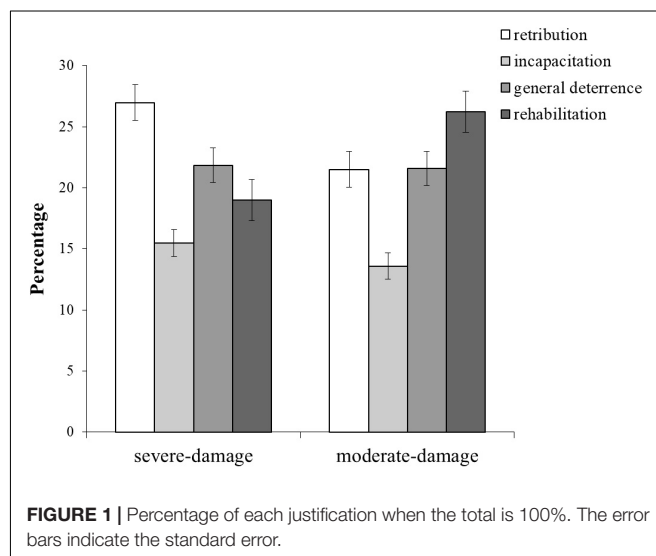

FIGURE 1 | Percentage of each justification when the total is 100%. The error bars indicate the standard error.

TABLE 2 | Regression analysis (severe-damage scenario).

	β	p	95% CI		VIF
Sex	0.045	0.593	−0.123	0.214	1.102
Age	−0.131	0.157	−0.312	0.051	1.281
Parental status	−0.070	0.439	−0.248	0.109	1.237
Retribution	0.114	0.406	−0.157	0.385	2.862
Incapacitation	−0.314*	0.011	−0.554	−0.075	2.232
General deterrence	0.116	0.337	−0.123	0.354	2.211
Rehabilitation	−0.450**	0.002	−0.733	−0.167	3.106
R^2	0.320	**			
Adjust R^2	0.274	**			

$F(7, 104) = 6.979, p < 0.001, AIC = 441.813, BIC = 466.279.$
$**p < 0.01, *p < 0.05, +p < 0.10.$
Coefficients represent standardized coefficients, CI, confidence interval; VIF, Variance Inflation Factor; AIC, Akaike information criterion; BIC, Bayesian information criterion.

TABLE 3 | Regression analysis (moderate-damage scenario).

	β	p	95% CI		VIF
Sex	0.114	0.243	−0.078	0.306	1.274
Age	0.082†	0.457	−0.135	0.298	1.622
Parental status	−0.030	0.752	−0.221	0.160	1.252
Retribution	0.190	0.172	−0.084	0.465	2.603
Incapacitation	0.123	0.307	−0.115	0.360	1.947
General deterrence	0.063	0.644	−0.206	0.332	2.506
Rehabilitation	−0.257	0.154	−0.613	0.098	4.363
R^2	0.227	**			
Adjust R^2	0.175	**			

$F(7, 105) = 4.401, p < 0.001, AIC = 497.941, BIC = 522.487.$
$**p < 0.01, *p < 0.05, +p < 0.10.$
Coefficients represent standardized coefficients, CI, confidence interval; VIF, Variance Inflation Factor; AIC, Akaike information criterion; BIC, Bayesian information criterion.

Scenario Manipulation Check

Based on a manipulation check item used to compare the mean ratings of the pain suffered by the child, the severe-damage scenario scored significantly higher than the moderate-damage scenario [4.68 (SD = 1.32) vs. 4.33 (SD = 1.26)], suggesting that the damage was considered more severe, [$t(223) = 2.03$, $p = 0.044$, and $d = 0.532$]. The mean rating of the defendant's rehabilitative potential was not significant but was higher for the moderate-damage scenario [3.59 (SD = 1.28) vs. 3.81 (SD = 1.25)] $t(223) = -1.28$, $p = 0.203$, and $d = -0.091$. Differences in the scores of other items were also not significant [incapacitation: 3.29 (SD = 1.43) vs. 2.99 (SD = 1.33), $p = 0.101$; general deterrence: 4.37 (SD = 1.34) vs. 4.43 (SD = 1.29), $p = 0.700$].

DISCUSSION

Availability and Theoretical Suitability of the Measure

As hypothesized, the results indicated that the weighting of justification changed according to the emphasized content, and the change could be detected by the new measure developed as part of this research. The results were completely symmetrical, with a higher rate of retribution observed in the severe-damage scenario and a higher rate of rehabilitation found in the moderate-damage scenario. Furthermore, it was found that all four justifications of retribution, incapacitation, general deterrence, and rehabilitation (although weighted differently) were considered in a certain ratio in the determination of punishment. Thus, the theoretical assumption of the four-hybrid structure of sentencing justification (Robinson, 1987; Exum, 2017; Hoskins, 2020) was supported. Furthermore, by showing that the weighting of retribution and rehabilitation was completely reversed in different scenarios, this study highlighted scenarios when retribution, considered the default justification, surrenders its place. Thus, the results suggest that the weighting of the hybrid structure can be flexible.

Coexistence of Retribution and Utilitarian Justification

Prior research has consistently demonstrated the tendency among the general population to make sentencing decisions based on retribution (e.g., Weiner et al., 1997; Carlsmith, 2006; Keller et al., 2010; Gerber and Jackson, 2013; Twardawski et al., 2020). The results of this study also confirmed that the weighting of retribution in sentencing decisions in serious cases can be somewhat predominant but did not support retribution as the only justification. Since retribution is the default approach (Carlsmith, 2006), it is more likely to be weighted, but that does not negate utilitarian justification from consideration. In fact, even the severe-damage scenario, where the child was killed by the defendant who should have protected him, did not result in retribution alone. The finding that general deterrence was the second-highest weighted in both scenarios clearly indicates a Japanese tendency to emphasize general deterrence (Gollwitzer and Bücklein, 2007; Jiang et al., 2010a), suggesting that this measure could also reflect cultural characteristics, despite the similarity between Japan and the West in terms of the dominance of retribution for serious crimes. The consistent preference for general deterrence is indirect evidence of Japanese perceptions of punishment, such as a society that is more likely to maintain cohesion through applying severe punishments such as the death penalty (Johnson, 2020) and a tight culture with a low tolerance for deviant behavior (Gelfand et al., 2011). This finding is consistent with existing research results stating that punishment is motivated by both retribution and utilitarianism (Crockett et al., 2014) and that the mixed justification approach is most supported (Spiranovic et al., 2012). In addition, this study demonstrated the specific ratio of the mixed justification. Once the specific ratio is clarified, the difference in the weighting of justification might explain why the severity of the sentence varies depending on the cases and the judges.

Advantages Over a Likert Scale

Almost all the scores on a Likert Scale will lie within the middle to high range (i.e., acquiescence-response bias; Krosnick, 1999),

which makes it difficult to determine the importance of each item. Previous studies that examined the relationship between punishment and justification have been plagued by this problem. The measure developed in this study makes it possible to compare the four weightings. It is possible to test whether the factor loadings are different by using a Likert Scale, for example, by performing invariance tests between each scenario. Importantly, however, this ratio-type measure is effective even when justifications are conflicting. In this study, we manipulated damage levels and rehabilitation potential; thus, various combinations can emerge as there are multiple factors affecting justifications and sentencing decisions. Depending on the combination, it can become difficult to analyze loadings using Likert scale scores. For example, if a young person with multiple prior convictions commits a burglary, or an adult with no prior convictions commits a serious assault, the Likert Scale would yield similar scores for both retribution and rehabilitation (Spiranovic et al., 2012). As a result, both may appear to have been given equal weightage. However, in reality, people will always have to choose between maintaining the emphasis on retribution (as the default) or believing in the rehabilitative potential of the offender and emphasizing rehabilitation. The present measure can determine which aspect is given more weighting and loading during sentencing. Furthermore, it would be useful to examine not only individual cases but also attitudes toward the judicial system. In a study that examined the correlation between the death penalty and justification using a Likert Scale, it was found that the higher the weighting for retribution and general deterrence, the higher the support for the death penalty in the United States, Japan, and China (Jiang et al., 2010b). For such a study on the justice system, the present scale may provide a clearer picture of how people's attitudes are determined based on any hybrid structure of justification.

Prediction of Sentencing Decisions

The results of the multiple regression analysis showed that justification, which was the main factor in each scenario, did not predict the sentence severity; in the severe-damage scenario, the higher the ratio of incapacitation to rehabilitation, the lighter the sentence, while retribution did not predict the sentence. In the moderate-damage scenario, rehabilitation did not predict sentence severity. At first glance, these results seem to contradict the predictions. However, they may be rather consistent considering that retribution is considered the default approach (Carlsmith, 2006). In the severe-damage scenario, most participants might have been unable to deviate from the tendency to opt for retribution by default and thus increased the severity of their sentences, while some exceptional participants who emphasized rehabilitation and incapacitation could have lowered the severity of their sentences. The negative impact of incapacitation, which was not focused on in this study, on sentencing in the severe-damage scenario suggested that the risk of having another child and abusing that child is so low that a longer sentence may not be necessary. Even in the moderate-damage scenario, a certain weight was placed on retribution (21.5%), leading to a competition between rehabilitation and retribution and neither individually predicting the sentence.

Thus, in both scenarios, the extent to which the weight of other justifications can be increased against the default weight of retribution was crucial to predict the sentence.

Implications

Empirical data using this new measure as a "litmus test" can be applied to trial procedures. A test similar to the present study can identify jurors with extremely biased justifications in the selection process and detect the impact of specific evidence (e.g., gruesome evidence or victim impact statements) on jurors from changes in the justification ratio. The more jurors can visualize the balance that individuals place on justification, the easier it will be for them to work toward a consensus. If the balance between retribution and utilitarian justification is important in sentencing (as in many countries), this measure can help examine whether the public is actually making judgments in accordance with this principle and suggest necessary improvements. Recently, some studies have measured physiological indicators and psychological benefits to victims to understand restorative justice, which aims to repair the relationship between victims and offenders (e.g., Lloyd and Borrill, 2020; Witvliet et al., 2020). While there are some studies (e.g., Daly, 2002) that theoretically compare the similarity of restorative justice with retribution, to understand the concept of restorative justice, a new angle should be to examine the proportion of the hybrid structure of justification when restorative justice is supported more.

Limitations and Future Research

The current study has some limitations. To test the new measure, cases of child neglect were deliberately chosen over general violent crimes. This was considered to manipulate and make the severity and rehabilitative potential of the cases more pronounced. Therefore, further research needs to verify whether the measure can detect changes in other cases that fall within the purview of the criminal justice system as well. While they were combined as a set in this test, crime severity and predictors of rehabilitation (i.e., selfishness and poverty) could have been manipulated separately in a 2 × 2 design to examine the effect of rehabilitation on sentencing decisions. In addition, it is necessary to examine the possibility of detecting changes in justifications other than retribution and rehabilitation. Based on Carlsmith's (2006) discussion, the weighting of general deterrence may change with the manipulation of the frequency or detection rate of the crime, and that of incapacitation may change with the manipulation of the likelihood of defendant violence. Thus, it will be necessary to explore the applicability of the new measure by manipulating various factors separately and using a variety of cases. This measure should be widely tested outside Japan. The findings that the difference between retribution and rehabilitation was not significant in the moderate-damage scenario and that retribution did not predict the sentence may have been related to the cultural characteristics of the Japanese sample, which emphasizes general deterrence. It would be useful to understand the basic principles of the theory of punishment by examining how much the ratio of retribution, "the default," and the ranking of other justifications are common across cultures and countries. Furthermore, with regards to collating responses, it may be more

effective to use the method of dragging and moving a slider bar rather than the approach of entering numerical values as used in this study. The Summation Model (Hollands and Spence, 1998) was followed to implement the theoretical concept of the four hybrid constructs in the new measure. If the slider bar is a better fit to this model, which we assume may lead to an improvement, participants will be more likely to respond, and their responses will more closely reflect their sentencing justification. This study reflected high multicollinearity because there was only one dummy in the current study (i.e., precedent; see **Tables 2**, **3**). Accordingly, it would be effective to include multiple dummies to mitigate multicollinearity, which would be easier using a slider bar.

CONCLUSION

Sentencing justification among the public follows the hybrid structure of retribution, incapacitation, general deterrence, and rehabilitation. In the present study, a ratio-type measure was developed to access this structure, and its usefulness was tested on a single type of crime. The study succeeded in detecting changes in the weighting of justification, which was previously not detected by the existing form of assessment involving the Likert Scale. In addition to the finding of previous studies that retribution is the most important justification in sentencing decisions, the present study found that retribution is not the only justification and that other justifications are also considered—although retribution is more likely to be weighted as the default.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://osf.io/n3a6s/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethical Review Committee for Behavioral Sciences, Graduate School of Human Sciences, Osaka University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

EW conducted the material preparation and data collection, performed the analysis, and wrote the first draft of the manuscript. EW and TI conducted the analysis and developed the discussion. TI and TW commented on previous versions of the manuscript and made modifications. All authors contributed to the study conception and design, read and approved the final manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.761536/full#supplementary-material

REFERENCES

Alter, A. L., Kernochan, J., and Darley, J. M. (2007). Transgression wrongfulness outweighs its harmfulness as a determinant of sentence severity. *Law Hum. Behav.* 31, 319–335. doi: 10.1007/s10979-006-9060-x

Andreescu, V., and Hughes, T. (2020). Public opinion and the death penalty in Japan. *Punishm. Soc.* 22, 573–595. doi: 10.1177/1462474520915572

Berryessa, C. M. (2018). The effects of psychiatric and "biological" labels on lay sentencing and punishment decisions. *J. Exp. Criminol.* 14, 241–256. doi: 10.1007/s11292-018-9322-x

Berryessa, C. M. (2021). The potential influence of criminological rationales in considering childhood abuse as mitigating to sentencing. *Child Abuse Negl.* 111:104818. doi: 10.1016/j.chiabu.2020.104818

Brubacher, M. R. (2019). Third-party views of incarceration: justice, desistance, and offender reintegration. *Psychiatr. Psychol. Law* 26, 693–705. doi: 10.1080/13218719.2019.1618754

Cañadas, E., Rodríguez-Bailón, R., and Lupiáñez, J. (2015). The effect of social categorization on trust decisions in a trust game paradigm. *Front. Psychol.* 6:1568. doi: 10.3389/fpsyg.2015.01568

Carlsmith, K. M. (2006). The roles of retribution and utility in determining punishment. *J. Exp. Soc. Psychol.* 42, 437–451. doi: 10.1016/j.jesp.2005.06.007

Carlsmith, K. M. (2008). On justifying punishment: the discrepancy between words and actions. *Soc. Justice Res.* 21, 119–137.

Carlsmith, K. M., and Darley, J. M. (2008). Psychological aspects of retributive justice. *Adv. Exp. Soc. Psychol.* 40, 193–236. doi: 10.1016/S0065-2601(07)00004-4

Carlsmith, K. M., Darley, J. M., and Robinson, P. H. (2002). Why do we punish? Deterrence and just deserts as motives for punishment. *J. Pers. Soc. Psychol.* 83, 284–299. doi: 10.1037/0022-3514.83.2.284

Cohen, J. (1988). "Multiple Regression and Correlation Analysis," in *Statistical Power Analysis for the Behavioral Sciences*, ed. J. Cohen (Hillsdale, NJ: Lawrence Erlbaum Associates), 407–465.

Cotton, M. (2000). Back with a vengeance: the resilience of retribution as an articulated purpose of criminal punishment. *Am. Crim. Law Rev.* 37, 1313–1362.

Crockett, M. J., Özdemir, Y., and Fehr, E. (2014). The value of vengeance and the demand for deterrence. *J. Exp. Psychol. Gen.* 143, 2279–2286. doi: 10.1037/xge0000018

Daly, K. (2002). Restorative justice: the real story. *Punishm. Soc.* 4, 55–79. doi: 10.1177/14624740222228464

Exum, J. J. (2017). Should death be so different: sentencing purposes and capital jury decisions in an era of smart on crime sentencing reform. *Ark. Law Rev.* 70:227.

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., et al. (2011). Differences between tight and loose cultures: a 33-nation study. *Science* 332, 1100–1104. doi: 10.1126/science.1197754

Gerber, M. M., and Jackson, J. (2013). Retribution as revenge and retribution as just deserts. *Soc. Justice Res.* 26, 61–80. doi: 10.1007/s11211-012-0174-7

Goldman, A. H. (1982). Toward a new theory of punishment. *Law Philos.* 1, 57–76. doi: 10.1007/BF00143146

Gollwitzer, M., and Bücklein, K. (2007). Are "we" more punitive than "me"? Self-construal styles, justice-related attitudes, and punitive judgments. *Soc. Justice. Res.* 20, 457–478. doi: 10.1007/s11211-007-0051-y

Gromet, D. M., and Darley, J. M. (2009). Punishment and beyond: achieving justice through the satisfaction of multiple goals. *Law Soc. Rev.* 43, 1–38. doi: 10.1111/j.1540-5893.2009.00365.x

Hollands, J. G., and Spence, I. (1998). Judging proportion with graphs: the summation model. *Appl. Cogn. Psychol.* 12, 173–190. doi: 10.1002/(SICI)1099-0720(199804)12:2<173::AID-ACP499<3.0.CO;2-K

Hoskins, Z. (2020). "Hybrid theories of punishment," in *The Routledge Handbook of the Philosophy and Science of Punishment*, eds B. Waller, E. Shaw, and F. Focquaert (New York, NY: Routledge), 37–48. doi: 10.4324/9780429507212

Jiang, S., Pilot, R., and Saito, T. (2010b). Why Japanese support the death penalty? *Int. Crim. Justice Rev.* 20, 302–316. doi: 10.1177/1057567710373276

Jiang, S., Lambert, E. G., Wang, J., Saito, T., and Pilot, R. (2010a). Death penalty views in China, Japan and the US: an empirical comparison. *J. Crim. Justice* 38, 862–869. doi: 10.1016/j.jcrimjus.2010.06.001

Johnson, D. T. (2020). "The death penalty and democracy," in *The Culture of Capital Punishment in Japan. Palgrave Advances in Criminology and Criminal Justice in Asia*, ed. D. T. Johnson (Cham: Palgrave Pivot). doi: 10.1007/978-3-030-32086-7_6

Keller, L. B., Oswald, M. E., Stucki, I., and Gollwitzer, M. (2010). A closer look at an eye for an eye: laypersons' punishment decisions are primarily driven by retributive motives. *Soc. Justice Res.* 23, 99–116. doi: 10.1007/s11211-010-0113-4

Kita, M., and Johnson, D. T. (2014). Framing capital punishment in Japan: avoidance, ambivalence, and atonement. *Asian J. Criminol.* 9, 221–240. doi: 10.1007/s11417-014-9189-3

Kitayama, S., Park, H., Sevincer, A. T., Karasawa, M., and Uskul, A. K. (2009). A cultural task analysis of implicit independence: comparing North America, Western Europe, and East Asia. *J. Pers. Soc. Psychol.* 97, 236–255. doi: 10.1037/a0015999

Krosnick, J. A. (1999). Survey research. *Annu. Rev. Psychol.* 50, 537–567. doi: 10.1146/annurev.psych.50.1.537

Lloyd, A., and Borrill, J. (2020). Examining the effectiveness of restorative justice in reducing victims' post-traumatic stress. *Psychol. Inj. Law* 13, 77–89. doi: 10.1007/s12207-019-09363-9

McCorkle, R. C. (1993). Research note: punish and rehabilitate? Public attitudes toward six common crimes. *Crime Delinq.* 39, 240–252. doi: 10.1177/0011128793039002008

McFatter, R. M. (1978). Sentencing strategies and justice: effects of punishment philosophy on sentencing decisions. *J. Pers. Soc. Psychol.* 36, 1490–1500. doi: 10.1037/0022-3514.36.12.1490

McKee, I. R., and Feather, N. T. (2008). Revenge, retribution, and values: social attitudes and punitive sentencing. *Soc. Justice Res.* 21:138. doi: 10.1007/s11211-008-0066-z

McMunigal, K. C. (1998). Desert, utility, and minimum contacts: toward a mixed theory of personal jurisdiction. *Yale Law J.* 108, 189–235. doi: 10.2307/797473

Nagin, D. S. (1998). "Deterrence and incapacitation," in *The Handbook of Crime and Punishment*, ed. M. Tonry (New York, NY: Oxford University Press), 345–368.

Niang, A., Leclerc, C., and Testé, B. (2020). When punishment goals moderate and mediate the effect of clinical reports on the recidivism risk on prison sentences. *Psychiatry Psychol. Law* 1–15. doi: 10.1080/13218719.2020.1805811

Okimoto, T. G., Wenzel, M., and Feather, N. T. (2009). Beyond retribution: conceptualizing restorative justice and exploring its determinants. *Soc. Justice Res.* 22, 156–180. doi: 10.1007/s11211-009-0092-5

Orth, U. (2003). Punishment goals of crime victims. *Law Hum. Behav.* 27, 173–186. doi: 10.1023/a:1022547213760

Oswald, M. E., Hupfeld, J., Klug, S. C., and Gabriel, U. (2002). Lay-perspectives on criminal deviance, goals of punishment, and punitivity. *Soc. Justice Res.* 15, 85–98. doi: 10.1023/A:1019928721720

O'Toole, M. J., and Fondacaro, M. R. (2017). When school-shooting media fuels a retributive public: an examination of psychological mediators. *Youth Violence Juv. Justice* 15, 154–171. doi: 10.1177/1541204015616664

Roberts, J. V., and Gebotys, R. J. (1989). The purposes of sentencing: public support for competing aims. *Behav. Sci. Law* 7, 387–402. doi: 10.1002/bsl.2370070308

Robinson, P. H. (1987). Hybrid principles for the distribution of criminal sanctions. *Northwest. Univ. Law Rev.* 82, 19–42.

Robinson, P. H., and Darley, J. M. (1995). *Justice, Liability, and Blame: Community Views and the Criminal Law*. New York, NY: Routledge. doi: 10.4324/9780429039812

Rucker, D. D., Polifroni, M., Tetlock, P. E., and Scott, A. L. (2004). On the assignment of punishment: the impact of general-societal threat and the moderating role of severity. *Pers. Soc. Psychol. Bull.* 30, 673–684. doi: 10.1177/0146167203262849

Shimizu, H. (2016). An introduction to the statistical free software HAD: suggestions to improve teaching, learning and practice data analysis. *J. Media Inf. Commun.* 1, 59–73.

Spiranovic, C. A., Roberts, L. D., Indermaur, D., Warner, K., Gelb, K., and Mackenzie, G. (2012). Public preferences for sentencing purposes: what difference does offender age, criminal history and offence type make? *Criminol. Crim. Justice* 12, 289–306. doi: 10.1177/1748895811431847

Takikawa, K. (2019). Is child abuse increasing? *Keishin Res. J.* 3, 1–8. doi: 10.24759/vetrdi.3.2_1

Templeton, L. J., and Hartnagel, T. F. (2012). Causal attributions of crime and the public's sentencing goals. *Can. J. Criminol. Crim. Justice* 54, 45–65. doi: 10.3138/cjccj.2010.E.29

Twardawski, M., Tang, K. T., and Hilbig, B. E. (2020). Is it all about retribution? The flexibility of punishment goals. *Soc. Justice Res.* 33, 195–218. doi: 10.1007/s11211-020-00352-x

Vidmar, N., and Miller, D. T. (1980). Socialpsychological processes underlying attitudes toward legal punishment. *Law Soc. Rev.* 14, 565–602.

Watamura, E., Wakebe, T., and Maeda, T. (2011). Can jurors free themselves from retributive objectives? *Psychol. Stud.* 56, 232–240. doi: 10.1007/s12646-011-0079-9

Weiner, B., Graham, S., and Reyna, C. (1997). An attributional examination of retributive versus utilitarian philosophies of punishment. *Soc. Justice Res.* 10, 431–452. doi: 10.1007/BF02683293

Witvliet, C. V., Root Luna, L., Worthington, E. L. Jr., and Tsang, J. A. (2020). Apology and restitution: the psychophysiology of forgiveness after accountable relational repair responses. *Front. Psychol.* 11:284. doi: 10.3389/fpsyg.2020.00284

Yamamoto, S., and Maeder, E. M. (2021). What's in the box? Punishment and insanity in the Canadian jury deliberation room. *Front. Psychol.* 12:2442. doi: 10.3389/fpsyg.2021.689128

Police Decision-Making in the Absence of Evidence-Based Guidelines: Assessment of Alcohol-Intoxicated Eyewitnesses

Daniel Pettersson, Magnus Bergquist and Angelica V. Hagsand*

Department of Psychology, University of Gothenburg, Gothenburg, Sweden

Regarding police procedures with alcohol-intoxicated witnesses, Swedish police officers have previously reported inconsistent and subjective decisions when interviewing these potentially vulnerable witnesses. Most officers have also highlighted the need for national policy guidelines aiding in conducting investigative interviews with intoxicated witnesses. The aims of the two studies presented here were to investigate whether (1) police officers' inconsistent interview decisions are attributable to a lack of research-based knowledge; (2) their decision to interview, as well as their perceptions of the witnesses' credibility could be influenced by scientific research; and (3) police officers decision-making and perceptions of witness credibility are biased by pre-existing social norms. In two separate randomized online experiments, police professionals and recruits (Study 1, $N=43$; Study 2, $N=214$) watched a recorded fictive witness interview to which they were asked to rate the probability of interviewing the witness, the witness' credibility, and to estimate the witness' level of intoxication. Results showed that interview probability and perceived witness credibility were affected by witness intoxication level. While it cannot be stated definitely from the present research, these findings provided indications that police officers and recruits lacked research-based knowledge. Results also showed that interview probability, but not perceptions of credibility, was influenced by a research-based message. In line with research, interview probability for the most intoxicated witness increased after reading the message. Unexpectedly, neither interview probability nor witness credibility was affected by social norms. The current findings added to the legal psychology literature by showing that a breath alcohol concentration (BrAC) as low as .04% was enough for police officers and recruits to consider intoxicated witnesses less credible than sober witnesses. Findings also indicated that, despite the lower credibility assessment, police may have some understanding that these witnesses can be interviewed at low intoxication levels (i.e., around .04%). However, this willingness to interview intoxicated witnesses ceased at a BrAC lower than the levels where research has found intoxicated witnesses as reliable as sober witnesses (i.e., BrAC<.10%). Future directions for research and policy development as well as theoretical and practical implications of the present findings are discussed.

Keywords: police, decision-making, social norms, intoxication, witness, intervention

INTRODUCTION

In the past, the alcohol and memory literature have often found evidence of detrimental memory impairments being caused by alcohol-intoxication (e.g., Parker et al., 1976; Mintzer, 2007). This provides a rationale for the prevalent perception among legal practitioners (Kassin et al., 2001; Evans et al., 2009; Crossland et al., 2018; Hagsand et al., 2021, 2022; Monds et al., 2021a) and lay people (Evans and Schreiber Compo, 2010; Monds et al., 2021b) that intoxicated witnesses are less credible than sober witnesses. However, the effects of memory impairments caused by alcohol-intoxication have often not been replicated within the applied legal context of eyewitness memory (see Altman et al., 2019, for a review). When assessing intoxicated eyewitnesses' memory, research has distinguished between the completeness of recall (i.e., total number of details recalled) and the accuracy of recall (i.e., number of correct details recalled; Schreiber Compo et al., 2019). A recent meta-analysis observed a dose-response relationship (i.e., memory impairments linearly increased with alcohol-intoxication) for completeness of recall but observed little effect of alcohol on the accuracy of intoxicated witnesses' recall (see Jores et al., 2019), indicating the reliability of this witness group. Moreover, the effects of alcohol on eyewitness memory are nuanced and depending upon the intoxication level (see Altman et al., 2018 for an experimental bar study). Recent research show that intoxicated witness accounts often are reliable when breath alcohol concentration (BrAC) is approximately below .10%, with memory impairment increasing with higher levels of intoxication (see Altman et al., 2019, for a review). When interviews are conducted immediately after the witnessing of a crime, low to moderately intoxicated (BrAC < .10%) individuals reportedly are no less susceptible to suggestive leading questions than sober individuals (Mindthoff et al., 2021), and intoxicated persons also give complete and accurate statements (e.g., Hagsand et al., 2017; Mindthoff et al., 2019). In contrast, when the interview was postponed, both sober and intoxicated witnesses (e.g., Hagsand et al., 2017; Hildebrand Karlén et al., 2017; Schreiber Compo et al., 2017; Evans et al., 2018) gave less accurate information. Complete and accurate statements of eyewitnesses can be of central importance in criminal investigations (Kebbell and Milne, 1998), where criminal charges as well as subsequent court convictions often rely on evidence gathered from police interviews (Howe and Knott, 2015). Despite the known importance of obtaining information from witness interviews, police officers have reported inconsistent decisions to interview intoxicated witnesses and furthermore report varying procedures of engaging with them (i.e., waiting until witness is sober vs. interviewing immediately; Evans et al., 2009; Crossland et al., 2018; Monds et al., 2021a; Hagsand et al., 2022).

One way to further understand police officers' inconsistent decisions to interview intoxicated witnesses has been studied through police surveys which were conducted both in Sweden and internationally (Evans et al., 2009; Crossland et al., 2018; Monds et al., 2021a; Hagsand et al., 2022). From these studies, several contributing factors to the officers' inconsistent decisions can be derived. Firstly, many police officers perceived intoxicated witnesses as less credible than sober witnesses, which may have rendered them reluctant to interview these individuals (Evans et al., 2009; Crossland et al., 2018; Monds et al., 2021a; Hagsand et al., 2022). As the research described above has proposed otherwise, these findings suggested that police officers perhaps lacked research-based knowledge, highlighting the need for guidelines regarding intoxicated witnesses' ability to accurately recall criminal events (Hagsand et al., 2022).

Secondly, procedural differences between officers as well as departments were evident across the studies mentioned above. The majority (69.9%) of UK police officers reported having different procedures for conduct with intoxicated and sober witnesses (Crossland et al., 2018), whereas approximately 40% of officers in Sweden (Hagsand et al., 2022) and 40% of officers in the US (Evans et al., 2009) reported different departmental procedures for these witness groups. In Sweden, the relatively low percentage of officers reporting different procedures could reflect the absence of an official research-based national policy that could provide guidelines with regards to proper conduct when encountering intoxicated witnesses.

A third contributing factor to the inconsistent interview decisions could be police officers self-reported reliance on observational methods to assess alcohol-intoxication (Evans et al., 2009; Crossland et al., 2018; Hagsand et al., 2022). Legality issues have commonly been cited as the reason for underutilized objective measurements (i.e., portable breathalyzer) when assessing witness intoxication levels. Instead of using objective measures of intoxication, officers reportedly make subjective assessments using observational methods (e.g., alcohol-odor, judging the behavior of the witness, using the standard field sobriety test, and conversational tests; Evans et al., 2009; Crossland et al., 2018; Hagsand et al., 2022). This raises the question of the accuracy of police officers' judgments. A field study conducted in the United States reported a 98% accuracy in detecting alcohol-intoxication ≥.08% but showed a noticeable decrease to 71% accuracy for BrAC levels <.08% (Stuster, 2006). Another US study found that police officers, viewing a video clip, could not accurately decide if a person had even consumed alcohol until BrAC reached >.15% (Brick and Carpenter, 2001). The authors suggested that the visual stimulus used in their study left out important cues (e.g., alcohol-odor). Furthermore, a review on the use of observational methods for assessing alcohol-intoxication (e.g., alcohol-odor, standard field sobriety test, impaired walking, distorted speech, and finger to nose) concluded that all techniques were unsubstantiated (Rubenzer, 2011). These studies implied that while it is generally difficult to assess alcohol-intoxication through observation, it may be especially difficulty at low to moderate intoxication levels (i.e., BrAC <.10%; Stuster, 2006; Rubenzer, 2011).

Though not apparent in every study, these three contributing factors were all evident within the Swedish survey study by Hagsand et al. (2022), which clearly highlighted the necessity of disseminating research-based knowledge among Swedish law enforcement. Since 2019, police officers in the United Kingdom operate under research-based guidelines when they encounter intoxicated witnesses (College of Policing, 2019). Collaborations between researchers in Sweden and the

Swedish Police Authority have begun, and the development of similar guidelines for intoxicated witnesses is underway (see Hagsand et al., 2020). However, this does not negate the fact that the police departments within Sweden and other countries currently still use inadequate methods of assessment when handling intoxicated witnesses. Indicative of this, 73% of police officers in Sweden (Hagsand et al., 2022), 74% of US officers (Evans et al., 2009), and 27% of officers in the United Kingdom (Crossland et al., 2018) reported that the decision to interview alcohol-intoxicated witnesses depended on the situation. For example, most Swedish police officers reported that they would consider the degree of the witness' alcohol-intoxication before conducting an interview (Hagsand et al., 2022). Where the absence of guidelines, unapplied research-based knowledge and unsubstantiated methods to assess intoxication contribute to police' behavior, a fourth factor may be the influence of social norms, generally applied in uncertain situations.

Social norms are informal rules shared by members of a particular social group and inform group members of behaviors and decisions (Cialdini and Trost, 1998; Legros and Cislaghi, 2020). These norms have been divided into descriptive and injunctive social norms, the former representing typical behaviors, the latter representing behaviors that are socially acceptable (Cialdini et al., 1990; Cialdini, 2012). A plethora of reasons exist for people's compliance to social norms, some of which are out of a desire to hold accurate beliefs about the world, to maintain favorable concept of the self and with others (Cialdini and Goldstein, 2004), out of practicality (Anderson and Dunning, 2014), in anticipation of positive or negative social sanctions (Legros and Cislaghi, 2020), and importantly, as guidance in uncertain situations (Bell and Cox, 2015). The *focus theory* states that social norms must be focal in attention (i.e., made salient) if they are to effectively induce compliance (Cialdini et al., 1990). When combined, descriptive and injunctive norms can amplify the effect of normative influence on behavior (Miller and Prentice, 2016). However, when incongruous, the more focal norm would elicit compliance (Cialdini, 2012). Using social norms as a fourth contributing factor is of relevance to the current study, as these have shown to have a widespread and well-documented influence on behavior (e.g., Cialdini and Trost, 1998; Crano, 2000; Legros and Cislaghi, 2020) in contexts related to legal psychology, such as petty crimes (Keizer et al., 2008; Keuschnigg and Wolbring, 2015) and thievery (Cialdini et al., 2006). Descriptive social norms in particular have been found to influence police officer's decision-making, where the knowledge of other officers intervening in a domestic violent situation through arrest increased the likelihood of police officers to do the same (Baldry and Pagliaro, 2014). Important to note is, however, that only those identifying strongly with their occupation were affected by the descriptive message.

This strong identification is often seen as a unique workplace culture, characterized by a strong sense of solidarity (Wieslander, 2019) and featured by external threats incongruent with most other workplaces (Marier and Moule, 2019). Because of this, group socialization processes, in which new recruits are encouraged to quickly adapt the tacit rules of the game (Gatto and Dambrun, 2012), have been argued to be especially strong within the Swedish Police Authority (Wieslander, 2019). These processes of socialization begin while recruits are still attending the Swedish Police Academy (Lander, 2013). As such, police culture develops professional and recruits alike. The information listed above would suggest that Swedish police officers, as well as the police officers within the other studies, could have been influenced by social norms for policing when assessing whether to interview intoxicated witnesses. This is an area within the field of legal psychology that has previously not been studied in light of police interviews with intoxicated witnesses and therefore represents a novel combination between the areas of legal psychology and social psychology.

The aims of the two studies presented here were to investigate whether (1) police officers' inconsistent interview decisions are attributable to a lack of research-based knowledge; (2) their decision to interview, as well as their perceptions of the witnesses' credibility could be influenced by scientific research; and (3) police officers decision-making and perceptions of witness credibility are biased by pre-existing social norms.

STUDY 1

The purpose of Study 1 was to pilot-test the experimental procedure in a small sample of police officers and police recruits. An online experimental study using a mixed design was conducted. While there was no explicit test of participants research-based knowledge, the first aim was investigated by combining participants estimates of witness intoxication level, their stated interview probability, and their perception of witness credibility. The second aim was investigated by having participants read a short research-based message before responding to the dependent measures. The third aim was investigated using both a short descriptive normative message and by measuring levels of pre-existing descriptive and injunctive norms before they responded to the dependent measures.

Materials and Methods
Participants
Professional police officers were recruited by invitations sent to all four regional police departments across Sweden. Police recruits were obtained by invitation sent to universities which managed police education in Sweden. A total of 84 participants clicked the invitation link. Some participants failed the attention checks ($n = 11$) or did not complete important study variables ($n = 30$) and were excluded. The final sample consisted of $N = 43$ participants. There was $n = 17$ (39.5%) police officers aged between 24 and 59 years old with almost an even split between men (52.9%) and women (47.1%). There was also $n = 26$ (60.5%) police recruits aged 19 to 45 years old, a majority of these were men (69%, women 31%).

Materials
The study was conducted online using the Qualtrics XM Platform. When starting the questionnaire, participants first

viewed a short film (videos were inspired by Hirn Mueller et al., 2015, with the addition of having intoxicated witnesses) depicting a fictious eyewitness police interview with two actors (female witness and male interviewer) seated at a table in an interview room. In the scripted scenario, the woman had visited a bar with a friend, where she witnessed someone being stabbed in the stomach. The film depicted how the interviewer tried to elicit information from the witness regarding the event. There were three versions, each depicting the witness as either sober, moderately intoxicated, or highly intoxicated. The actor portraying the witness was given detailed instructions on how to simulate intoxication at the targeted BrAC levels (.00, .10, and .15%, respectively, for sober, moderate, and high intoxication levels), and the instructions were based on previous research concerning which behavior is present at different intoxication levels (Söderpalm, 2011). Examples of the instructions were as: the sober witness (0 alcoholic drinks) should not get distracted during the interview and should answer in a polite and straightforward way; the moderately intoxicated witness (approximately four alcoholic drinks) may be less focused, more easily distracted, with heightened emotions yet still in control; the highly intoxicated witness (approximately 8–12 alcoholic drinks) would be easily distracted, have somewhat slurred speech, be repetitive, act nauseous, and so forth. The only difference between the three films was the degree of witness alcohol-intoxication and each film was approximately 1 min in length.

Validation of the Stimulus Material

Estimates of witness intoxication were conducted first in a sample of psychology students from the University of Gothenburg, Sweden and additionally in a sample of Swedish police officers and recruits. The BrAC levels depicted in the films were .00% (sober), .10% (moderate), and .15% (high). Both university students and police officers were asked to estimate witness intoxication level on a 7-point Likert response format. Among both the university students ($N = 102$) and the police officers ($N = 114$), a significant effect of witness intoxication on estimates of intoxication was found, $F_{student} (2,99) = 34.87$, $p < .001$, $\eta_p^2 = .41$ and $F_{officer} (2,111) = 35.73$, $p < .001$, $\eta_p^2 = .39$. With university students, planned simple contrasts found that the highly intoxicated witness ($M = 3.73$, $SD = 1.15$) was estimated to be significantly more intoxicated compared to the moderately intoxicated ($M = 2.51$, $SD = .79$) and sober witnesses ($M = 1.87$, $SD = .80$, $p < .001$), and with a significant difference between the sober and the moderately intoxicated witnesses ($p = .005$). Similar effects were mirrored with police officers ($M_{high} = 4.45$, $SD_{high} = 1.24$; $M_{moderate} = 2.81$, $SD_{moderate} = 1.18$; $M_{sober} = 2.17$, $SD_{sober} = 1.12$), with a significant difference between the highly intoxicated witness and the sober and moderately intoxicated witnesses ($ps < .001$), and a significant difference between the sober and moderately intoxicated witnesses ($p = .021$). University students ($N = 98$) were also asked to estimate the BrAC level of the witness. A significant effect of witness intoxication on estimates of BrAC was found $F(2,95) = 9.50$, $p < .001$, $\eta_p^2 = .17$. Planned simple contrasts found that the highly intoxicated witness ($M = .17$, $SD = .10$) was estimated to be significantly more intoxicated compared with the moderately intoxicated ($M = .10$, $SD = .07$) and sober witnesses ($M = .08$, $SD = .09$, $p < .001$). Here, however, there was a non-significant difference between the sober and the moderately intoxicated witnesses ($p = .566$).

Procedure

Once participants gave their consent, they were given a battery of background questions. Thereafter, participants were randomized to one of four between-subject conditions: (1) control, (2) measured social norms, (3) induced descriptive social norm, and (4) research-based message. In the measured social norms condition, participants were asked about the extent to which they believed their colleagues would interview the witness and the extent to which their colleagues would approve if the participants themselves interviewed the witness. A 7-point Likert response format ranging from 1 = none/disapprove to 7 = all/approve was used to capture responses. In the induced descriptive norm condition, participants read that police officers considered intoxicated witnesses to be credible if BrAC is less than .10% and if open-ended questions were used.[1] In the research-based message condition, participants read that research supported the view that intoxicated witnesses are credible if BrAC is less than .10% and if open-ended questions are used. All participants viewed all three films (depicting all intoxication levels) but in a counter-balanced order. All dependent variables were measured repeatedly after each film. Participants were asked (1) How credible did you find the witness? (2) Would you interview the witness? and (3) How intoxicated do you think the witness were? A 7-point Likert response format ranging from 1 = Not at all/not likely at all/completely sober to 7 = Completely/most likely/extremely intoxicated was used to capture responses. At the end of the study, attention checks were made.

Hypotheses

Based on previous research (Evans et al., 2009; Crossland et al., 2018; Monds et al., 2021a; Hagsand et al., 2022), we expected interview probability (H1a) and perceived witness credibility (H1b) to decrease as witness intoxication level increased. Due to potential lack of scientific knowledge among police officers, as suggested previously (e.g., Hagsand et al., 2022), we expected to find an effect of the research-based message on both interview probability (H2a) and perceived witness credibility (H2b). Based on the extensive research on social normative influence (e.g., Baldry and Pagliaro, 2014; Miller and Prentice, 2016), we expected to find an effect of the induced descriptive norm on both interview probability (H3a) and perceived witness credibility (H3b). We also expected to find an effect of activating pre-existing norms on both interview probability (H4a) and perceived witness credibility (H4b).

---

[1]Open-ended question format is an interview technique in which the interviewee is encouraged to provide a free-form answer. In contrast, a closed question often requires a yes or no answer or impose some other limit on possible answers. This was a design element used in the research from which the film material originated and was not pursued further in the present research.

TABLE 1 | Means (SD) for participants witness intoxication estimates in Study 1.

Witness intoxication	Information				Total
	Control *n* = 12	Social norm *n* = 12	Police norm *n* = 7	RBM *n* = 12	*N* = 43
Sober	2.08 (1.00)	2.33 (.89)	3.00 (1.00)	2.42 (1.51)	2.40 (1.14)
Moderate	2.42 (.90)	2.92 (.79)	2.86 (1.10)	2.50 (1.17)	2.65 (1.00)
High	5.17 (.72)	5.08 (.52)	5.29 (1.11)	4.83 (1.30)	5.07 (.91)

Participants were asked to estimate how intoxicated the witness were. Response format ranged from 1—completely sober to 7—extremely intoxicated. RBM, research-based message.

TABLE 2 | Means (SD) for interview probability in Study 1.

Witness intoxication	Information				Total
	Control *n* = 12	Social norm *n* = 12	Police norm *n* = 7	RBM *n* = 12	*N* = 43
Sober	6.42 (1.17)	6.83 (.58)	6.14 (1.07)	6.17 (1.27)	6.42 (1.10)
Moderate	6.75 (.62)	6.83 (.58)	6.43 (.98)	6.08 (1.34)	6.53 (.96)
High	4.50 (1.73)	6.08 (1.51)	4.86 (1.86)	5.25 (2.14)	5.21 (1.86)

Response format ranged from 1—not likely at all to 7—most likely. RBM, research-based message.

Results

Manipulation Check of Witness Intoxication

There was a significant main effect of witness intoxication on police officers' estimates of intoxication $F(2,78) = 153.13$, $p < .001$, $\eta_p^2 = .80$. Planned simple contrasts showed that participants considered the highly intoxicated witness to be significantly more intoxicated compared with the sober witness $F(1,39) = 222.56$, $p < .001$, $\eta_p^2 = .85$. There was no significant distinction between the sober and moderate witnesses $F(1,39) = 1.54$, $p = .222$, $\eta_p^2 = .04$. There was no main effect of information on intoxication estimate $F(3,39) = .67$, $p = .576$, $\eta_p^2 = .05$. Finally, there was a non-significant interaction effect of information and witness intoxication on intoxication estimate $F(6,78) = .87$, $p = .520$, $\eta_p^2 = .06$. Descriptive statistics are reported in **Table 1**. Results showed that participants estimated that the sober and moderately intoxicated witnesses were comparably intoxicated.

Main Analyses

For each dependent measure, a 4 (Information: control vs. measured norms vs. induced norm vs. research-based message) × 3 (Witness intoxication: sober vs. moderate vs. high) mixed design ANOVA's with repeated measures on the second factor were conducted.[2]

2Data availability: All SPSS data sets and outputs can be downloaded from the Open Science Framework (OSF). https://osf.io/qv8tn/?view_only=eb19536 2d8ab49308e824dc4743bb3af

Interview Probability

There was a significant main effect of witness intoxication on interview probability $F(2,78) = 30.64$, $p < .001$, $\eta_p^2 = .44$. Planned simple contrasts showed that participants were significantly less likely to interview the highly intoxicated witness compared with the sober witness $F(1,39) = 30.36$, $p < .001$, $\eta_p^2 = .44$; however, they made no significant distinction between the sober and moderate witnesses $F(1,39) = 1.90$, $p = .176$, $\eta_p^2 = .05$. This partially supported hypothesis 1a. There was a no main effect of information on interview probability $F(3,39) = 1.23$, $p = .313$, $\eta_p^2 = .09$. This contradicted hypotheses 2a, 3a, and 4a. Finally, there was no interaction between information and witness intoxication on interview probability $F(6,78) = 2.15$, $p = .056$, $\eta_p^2 = .14$. Descriptive statistics are reported in **Table 2**. Results showed that participants were just as likely to interview the sober witnesses as the moderately intoxicated witnesses.

Witness Credibility

There was a significant main effect of witness intoxication on witness credibility $F(2,78) = 35.97$, $p < .001$, $\eta_p^2 = .48$. Planned simple contrasts showed that the highly intoxicated witness was rated significantly less credible than the sober witness $F(1,39) = 37.84$, $p < .001$, $\eta_p^2 = .49$ and the moderately intoxicated witness $F(1,39) = 60.64$, $p < .001$, $\eta_p^2 = .61$. However, there was no significant distinction between the sober and moderately intoxicated witnesses $F(1,39) = .16$, $p = .691$, $\eta_p^2 = .00$. This partially supported hypothesis 1b. There was no significant main effect of information on witness credibility $F(3,39) = .47$, $p = .702$, $\eta_p^2 = .04$. This contradicted hypotheses 2b, 3b, and 4b. Finally, there was a non-significant effect of witness intoxication interaction on witness credibility $F(6,78) = .90$, $p = .502$, $\eta_p^2 = .06$. Descriptive statistics are reported in **Table 3**. Results showed that participants perceived the sober and moderately intoxicated witnesses as comparably credible.

Discussion

The first aim of the present research was to investigate whether the previously reported (e.g., Evans et al., 2009; Hagsand et al., 2022) inconsistent interview decisions could be attributable to a lack of research-based knowledge. In the present context, we would conclude knowledgeability under three conditions: (1) if estimates of witness intoxication differed between the sober and moderately intoxicated witnesses, (2) if interview probability were comparable for the sober and moderately intoxicated witnesses but differed for the highly intoxicated witness, and (3) if credibility ratings were comparable for the sober and moderately intoxicated witnesses but differed for the highly intoxicated witness. Although no explicit test of research-based knowledge was used, if these three conditions were met it would be very likely that participants possessed knowledge of scientific research. The results showed comparable interview probability, and perceived witness credibility, for the sober and moderately intoxicated witnesses. The results also revealed a decrease in interview probability and perceived witness credibility for the highly intoxicated witness. In line with conditions 2 and 3, these results indicated that police

TABLE 3 | Means (SD) for witness credibility in Study 1.

Witness intoxication	Information				Total
	Control n = 12	Social norm n = 12	Police norm n = 7	RBM n = 12	N = 43
Sober	5.25 (1.71)	5.25 (.62)	5.57 (1.13)	5.33 (1.37)	5.33 (1.25)
Moderate	5.25 (1.26)	5.00 (.60)	5.71 (.76)	5.67 (1.07)	5.37 (.98)
High	3.67 (1.37)	4.25 (.87)	4.43 (1.27)	4.25 (1.77)	4.12 (1.35)

Response format ranged from 1—not credible at all to 7—completely credible. RBM, research-based message.

officers' and recruits' decision and perception aligned with scientific research. However, contrasting condition 1, participants made similar intoxication estimates for the sober and moderately intoxicated witnesses. Therefore, the results remained inconclusive because there was no way to determine if participants treated the sober and moderate witnesses the same because they had pre-existing knowledge or because the degree of intoxication for these witnesses was considered similar.

The second aim of the present research was to investigate whether participants decision to interview, as well as their perceptions of the witnesses' credibility, could be influenced by scientific research. Results showed that the research-based message had no significant impact on participants' decision to interview the witnesses or on their perception of witness credibility. This indicated that a research-based message may not be a viable way to disseminate research findings among police officers. However, real-world policy guidelines like those used by police in the United Kingdom (College of Policing, 2019) are more extensive (i.e., a half page to one page) than a single sentence statement. It is possible that the short message used in Study 1 was insufficient to influence the participants. It is also possible that they already possessed this information and that the research-based message was not additionally helpful to them; however, results were inconclusive regarding participants knowledge base.

The third aim was to investigate whether police officer's decision-making and perceptions of witness credibility were biased by social norms. Neither the induced descriptive norm nor the activation of pre-existing social norms influenced the interview decision or the perception of witness credibility.

STUDY 2

In Study 1, it was not possible to determine if participants treated the sober and moderate witnesses the same because they had pre-existing knowledge or because they made similar intoxication estimates for these two witnesses. Thus, Study 1 remained inconclusive in terms of the first study aim. In Study 2, an attempt to untangle this issue was made by asking participants to estimate witness intoxication level (i.e., BrAC) rather than on a 7-point Likert response format. A related issue in Study 1 was that the approximately 1-min-long films may not have provided enough time for participants to observe the witness behaviors. Subtle but important mannerism changes between the sober and moderately

intoxicated witnesses could have been difficult to detect. This may have contributed to the comparable intoxication estimates for the sober and moderate witnesses. Therefore, Study 2 included longer films to provide ample time to observe the witnesses.

Regarding the second study aim (i.e., whether participants decision to interview, as well as their perceptions of the witnesses' credibility could be influenced by scientific research), Study 1 found no such prospect. It is possible that the short message in Study 1 was insufficient to influence participants. To strengthen the manipulation, a more extensive and real-world research-based message (see College of Policing, 2019, for actual UK guidelines) was used in Study 2.

In Study 1, social norms had no influence on participants' decision and perception. One issue was that Study 1 did not account for how much participants identified with the reference group (i.e., other police and recruits). A strong identification with the reference group has been associated with a greater effect of a descriptive norm message (Baldry and Pagliaro, 2014; Liu et al., 2019). *Social identity theory* (Tajfel and Turner, 1979) states that in certain social contexts, people consider their group identity as more salient than their individual identity (Ellemers and Haslam, 2012). Consequently, people would be more likely to conform to social norms when there is a strong association between the individual and the group. Such a strong bond could be expected among police officers and recruits (Marier and Moule, 2019; Wieslander, 2019). Therefore, Study 2 included a measure of identification with the police occupation to investigate if the lack of social normative effect in Study 1 was related to social identity.

Overall, Study 1 was intended as a minor pilot study and had a smaller sample size which consequently meant lower power. Study 2 represented an improvement over Study 1 in several ways. In addition to collecting an adequate number of participants, the design in Study 2 was simplified. Since neither norm manipulation had any impact in Study 1 and the aim of the present research was initially to investigate pre-existing norms, the induced social norms condition (which was piloted in Study 1) was removed. This further simplified the research design. Moreover, the longer films meant participants could experience survey fatigue and therefore, a between-subjects design was used so each participant viewed only one film. The null findings in Study 1 prompted a revision of the response format used. Study 1 included a 7-point Likert format, but in order to increase sensitivity, Study 2 included a 10-point format.

Materials and Methods
Participants

Police officers were recruited by invitation that was sent to all seven regional police departments across Sweden and *via* the national human resources department as well as personal contacts of the research team. In addition, all five universities which managed police education in Sweden was contacted *via* email and asked to forward an invitation to their police recruits. Finally, the invitation was also sent to police aspirants who underwent in-service training. A total of 336 people clicked the invitation link. Participants were excluded if they (a) did not consent (n = 8), (b) had participated in the pilot study

FIGURE 1 | Overview of study procedure in Qualtrics.
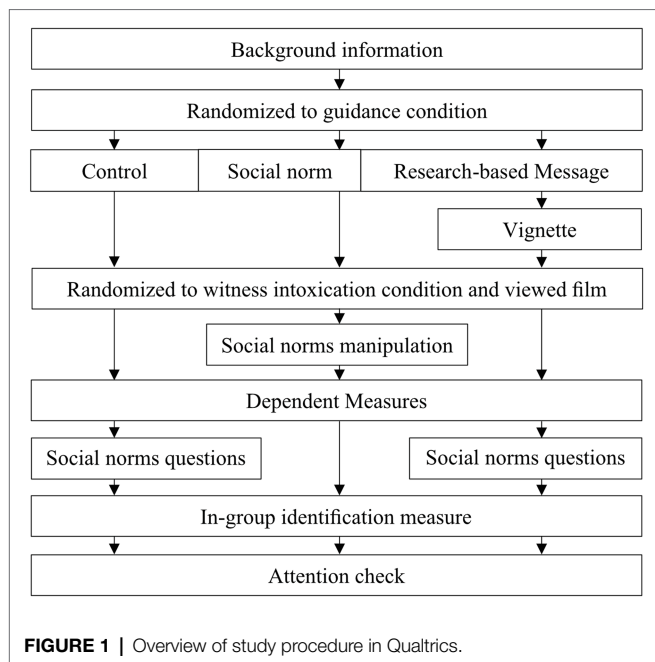
($n=2$), (c) failed the attention check ($n=2$), (d) did not complete the film viewing ($n=37$), or (e) had missing data on all dependent measures and could not be analyzed ($n=73$). Where data were available, attrition analyses showed that there was no significant gender difference between included and excluded participants ($p=.473$). Neither was there a significant difference in terms of how many police officers versus recruits were excluded ($p=.908$). Included and excluded police officers did not significantly differ in terms of experience working with witnesses ($p=.126$), and neither did recruits ($p=.336$). However, there was a significant mean difference in age [$t(199.93)=-2.33$, $p=.02$, Cohens $d=.28$]. Excluded participants ($M=35.59$, $SD=9.77$) were slightly younger than included participants ($M=38.56$, $SD=11.20$).

The final sample consisted of 214 participants. There was $n=152$ (71%) professional police officers, a majority of these were men (men 57.2%, women 42.1%, and other .7%), and the average age was 42 years ($SD=11.12$). Most (99.3%) professional police officers had experience interviewing witnesses ($M_{years}=13.16$, $SD=10.89$). All seven police regions in Sweden were represented in the sample (South 26.3%, West 20.4%, East 19.1%, Bergslagen 14.5%, Stockholm 12.5%, North 3.9%, Middle 3.3%). There was also $n=62$ (29%) police recruits, a majority of these were men (men 79%, women 21%), and the average age was 32 years ($SD=7.82$). Most recruits (74.1%) had been present for at least one witness interview. Most (80%) universities forwarded the invitation to their students (Linnaeus University; 30.6%, University of Borås; 29%, Södertörn University; 24.2%, and Malmö University; 16.1%).

Materials
To display more of the witness-interviewer interaction, the short films used in Study 1 were extended by editing together several films from the original set (inspired by Hirn Mueller et al., 2015)

to create longer versions. The three edited films varied in length with the sober film playing 3 min, 43 s, the moderately, and highly intoxicated films, 3:51 and 4:26, respectively. Differences in seconds between the films were the cause of the instructions to the actor playing the witness (e.g., telling the actor to make slower responses and be more easily distracted). The films were validated and pre-study analyses are reported under materials for Study 1.

In summary, the research-based message[3] stated that (1) level of intoxication greatly affects the extent of the memory impairments, (2) BrAC<.10% oftentimes does not affect witness memory, but in cases of negative effect, alcohol primarily affects the completeness of statements, and not the accuracy, (3) BrAC>.10% affects both completeness and accuracy, and (4) the most informative statements are obtained when witnesses are interviewed in close connection with the criminal event.

Inspired by previous research (Barreto and Ellemers, 2000; Baldry and Pagliaro, 2014), a short scale to measure social identity was constructed. It contained four propositional items (i.e., *being a member of the police is important to me, I feel like I am a part of the police, I feel good about being a part of the police, I feel the police occupation is the right fit for me*), presented in a 10-point Likert response format ranging from 1 = *completely disagree* to 10 = *completely agree*. A mean score across items was computed as a measure of identification with the police occupation (Cronbach's $\alpha=.84$).

Procedure
Study 2[4] was also conducted online using the Qualtrics XM Platform. Once they consented, participants were given a battery of background questions. They were then randomized to one of nine experimental conditions in a 3 (Information: control vs. social norm vs. research-based message) × 3 (Witness intoxication: sober vs. moderate vs. high) between-subjects experimental design (see **Figure 1**, for an overview of the study procedure in Qualtrics). Only participants in the research-based information condition read the research information after which participants in all conditions each saw one of the three films. After the film, only participants in the social norms condition were asked (in a 10-point Likert response format) the descriptive (i.e., *on a scale from 1 to 10, how many of your police colleagues or fellow police students would interview the witness?*) and the injunctive (i.e., *on a scale from 1 to 10, would your police colleagues or fellow police students approve/disapprove if you interviewed the witness?*) norm activation questions. All participants were then asked the dependent measures of how credible they found the witness and how probable it was that they would conduct an interview. Responses were captured on 10-point response formats which ranged from 1 = *not credible at all/not likely at all* to 10 = *most credible/very likely*. Participants were also asked to estimate the witness BrAC on a two decimal continuum which ranged from 0 to 4 and presented

---

[3]View the complete research-based message at Open Science Framework (OSF): https://osf.io/qv8tn/?view_only=eb195362d8ab49308e824dc4743bb3af
[4]The original Qualtrics questionnaire (in Swedish) used in Study 2 can be viewed at the Open Science Framework (OSF): https://osf.io/qv8tn/?view_only=eb195362d8ab49308e824dc4743bb3af

to participants in per mile (‰).[5] All participants were then asked how confident they were in their decision to interview the witness on a 10-point response format which ranged from 1—*not confident at all* to 10—*completely confident*. After this, participant in the control and research-based message conditions were asked the same social normative questions previously posed to participants in the social norms condition. All participants were then given the in-group identification measure. Finally, participants' attention during the film viewing was checked by asking them to identify the event described by the witness from two possible scenarios (one sentence long each). The two options had a slight but salient difference so that participants who paid attention should be able to pick the right option without much difficulty.

Power Calculations

To form a basis for interpretation, the frequently referenced Cohen's *d* guidelines which denoted a medium-sized effect as a mean difference of *d* = .50 (small *d* = .20, large *d* = .80; Cohen, 1988) was used. This was converted to Cohen's *f* (small *f* = .10, medium *f* = .25, and large *f* = .40) for use with G*Power 3.1 (Faul et al., 2007) to calculate main and interaction effects for ANOVA. If there were any significant effects to be found, Study 2 ($N = 214$, $\alpha = .05$) had a power of 91% to observe significant medium-sized main effects and a power of 84% to observe significant medium-sized two-way interaction effects.

Hypotheses

Based on the content of the research-based message (i.e., witnesses are generally reliable when BrAC is less than .10%) and the potential lack of research-based knowledge, we expected an interaction between information and witness intoxication. Compared to participants who received no information (i.e., control), participants who read the research-based message would be more likely to interview the moderately intoxicated witness compared to the sober witness (H1a) and less likely to interview the highly intoxicated witness compared to the sober witness (H1b). A similar interaction was expected for perceived witness credibility. We hypothesized that participants who read the research-based message (compared to participants in the control condition who received no information) would perceive the moderately intoxicated witness as more credible compared with the sober witness (H2a) and the highly intoxicated witness as less credible compared to the sober witness (H2b).

Based on the previous research demonstrating the effects of social normative influence (e.g., Miller and Prentice, 2016) and research which has found that police officers believed intoxicated witnesses were less credible than sober witnesses (Evans et al., 2009; Crossland et al., 2018; Monds et al., 2021a; Hagsand et al., 2022) as well as the potential lack of knowledge suggested previously (e.g., Hagsand et al., 2022), we predicted the following hypotheses for social norms. Compared to participants who received no information (i.e., control),

participants for whom pre-existing social norms were activated (i.e., considering what their colleagues would do or approve of doing) would be less likely to interview the moderately intoxicated (H3a) and the highly intoxicated (H3b) witness compared to the sober witness. We expected similar main effects for perceived witness credibility. Compared to those who received no information (i.e., control), participants for whom pre-existing social norms were activated would perceive the moderately intoxicated witness (H4a) and the highly intoxicated witness (H4b) as less credible compared to the sober witness.

Results

Manipulation Check of Social Norms

A strong positive correlation ($r = .68$, $p < .001$) between the descriptive and injunctive social norms measures indicated that what participants believed others would do and approve of doing, aligned well. A robust one-way ANOVA was conducted with information condition as independent variable and descriptive norm as dependent variable. There was a non-significant effect of information on descriptive norm [$F(2,204) = 1.53$, $p = .22$, $\eta^2 = .02$], indicating a shared view among participants regarding the actions of others. Another robust one-way ANOVA found a significant effect of information on the injunctive norm [$F(2,203) = 3.28$, $p = .04$, $\eta^2 = .03$]. *Post-hoc* comparisons showed a non-significant difference between the control ($M = 7.67$, $SD = 2.65$) and social norms conditions ($M = 7.37$, $SD = 3.46$, $p = .85$, $g = .10$, 95% CI [−.24, .44]). There was also a non-significant difference between the control and research-based message conditions ($M = 8.53$, $SD = 2.10$, $p = .09$, $g = .36$, 95% CI [.03, .70]). However, there was a significant difference between the social norms and research-based message conditions ($p = .05$, $g = .41$, 95% CI [.07, .75]). Results indicated that participants in the research-based message condition, to a larger degree, believed that others would approve of them interviewing the witness. This was unsurprising given that the research-based message contained information on the reliability of intoxicated witnesses.

Manipulation Check of Witness Intoxication

There was a significant main effect of witness intoxication on BrAC estimates [$F(2,201) = 87.64$, $p < .001$, $\eta_p^2 = .47$]. Bonferroni adjusted planned comparison showed a significant difference between the sober and the moderate conditions ($p < .001$, $g = .67$, 95% CI [.33, 1.01]), between the sober and high conditions ($p < .001$, $g = 1.97$, 95% CI [1.56, 2.38]), and between the moderate and high conditions ($p < .001$, $g = 1.42$, 95% CI [1.06, 1.79]). There was a non-significant effect of information on BrAC estimates [$F(2,201) = 2.40$, $p = .094$, $\eta_p^2 = .02$]. There was also a non-significant interaction effect of information and witness intoxication on BrAC estimate [$F(4,201) = .18$, $p = .950$, $\eta_p^2 = .00$]. Descriptive statistics are reported in **Table 4**. The results showed that participants clearly distinguished between all witness intoxication levels independent of information condition. An issue that is important to immediately notice is that participants estimated that the highly intoxicated witness had an average BrAC of .09% (see **Table 4**). This is much lower than the expected .15% and has consequences for the interpretation of the main analyses.

TABLE 4 | Means (SD) for participants BrAC (%) estimates in Study 2.

Witness intoxication	Information			Total
	Control	Social norm	RBM	
Sober	$n = 18$	$n = 24$	$n = 23$	$n = 65$
	.01 (.02)	.02 (.04)	.02 (.02)	.02 (.03)
Moderate	$n = 30$	$n = 20$	$n = 25$	$n = 75$
	.04 (.03)	.05 (.03)	.05 (.03)	.04 (.03)
High	$n = 24$	$n = 20$	$n = 26$	$n = 70$
	.08 (.04)	.10 (.05)	.10 (.03)	.09 (.04)
Total	$n = 72$	$n = 64$	$n = 74$	$N = 210$
	.05 (.04)	.05 (.05)	.06 (.04)	.05 (.04)

Responses were captured on a continuous scale ranging from 0 to 4. RBM, research-based message.

TABLE 5 | Means (SD) for interview probability in Study 2.

Witness intoxication	Information			Total
	Control	Social norm	RBM	
Sober	$n = 18$	$n = 24$	$n = 23$	$n = 65$
	8.22 (3.15)	9.00 (2.57)	8.48 (2.31)	8.60 (2.64)
Moderate	$n = 30$	$n = 22$	$n = 25$	$n = 77$
	6.90 (3.32)	7.00 (3.31)	8.32 (1.87)	7.39 (2.96)
High	$n = 25$	$n = 21$	$n = 26$	$n = 72$
	5.12 (3.60)	5.81 (4.09)	7.69 (3.33)	6.25 (3.78)
Total	$n = 73$	$n = 67$	$n = 74$	$N = 214$
	6.62 (3.54)	7.34 (3.56)	8.15 (2.58)	7.37 (3.30)

Response format ranged from 1—not likely at all to 10—very likely. RBM, research-based message.

Main Analyses

For each dependent measures, a 3 (Information: control vs. social norms vs. research-based message) × 3 (Witness intoxication: sober vs. moderate vs. highly intoxicated) between-subjects factorial ANOVA was conducted[6,7].

Interview Probability

Descriptive statistics are reported in **Table 5**. There was a significant main effect of witness intoxication on interview probability $[F(2,205) = 9.67, \ p < .001, \ \eta_p^2 = .09]$. Bonferroni adjusted planned comparisons showed a non-significant difference between the sober and moderate conditions ($p = .07$, $g = .43$, 95% CI [.09, .76]), a significant difference between the sober and high conditions ($p < .001$, $g = .72$, 95% CI [.37, 1.06]), and a non-significant difference between the moderate and

---

[6]The present study violated some of the assumptions of the general linear model. While analysis of variance has demonstrated robustness against violations (e.g., Carifio and Perla, 2007; Norman, 2010), the use of parametric test with ordinal data has been a much-debated issue (Carifio and Perla, 2008). Unreported sensitivity analyses using the non-parametric Kruskal-Wallis test re-tested the main analyses and these results did not alter any of the parametric inferences.
[7]Data availability: All SPSS data sets and outputs can be downloaded from the Open Science Framework (OSF). https://osf.io/qv8tn/?view_only=eb19536 2d8ab49308e824dc4743bb3af

high conditions ($p = .08$, $g = .33$, 95% CI [.01, .66]). Participants were least likely to interview the highly intoxicated witness and most likely to interview the sober witness. There was also a significant main effect of information on interview probability, $[F(2,205) = 3.80, p = .02, \eta_p^2 = .04]$. Planned contrasts showed a non-significant difference between the control and social norm conditions ($p = .33$, $g = .20$, 95% CI [−.13, .54]); therefore, hypotheses 3a and 3b were not supported. There was a significant difference between the control and research-based message conditions ($p = .007$, $g = .49$, 95% CI [.17, .82]). Results showed that a research-based message, but not social norms, affected the interview probability. There was a non-significant information by witness intoxication interaction on interview probability $[F(4,205) = 1.24, \ p = .29, \ \eta_p^2 = .02]$; therefore, hypotheses 1a and 1b were not supported. Despite a non-significant interaction, visual inspection of the data showed a convergence of the slopes which indicated that in the research-based message condition only, witness intoxication had no effect. Unplanned simple effects analysis confirmed this research-based message $[F(2,205) = .44, \ p = .64]$, control $[F(2,205) = 5.40, p < .01]$, and social norm $[F(2,205) = 6.06, p < .01]$.

Witness Credibility

Descriptive statistics are reported in **Table 6**. There was a significant main effect of witness intoxication on witness credibility $[F(2,205) = 17.06, p < .001, \eta_p^2 = .14]$. Bonferroni adjusted planned comparisons showed a significant difference between the sober and moderate conditions ($p = .02$, $g = .52$, 95% CI [.18, .86]), a significant difference between the sober and high conditions ($p < .001$, $g = .95$, 95% CI [.60, 1.30]), and a significant difference between the moderate and high conditions ($p = .01$, $g = .45$, 95% CI [.12, .77]). Results showed that participants rated the sober witness the most credible and the highly intoxicated witness the least credible. There was a non-significant main effect of information on witness credibility, $[F(2,205) = 2.53, p = .08, \eta_p^2 = .02]$. Planned contrasts showed a non-significant difference between the control and social norm conditions ($p = .186$, $g = .13$, 95% CI [−.20, .46]); therefore, hypotheses 4a and 4b were not supported. There was a non-significant difference between the control and research-based message conditions ($p = .364$, $g = .18$, 95% CI [−.15, .50]). Results showed that neither research-based message, nor social norms, affected witness credibility ratings. There was a non-significant interaction effect of information and witness intoxication on witness credibility $[F(4,205) = 1.96, \ p = .102, \eta_p^2 = .04]$, and therefore, hypotheses 2a and 2b were not supported.

Supplementary Analyses
Social Identity

Inspired by previous research (Baldry and Pagliaro, 2014), a median split ($Md = 9.00$) divided participants into *low identifiers* ($M = 7.48$, $SD = 3.13$) and *high identifiers* ($M = 7.70$, $SD = 3.15$) with respect to social identity. A 3 (Information: control vs. social norm vs. research-based message) × 2 (Social identification: low vs. high) between-subjects factorial ANOVA with the probability of interviewing the witness as dependent variable was conducted. There was a non-significant main

TABLE 6 | Means (SD) for witness credibility in Study 2.

Witness intoxication	Information			Total
	Control	Social norm	RBM	
Sober	$n = 18$	$n = 24$	$n = 23$	$n = 65$
	7.94 (1.31)	8.04 (1.46)	7.78 (1.45)	7.92 (1.40)
Moderate	$n = 30$	$n = 22$	$n = 25$	$n = 77$
	7.23 (2.00)	6.68 (1.49)	7.24 (1.74)	7.08 (1.78)
High	$n = 25$	$n = 21$	$n = 26$	$n = 72$
	6.08 (2.02)	5.33 (2.18)	7.04 (1.87)	6.21 (2.10)
Total	$n = 73$	$n = 67$	$n = 74$	$N = 214$
	7.01 (1.98)	6.75 (2.03)	7.34 (1.71)	7.04 (1.91)

Response format ranged from 1— not credible at all to 10—completely credible. RBM = research-based message.

TABLE 7 | Means (SD) for identification with the police occupation in Study 2.

Information	Social identity	
	Low identifiers	High identifiers
Control	$n = 27$	$n = 31$
	6.63 (3.44)	7.71 (2.91)
Social norm	$n = 35$	$n = 28$
	7.66 (3.36)	6.89 (3.76)
Research-based message	$n = 29$	$n = 37$
	8.07 (2.37)	8.30 (2.76)
Total	$n = 91$	$n = 96$
	7.48 (3.13)	7.70 (3.16)

Composite score of the four-item scale. Response format ranged from 1—completely disagree to 10—completely agree.

effect of information on interview probability [$F(2,181) = 2.01$, $p = .14$, $\eta_p^2 = .02$]. There was a non-significant main effect of identification on interview probability [$F(1,181) = .16$, $p = .69$, $\eta_p^2 = .00$]. Finally, there was also a non-significant interaction effect of information and social identification on interview probability [$F(2,181) = 1.32$, $p = .27$, $\eta_p^2 = .01$]. Descriptive statistics are reported in **Table 7**. Results showed that identification with the police occupation had no impact on the probability of interviewing a witness, nor were participants who identified strongly with the police occupation influenced by social norms to a greater degree than those who identified less strongly.

Confidence Rating

A robust one-way ANOVA was conducted with witness intoxication level as independent variable and participants' confidence in their decision to interview the witness as the dependent variable. There was non-significant effect of witness intoxication on participants confidence ratings {$F(2,207) = 1.11$, $p = .33$, $\eta^2 = .01$, 95% CI [.00, .05]}. Descriptive statistics are reported in **Table 8**. Results showed that participants who saw the film with the highly intoxicated witness were no less confident in their decision to interview than those who saw the sober and moderate witnesses.

Discussion

The first aim of Study 2 was the same as in Study 1 to investigate whether police officers' inconsistent interview decisions could be attributed to a lack of research-based knowledge. Again, there was no explicit test of police officers and recruit's knowledge, instead such a conclusion, would be based on participants' behavior when responding to the questions. In Study 2, there was the additional concern that police officers and recruits estimated that the highly intoxicated witness had an average BrAC of .09% (see **Table 4**), which was much lower than the intended .15%. This means that participants based their answers to the questions on a BrAC level in the low to moderate range (i.e., BrAC < .10%). Therefore, any further interpretation of the results must account for this lower estimate. Because of this, in Study 2, a lack of knowledge would be concluded if (1) interview probability differed across witness intoxication level, even for the "highly" intoxicated witness, and (2) if perceived witness credibility differed across witness intoxication level, again even for the "highly" intoxicated witness. Contrary to Study 1, participants made clear distinctions between all three witnesses' intoxication levels. Interview probability remained the same for the sober and moderately intoxicated witnesses but differed significantly for the "highly" intoxicated witness. Perceived witness credibility significantly differed across all three levels of intoxication. Had participants possessed research-based knowledge, it should have been unlikely that they would have treated any of the witnesses differently because they all were estimated by the participants to have a BrAC level in the low to moderate range. A range where scientific research has found that intoxicated witnesses can be reliable (see Altman et al., 2019; Jores et al., 2019, for reviews and meta-analysis) and where the consequences of postponing an interview could lead to less complete and accurate statements (e.g., Hagsand et al., 2017; Hildebrand Karlén et al., 2017).

The second aim of Study 2 was again the same as in Study 1, to investigate whether their decision to interview, as well as their perceptions of the witnesses' credibility could be influenced by scientific research. In line with Study 1, perception of witness credibility was unaffected by the research-based message. In contrast with Study 1, Study 2 found that the highly intoxicated witness was more likely to be interviewed compared with the condition that received no information (i.e., control). The research-based message informed participants about research regarding the reliability of low to moderately intoxicated witness statements and the consequences of postponing the interview. As such, it was unexpected to find an increase in interview probability for the highly intoxicated witness. However, when accounting for participants inaccurate estimates of intoxication level, these results made sense. The highly intoxicated witness was considered by participants to be in the low to moderate range and therefore encompassed by the information in the message. These findings, therefore, indicated that a research-based message might assist police officers and recruits to make decisions that are more in line with research findings. Such a message is more likely to affect the decision to interview than it is to affect perceptions of witness credibility.

TABLE 8 | Means and standard deviations for confidence ratings across witness intoxication in Study 2.

Witness intoxication	*n*	*M*	*SD*
Sober	63	8.79	1.89
Moderate	76	8.25	2.33
High	71	8.38	2.41
Total	210	8.46	2.24

Table shows descriptive statistics for participants' confidence in their decision to interview the witness across witness intoxication level. Response format ranged from 1—not confident at all to 10—absolutely confident.

The third aim was again to investigate whether police officers' decision-making and perceptions of witness credibility are biased by pre-existing social norms. In line with Study 1, there was no statistically significant effect of social norms on interview probability in Study 2. People tend to comply with social norms more in uncertain situations where the right course of action is unclear (Bell and Cox, 2015). However, confidence ratings showed that all participants, regardless of witness intoxication level, were comparably confident in their decision to interview. Without the element of uncertainty, there may have been little reason for participants to look to others for guidance which may have diminished the impact of social norms. On the other hand, identification with the police occupation was high across all conditions which should have made compliance with the norm more likely (Baldry and Pagliaro, 2014; Liu et al., 2019).

GENERAL DISCUSSION

In recent surveys, Swedish police officers reported inconsistent individual interview decisions, absent policy guidelines, and subjective methods for assessing intoxication level among witnesses, victims, and suspects (Hagsand et al., 2021, 2022). Officers also reported perceptions of credibility contrary to research on this witness group (see Altman et al., 2019; Jores et al., 2019, for reviews and meta-analysis). This may produce uncertain situations in which the decision to interview might be unjustly influenced by social norms. Therefore, two studies were conducted to investigate whether (1) police officers' inconsistent interview decisions are attributable to a lack of research-based knowledge; (2) their decision to interview, as well as their perceptions of the witnesses' credibility could be influenced by scientific research; and (3) police officers decision-making and perceptions of witness credibility are biased by pre-existing social norms.

Prior to discussing the findings, it is necessary to mention again that participants' estimation of intoxication level did not align with the pre-study validation of the stimulus material. In Study 2, participants inaccurately perceived both intoxicated witnesses to be low to moderately intoxicated. It is interesting to note that the university students in Study 1 made a far more accurate assessments about the highly intoxicated witness than the police officers in Study 2. Students were, however, far less accurate when assessing the sober witness compared to police officers in Study 2. Explanations addressing these

issues surround the discussion of whether video clips are sufficient for making accurate estimates (Brick and Carpenter, 2001), and whether inaccuracies stem from using observational methods for assessing intoxication level which are ineffective (Rubenzer, 2011). Both the moderately and the highly intoxicated witnesses were estimated lower than what was intended in the research design, which could suggest the police frequent encounters with intoxicated people—many who are heavy drinkers (Evans et al., 2009; Crossland et al., 2018; Monds et al., 2021a; Hagsand et al., 2022), may have desensitized them to the behavioral effects of alcohol-intoxication. This might have resulted in the fact that no witness was perceived to be highly intoxicated by the police officers and recruits in Study 2. Any interpretations and implications that are made from the results therefore, must treat witnesses only from a sober to moderate intoxication level (i.e., <.10%), as these were the levels upon which participants based their answers to the survey questions.

The first aim was to investigate whether the previously reported inconsistent interview decision could be attributed to a lack of research-based knowledge (see also Hagsand et al., 2021, 2022). Although we did not explicitly test participants knowledge regarding research findings, we find it reasonable to expect cognizant police officers and recruits to consider low to moderately intoxicated witnesses comparably credible to sober witness, not hesitating to interview the former as much as the latter. This would be in line with research findings (e.g., Altman et al., 2019; Jores et al., 2019). Because all witnesses in Study 2 were considered low to moderately intoxicated, there should have been no variation in interview probability or perceived credibility across intoxication levels. However, participants rated the previously deemed highly intoxicated witness as less likely to be interviewed and less credible compared with the others. In addition, an unplanned simple effects analysis showed that after reading the research-based message about the reliability of low to moderately intoxicated witnesses, interview probability was less affected by degree of intoxication. This difference notes that participants did not make judgments based on pervious knowledge. These findings support previously self-reported survey results (e.g., Hagsand et al., 2022) as well as research which has found that police officers regarded intoxicated witnesses as less credible compared with sober witnesses (Evans et al., 2009; Crossland et al., 2018; Monds et al., 2021a; Hagsand et al., 2022). From the present data, it cannot be concluded whether police officers and recruits lacked prior knowledge in making judgments, but our results would favor the inference that this was in fact the case.

The second aim was to investigate whether their decision to interview, as well as their perceptions of the witnesses' credibility could be influenced by scientific research. What both studies found was that perceived witness credibility remained unaffected by the research-based message. That is, regardless of the research-based information participants received, they were not influenced in their credibility judgments of the witnesses. However, in Study 2, the research-based message did influence participant's willingness to interview the highly intoxicated witness, which was not the

case in Study 1. The more informative message used in Study 2 could account for this discrepancy between studies, as a one sentence long message was perhaps not sufficient to affect participants in Study 1. Unexpectedly, the moderately intoxicated witness, who was estimated to have an average BrAC of .04%, was not more likely to be interviewed compared to the sober witness. A possibility is that police officers and recruits did not consider such a low intoxication level a reason to postpone the interview. As such, the research-based message was not additionally helpful to them. These findings indicated that police officers and recruits may have a basic understanding that witnesses can be interviewed at low levels of alcohol-intoxication (i.e., around .04%), but that they believe that this ceases to be the case at a lower intoxication level than what scientific research has suggested. It is interesting to note that the decision to interview increased after reading the message, even though participants still considered both intoxicated witnesses to be less credible than the sober witness. It appears that the decision to interview was made despite internally held perceptions. Perhaps participants were more affected by the information regarding the consequences of postponing the interview than they were by the information about intoxicated witness reliability. As such, they may have decided to interview the witness to avoid losing important details to a crime but remained confident that intoxicated witnesses are less credible. In summary, a research-based message may be a key method to encourage the right procedure when deciding to interview an intoxicated witness. In addition, and in concurrence with previous literature, the results showed the tendency of the police to perceive witnesses as less credible, even with BrAC as low as .04% (as the current study has found; Evans et al., 2009; Crossland et al., 2018; Monds et al., 2021a; Hagsand et al., 2022).

The third aim was to investigate whether police officers' decision-making and perceptions of witness credibility are biased by pre-existing social norms. Neither study found that participants were biased by social normative influence, neither in their interview decision nor their estimates of witness credibility. Participants who were prompted to think about injunctive and descriptive norms were comparable to those who were not stimulated by such norms, and this trend was consistent across all intoxication conditions. The results seem to infer that social norm had little impact on both the decision to interview a witness, and perceived credibility. Considering the abundance of the general literature demonstrating social normative influence in various behaviors and contexts, these findings were unexpected (e.g., Rivis and Sheeran, 2003; Melnyk et al., 2010; Fischer et al., 2011; Baldry and Pagliaro, 2014; Bergquist et al., 2019), but more research within the police context is needed.

As a possible explanation for the null findings of social norms in Study 1, Study 2 included a measure of identification with the police occupation. *Social identity theory* (Tajfel and Turner, 1979) has suggested that a strong sense of in-group identification will incite people to act more in line with their group identity than their individual identity, resulting

in social norm influences being particularly effective when the group identification is strong. In contrast with previous research (Baldry and Pagliaro, 2014; Liu et al., 2019), the degree of identification with the police occupation did not impact the effect of social norms in Study 2. In addition, confidence in their decision to interview the witness remained the same across intoxication levels. All participants strongly identified with the police occupation which (apart from indicating possible ceiling effects) should have increased the social normative influence. Having the questionnaire at the end could have impacted the study in two competing ways. First, the study procedure itself could have made their police identities salient before they answered the questionnaire, which would explain the high average. Second, social norms remained non-significant, which perhaps indicated that their identities were not salient when they answered the dependent measures. Had the identification questions been included earlier in the study the participants' identities could have been salient when they made their decisions regarding the witnesses and the study procedure could not have affected their identification responses. As previously stated, social norms must be activated to influence decisions and behaviors (Cialdini et al., 1990), and they have a greater influence on those in uncertain situations (Bell and Cox, 2015). As an explanation for the present results, it may be possible that attentional salience is a necessary, but not sufficient condition, under which social norms exert their influence. Some other psychological motivation (e.g., uncertain situations) may also be necessary for the effects of social norms to emerge. Further research should explore such matters in the context of legal psychology and policing.

LIMITATIONS

The inclusion of both professional police officers and police recruits was a sound decision because it was reasonable to assume many similarities between these populations (Gatto and Dambrun, 2012; Lander, 2013; Wieslander, 2019). However, it is possible that differential experiences between these groups could have influenced the findings. In addition, despite research (*ibid.*) which has indicated strong socialization processes, professional police and police recruits could be groups with differing normative codes of conduct. Another limitation concerns the measure of social identity which was implemented shortly after data collection had commenced. Consequently, 30 participants completed the study before implementation. Also, since it was a measure of identification with the police occupation, it could be biased toward professional police officers and possibly have excluded recruits. Another limitation was that the small sample size in Study 1 restrained any firm conclusions; however, as this was designed as a pilot study, we believe that Study 1 fulfilled its purpose. Further on, although Study 2 did not have statistical power to detect small effects, it had power to detect medium- to large-sized effects. Further research could aim at trying to gain more police participants and build upon this study.

IMPLICATIONS AND FUTURE DIRECTIONS

The present research did not examine police officers and recruit's knowledge directly (i.e., *via* an explicit knowledge test). Future research should examine police knowledge explicitly by asking officers and recruits to complete a proper test of their knowledge regarding intoxicated witnesses' ability to recall events. In addition, in the present studies, participants were not asked about what training in assessing alcohol-intoxication they may have received. This limited the scope of the discussion around potential issues with using observational methods to assess intoxication level. We encourage researchers to examine this in future studies.

The present findings cautiously suggest a potential vacancy in the Swedish police education. Not only are national guidelines for professional police required, but future research should also investigate this potential gap concerning alcohol-intoxication and witness memory in the curriculum at Swedish Police Academies. This should not be taken as an indication that police officers, departments, police recruits, or the academies are solely responsible for this potential deficiency. Researchers also carry a responsibility to share knowledge in an accessible manner where bridging this gap is paramount for scientific research to become relevant outside of the scientific community (see Hagsand et al., 2020; Hagsand, 2021).

The research-based message impacted the interview decision but not the perception of credibility and future research should investigate why this was the case. Still, participants embraced the content of the message and decided to interview in line with research recommendations. Therefore, future implementation of national policy guidelines regarding alcohol-intoxicated witnesses could be disseminated *via* an informative message. Since the effects of interventions have tended to be strongest directly after implementation (e.g., Fernandes et al., 2014), future research should investigate the long-term effects of providing police officers and recruits with research-based information.

Because some of our findings contradicted the general trend within the field, future research should attempt to replicate these findings and examine if there are any circumstances under which social norms could influence police decisions and perceptions. One possibility is that self-selection bias (i.e., which participants decided to take part in the studies) may have skewed some results. While this is a common issue in any research design, we still recommend that future research replicate these findings in other samples of the population.

Due to the novelty of the current study, many additional advances within the study design are made available for future research. For example, different genders could act as witness and interviewers, and instead of using video clips for assessment, participants could view face-to-face interactions between interviewer and witnesses, and additional dependent variables. We encourage other researchers to not only replicate the proposed study (e.g., making it more generalizable to other countries), but also add additional variables and make other

adjustments to ultimately further the field of legal psychology in a meaningful direction.

CONCLUSION

The present findings suggested that police officers and police recruits might make decisions in the absence of research-based knowledge, leading to inconsistent interview decisions, as well as their ability to deem witnesses as credible. The results also highlighted that a research-based message, in the shape of procedure guidelines, could be a way to align the decision to interview with research recommendations, but only when there is enough information included in the message, as just a single sentence might not work. Regardless of intoxication level, witnesses were perceived as less credible when under the influence, and this judgment yet again appeared to be made in the absence of scientific research. Furthermore, social norms were found to be ineffective to influence police on their decisions to interview, and this invites further investigation. The current findings added to the legal psychology literature (Evans et al., 2009; Crossland et al., 2018; Monds et al., 2021a; Hagsand et al., 2022) by showing that the perception of intoxicated witness as less credible than sober witness is present at BrAC levels as low as .04%. These findings also indicated that police officers and recruits may have a basic understanding that witnesses can be interviewed at low levels of alcohol-intoxication (i.e., around .04%), but that they believe that this ceases to be the case at intoxication levels lower than what scientific research has suggested. This novel examination on social norms and research-based messages in the context of police studies on alcohol-intoxicated witnesses could help to inform future research endeavors to continue to build upon this knowledge and examine this area more closely.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at Open Science Framework (OSF) https://osf.io/qv8tn/?view_only=eb195362d8ab49308e824dc4743bb3af.

ETHICS STATEMENT

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors contributed to the conceptualization of research questions, the study design, the data collecting procedure, and approved the submitted version. DP carried out quantitative analyses

REFERENCES

Altman, C., Schreiber Compo, N., Hagsand, A. V., and Evans, J. R. (2019). "State of intoxication: A review of the effects of alcohol on witnesses' memory." in *Evidence-Based Investigative Interviewing: Applying Cognitive Principles*. eds. J. Dickinson, N. S. Compo, R. N. Carol, M. McCauley and B. Schwartz (NY: Routledge).

Altman, C., Schreiber Compo, N., McQuiston, D., Hagsand, A. V., and Cervera, J. (2018). Witnesses' memory for events and faces under elevated levels of intoxication. *Memory* 26, 946–959. doi: 10.1080/09658211.2018.1445758

Anderson, J. E., and Dunning, D. (2014). Behavioral norms: variants and their identification. *Soc. Personal. Psychol. Compass* 8, 721–738. doi: 10.1111/spc3.12146

Baldry, A. C., and Pagliaro, S. (2014). Helping victims of intimate partner violence: the influence of group norms among lay people and the police. *Psychol. Violence* 4, 334–347. doi: 10.1037/a0034844

Barreto, M., and Ellemers, N. (2000). You can't always do what you want: social identity and self-presentational determinants of the choice to work for a low-status group. *Personal. Soc. Psychol. Bull.* 26, 891–906. doi: 10.1177/01461672002610001

Bell, D. C., and Cox, M. L. (2015). Social norms: do we love norms too much?: social norms. *J. Fam. Theory Rev.* 7, 28–46. doi: 10.1111/jftr.12059

Bergquist, M., Nilsson, A., and Schultz, W. P. (2019). A meta-analysis of field-experiments using social norms to promote pro-environmental behaviors. *Glob. Environ. Chang.* 59:101941. doi: 10.1016/j.gloenvcha.2019.101941

Brick, J., and Carpenter, J. A. (2001). The identification of alcohol intoxication by police. *Alcohol. Clin. Exp. Res.* 25, 850–855. doi: 10.1111/j.1530-0277.2001.tb02290.x

Carifio, J., and Perla, R. J. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *J. Soc. Sci.* 3, 106–116. doi: 10.3844/jssp.2007.106.116

Carifio, J., and Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Med. Educ.* 42, 1150–1152. doi: 10.1111/j.1365-2923.2008.03172.x

Cialdini, R. (2012). "The focus theory of normative conduct," in eds. P. Van Lange, A. Kruglanski and E. T. Higgins *Handbook of Theories of Social Psychology*. Vol. 2 (United States: SAGE), 295–312.

Cialdini, R. B., Demaine, L. J., Sagarin, B. J., Barrett, D. W., Rhoads, K., and Winter, P. L. (2006). Managing social norms for persuasive impact. *Soc. Influ.* 1, 3–15. doi: 10.1080/15534510500181459

Cialdini, R. B., and Goldstein, N. J. (2004). Social influence: compliance and conformity. *Annu. Rev. Psychol.* 55, 591–621. doi: 10.1146/annurev.psych.55.090902.142015

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *J. Pers. Soc. Psychol.* 58, 1015–1026. doi: 10.1037/0022-3514.58.6.1015

Cialdini, R. B., and Trost, M. R. (1998). "Social influence: social norms, conformity, and compliance," in *The Handbook of Social Psychology*.

4th Edn. Vol. 1–2. eds. D. T. Gilbert, S. T. Fiske and G. Lindzey (New York: McGraw-Hill), 151–192.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. 2nd Edn. United States: L. Erlbaum Associates.

College of Policing (2019). Guidelines for first responders: UK Police report. Available at: https://www.college.police.uk/guidance/obtaining-initial-accounts/alcohol-intoxication (Accessed August 20, 2021).

Crano, W. D. (2000). Milestones in the psychological analysis of social influence. *Group Dyn. Theory Res. Pract.* 4, 68–80. doi: 10.1037/1089-2699.4.1.68

Crossland, D., Kneller, W., and Wilcock, R. (2018). Intoxicated eyewitnesses: prevalence and procedures according to England's police officers. *Psychol. Crime Law* 24, 979–997. doi: 10.1080/1068316X.2018.1474216

Ellemers, N., and Haslam, S. A. (2012). "Social identity theory," in eds. LangeP. Van, A. Kruglanski and E. T. Higgins *Handbook of Theories of Social Psychology*. Vol. 2 (United States: SAGE), 379–398.

Evans, J. R., and Schreiber Compo, N. (2010). Mock jurors' perceptions of identifications made by intoxicated eyewitnesses. *Psychol. Crime Law* 16, 191–210. doi: 10.1080/10683160802612890

Evans, J. R., Schreiber Compo, N., Carol, R. N., Nichols-Lopez, K., Holness, H., and Furton, K. G. (2018). The impact of alcohol intoxication on witness suggestibility immediately and after a delay. *Appl. Cogn. Psychol.* 33, 358–369. doi: 10.1002/acp.3502

Evans, J. R., Schreiber Compo, N., and Russano, M. B. (2009). Intoxicated witnesses and suspects: procedures and prevalence according to law enforcement. *Psychol. Public Policy Law* 15, 194–221. doi: 10.1037/a0016837

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146

Fernandes, D., Lynch, J. G., and Netemeyer, R. G. (2014). Financial literacy, financial education, and downstream financial behaviors. *Manag. Sci.* 60, 1861–1883. doi: 10.1287/mnsc.2013.1849

Fischer, P., Krueger, J. I., Greitemeyer, T., Vogrincic, C., Kastenmüller, A., Frey, D., et al. (2011). The bystander-effect: a meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychol. Bull.* 137, 517–537. doi: 10.1037/a0023304

Gatto, J., and Dambrun, M. (2012). Authoritarianism, social dominance, and prejudice among junior police officers: the role of the normative context. *Soc. Psychol.* 43, 61–66. doi: 10.1027/1864-9335/a000081

Hagsand, A. V. (2021). Eyewitness testimony: how much alcohol is too much? *Res. Outreach* 126, 166–169. doi: 10.32907/RO-126-1847382832

Hagsand, A. V., Evans, J. R., Pettersson, D., and Schreiber Compo, N. (2021). A survey of police officers encounters with sober, alcohol- and drug-intoxicated suspects in Sweden. *Psychol. Crime Law* 1–22. doi: 10.1080/1068316X.2021.1929978 [Epub ahead of print]

Hagsand, A. V., Pettersson, D., Evans, J. R., and Schreiber Compo, N. (2022). Police survey: procedures and prevalence of intoxicated witnesses and victims in Sweden. *Eur. J. Psychol. Appl. Legal Context* 14, 21–31. doi: 10.5093/ejpalc2022a3

Hagsand, A. V., Roos Af Hjelmsäter, E., Granhag, P. A., Fahlke, C., and Söderpalm Gordh, A. (2017). Witnesses stumbling down memory lane: the effects of

alcohol intoxication, retention interval, and repeated interviewing. *Memory* 25, 531–543. doi: 10.1080/09658211.2016.1191652

Hagsand, A., Stenman, B., and Sundqvist, J. (2020). Fler brott kan klaras upp med hjälp av alkoholpåverkades vittnesmål [More crimes can be resolved by using alcohol-intoxicated person's testimonies]. *Dagens Juridik*. Dagens Juridik (Daily Law). Available at: https://www.dagensjuridik.se/debatt/fler-brott-kan-klaras-upp-med-hjalp-av-alkoholpaverkades-vittnesmal/ (Accessed August 20, 2021).

Hildebrand Karlén, M., Roos Af Hjelmsäter, E., Fahlke, C., Granhag, P. A., and Söderpalm-Gordh, A. (2017). To wait or not to wait? Improving results when interviewing intoxicated witnesses to violence. *Scand. J. Psychol.* 58, 15–22. doi: 10.1111/sjop.12345

Hirn Mueller, D., Schreiber Compo, N., Molina, J., Bryon, A., and Pimentel, P. S. (2015). Productive and counterproductive interviewing techniques: do law enforcement investigators know the difference? *Psychol. Public Policy Law* 21, 295–308. doi: 10.1037/law0000047

Howe, M. L., and Knott, L. M. (2015). The fallibility of memory in judicial processes: lessons from the past and their modern consequences. *Memory* 23, 633–656. doi: 10.1080/09658211.2015.1010709

Jores, T., Colloff, M. F., Kloft, L., Smailes, H., and Flowe, H. D. (2019). A meta-analysis of the effects of acute alcohol intoxication on witness recall. *Appl. Cogn. Psychol.* 33, 334–343. doi: 10.1002/acp.3533

Kassin, S. M., Tubb, V. A., Hosch, H. M., and Memon, A. (2001). On the "general acceptance" of eyewitness testimony research. *Am. Psychol.* 56, 405–416. doi: 10.1037//0003-066X.56.5.405

Kebbell, M. R., and Milne, R. (1998). Police officers' perceptions of eyewitness performance in forensic investigations. *J. Soc. Psychol.* 138, 323–330. doi: 10.1080/00224549809600384

Keizer, K., Lindenberg, S., and Steg, S. (2008). Spreading of disorder. *Science* 322, 1681–1685. doi: 10.1126/science.1161405

Keuschnigg, M., and Wolbring, T. (2015). Disorder, social capital, and norm violation: three field experiments on the broken windows thesis. *Ration. Soc.* 27, 96–126. doi: 10.1177/1043463114561749

Lander, I. (2013). Obstacles for changes within the (Swedish) police force: professional motivations, homosociality, and ordering practices. *J. Scandinavian Stud. Criminol. Crime Prevent.* 14, 43–61. doi: 10.1080/14043858.2013.773691

Legros, S., and Cislaghi, B. (2020). Mapping the social-norms literature: an overview of reviews. *Perspect. Psychol. Sci.* 15, 62–80. doi: 10.1177/1745691619866455

Liu, J., Thomas, J. M., and Higgs, S. (2019). The relationship between social identity, descriptive social norms and eating intentions and behaviors. *J. Exp. Soc. Psychol.* 82, 217–230. doi: 10.1016/j.jesp.2019.02.002

Marier, C. J., and Moule, R. K. (2019). Feeling blue: officer perceptions of public antipathy predict police occupational norms. *Am. J. Crim. Justice* 44, 836–857. doi: 10.1007/s12103-018-9459-1

Melnyk, V., Van Herpen, E., and Van Trijp, J. C. M. (2010). The influence of social norms in consumer decision making: a meta-analysis. *Adv. Consum. Res.* 37, 463–464.

Miller, D. T., and Prentice, D. A. (2016). Changing norms to change behavior. *Annu. Rev. Psychol.* 67, 339–361. doi: 10.1146/annurev-psych-010814-015013

Mindthoff, A., Evans, J. R., Schreiber Compo, N., Polanco, K., and Hagsand, A. V. (2021). No evidence that low levels of intoxication at both encoding and retrieval impact scores on the Gudjonsson suggestibility scale. *Psychopharmacology* 238, 1633–1644. doi: 10.1007/s00213-021-05797-9

Mindthoff, A., Hagsand, A. V., Schreiber Compo, N., and Evans, J. (2019). Does alcohol loosen the tongue? Intoxicated persons' willingness to report transgressions or criminal behavior carried out by themselves or others. *Appl. Cogn. Psychol.* 33, 414–425. doi: 10.1002/acp.3480

Mintzer, M. Z. (2007). The acute effects of alcohol on memory: a review of laboratory studies in healthy adults. *Int. J. Disabil. Hum. Dev.* 6, 397–403. doi: 10.1515/IJDHD.2007.6.4.397

Monds, L. A., Cullen, H. J., Kloft, L., Sumampouw, N., van Golde, C., Harrison, A. W., et al. (2021a). Police perceptions of eyewitness impairment due to alcohol and other drug use: a cross-cultural comparison. *Police Pract. Res.* 23, 34–49. doi: 10.1080/15614263.2021.1979397

Monds, L. A., Cullen, H. J., Kloft, L., van Golde, C., Harrison, A. W., and Flowe, H. (2021b). Memory and credibility perceptions of alcohol and other drug intoxicated witnesses and victims of crime. *Psychol. Crime Law* 1–21. doi: 10.1080/1068316X.2021.1962871 [Epub ahead of print]

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Adv. Health Sci. Educ.* 15, 625–632. doi: 10.1007/s10459-010-9222-y

Parker, E. S., Birnbaum, I. M., and Noble, E. P. (1976). Alcohol and memory: storage and state dependency. *J. Verbal Learn. Verbal Behav.* 15, 691–702. doi: 10.1016/0022-5371(76)90061-X

Rivis, A., and Sheeran, P. (2003). Descriptive norms as an additional predictor in the theory of planned behaviour: a meta-analysis. *Curr. Psychol.* 22, 218–233. doi: 10.1007/s12144-003-1018-2

Rubenzer, S. (2011). Judging intoxication. *Behav. Sci. Law* 29, 116–137. doi: 10.1002/bsl.935

Schreiber Compo, N., Carol, R. N., Evans, J. R., Pimentel, P., Holness, H., Nichols-Lopez, K., et al. (2017). Witness memory and alcohol: the effects of state-dependent recall. *Law Hum. Behav.* 41, 202–215. doi: 10.1037/lhb0000224

Schreiber Compo, N., Vallano, J., Rivard, J., Hagsand, A. V., Pena, M., and Altman, C. (2019). "Methods of studying eyewitness memory," in *Handbook of Research Methods in Human Memory*. eds. H. Otani and B. L. Schwartz. 1st ed (New York: Routledge), 253–266.

Söderpalm, B. (2011). "Alkohol," in *Beroendemedicin*. eds. J. Franck and I. Nylander (Lund, Sweden: Studentlitteratur AB), 93–112.

Stuster, J. (2006). Validation of the standardized field sobriety test battery at 0.08% blood alcohol concentration. *Hum. Factors* 48, 608–614. doi: 10.1518/001872006778606895

Tajfel, H., and Turner, J. (1979). "An integrative theory of intergroup conflict," in *The Social Psychology of Intergroup Relations*. eds. W. G. Austin and S. Worchel (United States: Brooks/Cole), 33–47.

Wieslander, M. (2019). Learning the (hidden) silence policy within the police. *Stud. Contin. Educ.* 41, 308–325. doi: 10.1080/0158037X.2018.1497592

Check for updates

Assigning Punishment: Reader Responses to Crime News

Kat Albrecht[1]* and Janice Nadler[2]

[1] *Department of Criminal Justice and Criminology, Andrew Young School of Policy Studies, Georgia State University, Atlanta, GA, United States, [2] Northwestern Pritzker School of Law, American Bar Foundation, Chicago, IL, United States*

In this study we test how the composition of crime news articles contributes to reader perceptions of the moral blameworthiness of vehicular homicide offenders. After employing a rigorous process to develop realistic experimental vignettes about vehicular homicide in Minnesota, we deploy a survey to test differential assignments of suggested punishment. We find that readers respond to having very little information by choosing neutral or mid-point levels of punishment, but increase recommended punishment based on information about morally charged conduct. By contrast, information about the perpetrator's immigration status caused respondents to split into two groups on whether the offense deserves neutral or increased punishment. We find that political ideology strongly influences recommendations for more severe punishment when the immigration status of the perpetrator is revealed. We argue that this difference represents a moral dimension to punishment and blameworthiness that incorporates factors outside the active offense and therefore reveals the social influence of differential reporting in shaping public perception.

Keywords: blameworthiness, homicide, punishment, crime news, political ideology

INTRODUCTION

The content and construction of crime news provides an important resource for examining social inequality. American media produces a large quantity of news about crime, and this reporting resonates with Americans (Boulahanis and Heltsley, 2004; Norman, 2018). Importantly, the news is not a monolith; instead, it is a shared universe of interactive creation, allowing us to digest information from the world around us and extract value from it (Pan and Kosicki, 1993; Berkowitz, 1997; Lu, 2012). News shapes our perception of the world – not by providing an objective reflection of facts, but rather by filtering information through a lens of news creation constructed by news reporters (Schudson, 2011). By studying the filtering process through which information becomes news stories, we can understand how readers form beliefs and opinions about guilt and innocence in crime news.

In this study we analyzed how the construction of news stories can change the perceptions of news readers. Specifically, we tested how altering both the quantity and the nature of the information presented can change perceptions of blameworthiness and punishment. First, we conducted a detailed content analysis of homicide news articles in Minnesota to develop three news vignettes that cue different levels of moral culpability of vehicular homicide offenders. Next, we conducted a survey experiment using the news vignettes to measure perceptions of punishment. We observed differing punishment recommendations that varied according to political views and other demographic factors. The results suggest a link between news and the current political climate, specifically invoking beliefs about morality as guiding belief in punishment.

THE IMPORTANCE OF CRIME NEWS

Newspapers do not exist in a vacuum; they are created for and digested by an audience who themselves exist in the social world. Thus, the flow of information from news media is not unidirectional; rather it is a socio-cognitive relationship involving multiple actors. Pan and Kosicki (1993) describe the shared cultural universes of sources, journalists and audiences in the dissemination of news media with particular emphasis on the role of the audience as both readership and financial life-force for the institution of news. Shoemaker (2006) explains the logistics of this system of news and the interactive roles of its constituents.

"News is a commodity. It can be bought, sold, and traded. Journalists manufacture the news. Public relations firms manipulate the news. The audience consumes the news. Advertisers pay to place their products next to the news. News travels by word of mouth, across the Internet and other mass media. Professional associations focus on the production of news and on social science research about news. Televised news shouts at us in airport waiting rooms. News is ubiquitous" (106).

These tensions are not about fabricating news, but rather characterize news as a social institution shaped by economics, technology, politics, culture, and organizational structures (Schudson, 2011). This perspective helps us transcend the logistical process of reporting news and instead intuit value from its actual construction (see Berkowitz, 1997; Lu, 2012). Fishman (1988) argues that the news is in fact socially constructed, employing the example of a 1976 crime wave against elderly New Yorkers. This particular crime wave, while made up of real criminal incidents – was not actually an increase in crime from the same period in the previous year. Fishman explained that reporters did not fabricate the news, rather "they gave a determinate form and content to the incidents they report(ed)" (1988:10–11). This explanation gives reporters greater status than inscribers of rote fact – instead they interpret and ascribe meaning to events in the way that they report them. Indeed, reporters are quite cognizant of the social meaning of the events they report about even though news is very subjective (Gieber, 1964). The shaping of news is important because of its influence in the everyday lives of consumers. Ninety-three percentage of Americans say they follow the news at least occasionally, a large majority of them reporting that they do so for reasons that are primarily due to social interactions and civic responsibility (Purcell et al., 2010). In this way, the very circulation of news is dependent on the same society it reports about.

Crime news is one of the most prevalent types of reported news, but numerous studies have concluded crime news does not correlate with actual crime rates (Graber, 1979; Dorfman et al., 2001; Boulahanis and Heltsley, 2004). For example, a 2001 study of crime reports in the LA Times concluded that 80% of murders were reported on, but only 2% of physical and sexual assaults received news coverage (Dorfman et al., 2001). This creates a news-scape where some crime news is disproportionately reported, with a particular emphasis on murders. While the sheer volume of crime reporting as a percentage of space may be high, researchers conclude that this deluge of crime content may actually be keeping pace with the readerships desire to read

about crime (Graber, 1979). So, in this sense, news about crime is reported to the same extent that readers want to read about crime rather than in proportion to its actual occurrence.

The prevalence and construction of crime news matters because of its connection to negative consequences on attitudes, including racial stereotyping, public mis-perceptions of certain people as super-predators, and fostering fear of crime that does not accurately reflect the real spatial/demographic picture of crime (Barlow et al., 1995; Gilliam et al., 1996; Sorenson et al., 1998; Thorson, 2001; Boulahanis and Heltsley, 2004). These effects are attributable not only to the simple dichotomy of which cases are covered and which ones are not, but also to the way in which cases are covered and constructed. In one study, researchers found that the way news is reported implies that minority persons, unemployed persons, and male youths are more often members of deviant social groups (Humphries, 1981; Meyers, 2004; Dixon, 2006).

One theory about variation in reporting focuses on the concept of newsworthiness and efforts to make content newsworthy. Surette (1998) usefully defined newsworthiness as essentially "...the criteria by which news producers choose which of all known events are to be presented to the public as news events (60)." Chermak (1995) presented some of the earliest evidence that news reporters consciously select crime stories for reporting based on how newsworthy they were. Importantly, Chermak noted that not only are not all crimes newsworthy, even some extreme crimes like homicide were deemed "not interesting enough" to be covered by the media (1998). This further illustrates the shared space of journalist and reader where anticipated reader response can help drive reporting decisions.

Katz (1987) proposes that for something to be newsworthy it must transgress a moral boundary as internalized by society. Increased attention to crime news can produce harsher blameworthiness evaluations for Black suspects compared to White suspects (Dixon, 2008), demonstrating that boundaries of morality are subject to and derivative of other biases in society. This poses difficult and important questions for why certain victims are more sympathetic and certain offenders are perceived as guiltier. We explore these questions here through the lens of criminal law, using vignettes designed to trigger moral judgments, such as drunk driving and illegal immigration.

BLAMEWORTHINESS AND CRIMINAL LAW

Psychological judgments about blame rely on both the harm that the agent causes as well as the mental state of the agent at the time she caused the harm (Cushman, 2008). Thus, two friends who walk out of a bar and who each crash while driving home in the snow are blamed differently depending on the harm they cause. We blame and punish more severely a drunk driver who injures a person than a drunk driver who damages a tree, even if all else is equal (Cushman, 2008). In addition, we blame and punish a person who intentionally causes harm more severely than a person who unintentionally causes the exact same harm (Alicke and Davis, 1989; Alicke et al., 1994; Robbennolt, 2000). These

psychological judgments arise from intuitions about blame and punishment, including attitudes about how severely to punish and for what purpose (Carlsmith et al., 2002; Carlsmith and Darley, 2008; Pizarro and Tannenbaum, 2012; Bilz, 2016).

At the same time, much blame and punishment occur within a social context, imposed by institutions and subject to guidelines or regulations. Governments, schools, firms, and the like typically have formal blame and punishment systems – formal rules are enforced by designated individuals, and the process is highly institutionalized (Cushman, 2014). At the same time, institutionalized blame and punishment relies heavily on our intuitive sense of justice (Robinson and Darley, 1995; Mikhail, 2007). The prototypical example of regularized blame and punishment is the criminal legal system. In criminal law, blameworthiness is codified into law by a set of standards that include the consideration of mens rea, or guilty mind, and actus reus, or wrongful act. Historically derived from Christianity, generally immoral conduct was sufficient to prove mens rea (Robinson, 2002). By the middle of the 13th century, it was well established that "justifiable punishment is premised on and proportional to moral guilt" (Gardner, 1993; p. 655). Historically, punishment was thus intrinsically connected to moral blameworthiness, and contemporary philosophical conceptions of punishment include moral responsibility as a central condition for punishment (Bennett and Brownlee, 2020). While current systems of criminal law have developed into a less explicitly normative inquiry into the offender's state of mind (Nadler, 2022), even contemporary conceptions of mens rea reflects the attachment of moral blame and the offender's state of mind at the time of the offense (Gardner, 1993; Nadler, 2022). Blameworthiness intuitions continue to influence our justice system not only in assigning guilt, but also in prescribing punishment. The degree of resulting harm influences judgments of punishment as well as the perceived wrongfulness of the act, although the magnitude of the resulting influence is the subject of some debate (Cushman, 2008; Kneer and Machery, 2019). Severity of harm does not solely determine punishment, of course – for example, some homicides are punished less severely than others – even if the outcome of death is the same. We see this frequently in the contemporary justice system where we distinguish justifiable and non-justifiable killings, but also divide non-justifiable killings into degrees that call for less punishment based on less intent and mitigating circumstances.

Assessments of severity of harm, the actor's role in causing or contributing to the harm, and the actor's intentionality are not made in a vacuum. Often, judgments of these aspects of an actor's role are made under uncertainty: How much intent did the actor have? How strongly causal was the actor's role in the harm? Alicke's (2000) theory of culpable control posits that when people assess blame, they try to assess how much control the actor exercised over the harm. If an actor intentional conduct directly causes the harm, then the actor is perceived to have high control. But under uncertainty, these perceptions of intent and harm are directly influenced by our initial affective reaction to the harm situation. For example, if John crashes while speeding home to hide an anniversary present for his parents, he is judged less harshly than if he is hiding a vial of cocaine he left out in the open, even though the harm (injuring another driver) and the intentionality (less than intentional, but unreasonably disregarding risk) is the same in both scenarios (Alicke, 1992; see also, Nadler, 2012; Nadler and McDonnell, 2012). John-the-cocaine-hider evoked stronger initial affective reactions, which motivated a desire to understand the conduct as more blameworthy than that of John-the-present-hider. On this account, we engage in "blame validation" – we make blame attributions spontaneously according to how strongly negative our gut reaction is, and then we validate our blame assessment by adjusting evaluations of intention and causation accordingly.

The standard theoretical inputs for punishment and blame judgments – such as intent and severity of harm – are therefore themselves influenced by our perceptions of what kind of person the actor is, including the actor's motives for acting and her character (Uhlmann et al., 2015; Siegel et al., 2017). Alicke's culpable control model posits that we constantly evaluate other people to determine which individuals are trustworthy in the sense of promoting rather than threatening our own physical and psychological well-being (Hieronymi, 2004; Alicke, 2014). According to person-based theories of moral blame, we spontaneously evaluate wrongdoing based on features of the person before having the opportunity to carefully weigh the legally central features of mental state and resulting harm. Evaluating features of the person might include legitimate considerations of motive (e.g., a person driving through a red light to rush someone to the hospital is legitimately blamed less for causing harm than a person engaging in the same conduct to show off for friends). But less legitimate features of the person also influence perceptions of blame, intentionality, and causal role in harm, such as perceived moral character (Nadler, 2012; Nadler and McDonnell, 2012). And other features of the person are completely illegitimate (such as race, national origin, religion) but might nevertheless influence blame and punishment judgments *via* the culpable control pathway posited by Alicke (2000, 2014).

Blame by nature relies on causal responsibility by a human agent, and so invokes a judgment of responsibility that is moral in nature (Coates and Tognazzini, 2012). For this reason, the conduct to which we attach blame reflects poorly on the actor as a moral agent and leads us to infer moral character that lacks loyalty, integrity, or the like (Coates and Tognazzini, 2012). At the same time, prior judgments of moral character can themselves influence degree of blame, as we just discussed.

In the studies reported here, we test the effect of two such person-based factors – one legitimate and one illegitimate – on perceptions of blame and punishment. We do this by cuing morality in vignettes about drinking and driving and illegal immigration, which we describe in further detail below the section "Site of the Research." Moral Attitudes, Blame, and Punishment When an agent causes harm in a context that the public views as morally objectionable, people view the conduct causing harm in a negative light. We saw this earlier in the vignette about John-the-cocaine-hider. Because possession and use of illegal drugs is viewed by many as morally objectionable, John's conduct that led to the accident was viewed negatively. At the same time, when the agent is a member of certain social outgroups (for example, homeless people, undocumented

migrants), that agent is viewed as less competent and trustworthy and their conduct more blameworthy (Fiske, 2018). We next develop examples of morally objectionable conduct (drunk driving) and a morally derogated outgroup (immigrants) that we use to form the basis of the experimental study on assigning punishment that we report below.

Drunk Driving and Moral Attitudes

Fifty years ago, the decision to get behind the wheel of a car after drinking alcohol was considered mostly a matter of personal preference. In the ensuing years, the issue of driving while impaired by alcohol underwent a radical change and moved into the domain of morality. During the 1980s, activists grew the number of local anti-drunk-driving groups from a few dozen to over 400. Their goal was to reduce drunk driving in their respective communities (McCarthy and Wolfson, 1996). Aided by national umbrella organizations, local activists focused on moralization of the issue with the message "You can make a difference" – a slogan plainly designed to appeal to the American ethic of individual responsibility. At the same time, the success of the effort to move drunk driving into the consciousness of the public and into the domain of the moral depended on tapping into and managing intense emotions, like fear. Mothers Against Drunk Driving (MADD) is the highest profile organization of its kind in the United States, and its very name evokes the tragic image of a mother grieving for a dead child, "a threat to something sacred in society: the relationship of mother and child…" (Schmidt, 2014).

The fear of a drunk driving crash in the future presents the looming potential of losing one's own life, losing a loved one, or taking another person's life (Schmidt, 2014). Drunk driving injuries and deaths are shaped into narratives involving a binary moral discourse involving immoral, anti-civil perpetrators acting upon innocent victims. Collectively the acts performed by these individual perpetrators – driving vehicles while under the influence of alcohol – represent a challenge to the moral foundations of society (Schmidt, 2014). At the same time, because drunk driving is a behavior that is ongoing and strikes randomly, there is the possibility that any one of us could become a victim in the future.

Perpetrators of drunk driving accidents are framed as individuals who make a choice: they put the key in the ignition. By choosing to insert the key, the individual is portrayed as choosing not to care about others and instead to put them at risk – a fundamental lack of compassion. The MADD narrative presses us to empathize with the anguish of a mother whose young adult child's life has suddenly ended. The individual who chooses to insert the key after drinking is portrayed as displaying a complete disregard for that anguish. By disregarding this pain and sorrow, the drunk driver is perceived as rejecting this sacred value of motherhood and is rendered a moral monster.

Strong moral reactions can result from harm that is diagnostic of the actor's moral character. For example, a CEO who spent company funds redecorating his office while the company was cutting thousands of jobs provoked public scorn not because the act of redecorating was particularly harmful but because in context the act was seen as indicative of the CEO's character

(Tannenbaum et al., 2011). When evaluating wrongs and harmful acts, people care about what kind of person the actor is: who that person is and not just what they have done (Nadler, 2012; Nadler and McDonnell, 2012). Certain acts are viewed as highly informative of character: these include animal cruelty, racist speech, and to some extent in recent decades, drunk driving, especially when it results in injury or death.

Moral Attitudes Toward Immigrants

In the past few decades, immigration patterns in the United States shifted such that immigrants now live in communities throughout the nation, rather than being concentrated in a handful of regions. Many Americans have negative attitudes toward immigrants as a group – most commonly that immigrants cause problems and should be kept out of the country. At the same time many people hold positive attitudes toward immigrants, including the belief that they are hard-working and enrich American culture. Sometimes these conflicting negative and positive views are held by the same individuals (Ostfeld, 2017). White Americans' attitudes toward immigrants tend to track with their racial attitudes, and individuals who hold more ethnocentric views are more hostile toward immigrants who come from countries outside of Europe (Hainmueller and Hopkins, 2014). Racially resentful whites would like to see restrictions on the flow of immigrants as well as government services denied to immigrants (Kinder et al., 1996; p. 123). Immigrants who entered the country without authorization are viewed negatively, especially by ideological conservatives (Hainmueller and Hopkins, 2014).

Racial resentment among whites increases when the presence of non-whites is perceived to affect their own community. "In the view of many Whites, Blacks in the neighborhood threaten property values and safe schools; Blacks at church violate definitions of community; Blacks at work stir up apprehensions about lost jobs and promotions. At the same time, distance from Blacks allows Whites the luxury of expressing racial tolerance" Kinder and Mendelberg (2000; p. 404). Experimental work has demonstrated that whites are less comfortable with immigrants living near them, working with them, and marrying into their family when those immigrants are depicted as darker skinned compared to when they are depicted as lighter skinned (Ostfeld, 2017). This finding was independent of whether the individual immigrants in question were more assimilated or less assimilated in American culture.

There is a significant literature discussing the morality of immigration, with a particular emphasis on illegal immigration. Importantly, scholars argue that illegal immigration is not always morally wrong depending on the larger belief structures and the incompatibility of multiple legal, social, and protective obligations. For example, if a country limits immigration more than it morally should, the illegal immigration may be a legitimate response rather than a moral breach (Risse, 2008; Taylor, 2008). Many of these writings in law and philosophy tie the moral obligation back to the state, but there is less work analyzing how a layperson in America might interpret the morality of illegal immigration. We do know that Americans are divided on the issue of illegal immigration and that

ways of framing illegal immigration as an issue vary across the country. Discourse in border adjacent regions tends to focus on illegality in immigration (as opposed to immigration more broadly) and to be significantly racialized (Branton and Dunaway, 2009; Ramakrishnan et al., 2010; Merolla et al., 2013). Much of this framing plays out in the news, with different rhetoric and framing characterizing liberal/progressive versus conservative news sources (Merolla et al., 2013), though the changes in laypeople's decision making as a result of those frames is less studied.

SITE OF THE RESEARCH

In this study, we survey readers in state of Minnesota in the United States due to a confluence of salient situational factors and a more general need for increased homicide research outside the largest urban settings[1]. First, we prioritized a location with a relatively high rate of occurrence of vehicular homicides, but that had varied sentencing outcomes. According to the Minnesota Sentencing Commission, while the sentencing guidelines under MN Statute 609.2112 recommend up to 10 years in prison for all vehicular homicide offenders, a substantial portion of vehicular homicide offenders receive stayed sentences or local confinement for a relatively short period (Minnesota Sentencing Guidelines Commission [MSGC], 2016/2017). This wide range primed readers with the realistic ability to make varied choices in punishment outcomes. Second, we chose a location with a standardized type of media coverage, i.e., one main news outlet that covers criminal news across the region. This increases the likelihood that participants will have seen news disseminated in a similar format.

DATA AND METHODOLOGY

This study had two phases of data collection: the purpose of the first phase was to understand the standard formulation of news articles about Minnesota homicides, and in the second we constructed and deployed a vignette experiment. The survey experiment was designed to assess how readers assign punishment to perpetrators along two different dimensions – characteristics of the person (immigrant subject to deportation order, or non-immigrant) and characteristics of the conduct (driving while impaired by alcohol, or not). The phase 1 findings informed the design of the vehicular manslaughter vignettes used in the subsequent experiment[2].

[1]Studies often focus instead on cities that have the most homicide, ostensibly to get a robust picture of homicides overall (see Lattimore, 1997). In our case, we are less interested in homicide as a nationwide phenomenon, so we take this opportunity to focus on an understudied context.

[2]Not all vignette-based work requires as much content analysis and adherence to real-world scenarios as we conducted here. However, in this case, the localized nature of the research required us to replicate reality as closely as possible to approximate news articles with appropriate verbiage, content, and tone. Notably, 66.7% of participants reported reading crime news from Minnesota (the context modeled in the vignettes) sometimes, often, or always, demonstrating the likely familiarity of the participant pool with a particular type of crime news.

Phase 1: Constructing the Experimental Vignettes

Using the Minneapolis Star Tribune, the largest newspaper in Minnesota[3], we gathered 600 articles that met our criteria for potentially being about a homicide[4]. We screened the articles for relevance and established a 3-month cut point for analysis, leaving us with a final corpus of 177 test articles. We examined a 3-month period (March 18, 2019–June 18, 2019) in which we coded 110,250 words of text in 177 articles, covering 83 separate cases and 93 victims (seven cases involved multiple victims) of homicide.

We collected metadata about each article including date of publication, article title, author, and total word count. We also collected case-level information about the number of actors, the type of killing, any specific homicide-related charges, and the location of the incident. Finally, we also collected victim-level and offender-level information like age, gender, race, and the relationship between victim and offender.

We used the information gleaned from the corpus of 177 news articles to design our experimental vignettes. In our population of articles, victim and offender gender were mentioned a vast majority of the time (86.44 and 85.31% of the articles, respectively). The age of the offender was also usually mentioned (79.66% of articles), though the age of victims was reported only about half the time (53.11% of articles). It was much less common for race to be mentioned in the article with offender race mentioned around 17.51% of the time and victim race mentioned 18.64% of the time. Consequently, in our manufactured vignette we opted to report both victim and offender gender, offender age and one victim's age, and no race information.

The most common type of killings reported in this period were shootings (42) and vehicular manslaughter (24). While we considered selecting shootings for our vignettes, we instead chose vehicular manslaughter because it lacks many confounding characteristics of other homicide types. In vehicular manslaughters there are less frequently pre-existing relationships between parties, neighborhood effects, or complicated motives that might not be clear from a news article in vehicular homicide cases. The fact that nearly 1/4 of homicides in the 3-month period were vehicular indicated that this time of crime would be plausible in the Minnesotan context. Importantly, vehicular manslaughter can also be framed as purely accidental or as accidental with compounding factors which gave us more flexibility in designing the vignettes.

In conducting a close code of all 177 articles we were also able to familiarize ourselves with the verbiage used in reporting about vehicular manslaughter. To replicate actual news stories as closely as possible, we selected two articles which formed the basis for our experimental vignettes (see **Supplementary Material**). We designed three vignettes derivative of the same vehicular manslaughter scenario (see **Supplementary Material**). The scenarios are as similar as possible in wording and keep

[3]The Star Tribune has a daily circulation of 288,315, a Sunday circulation of 581,063, and a digital subscription rate of 50,000.

[4]Using the World Access News Database, we used one inclusive Boolean search function gather articles (kill* OR homicide* OR slay* OR murder*).

TABLE 1 | News engagement descriptives (%).

	Read news	Read MN crime news	Watch TV news
Never	0.57	2.84	14.2
Rarely	10.8	20.45	30.11
Sometimes	36.36	39.2	22.73
Often	38.07	27.27	23.86
Always	14.2	10.23	9.09
N	176	176	176

offender and conduct characteristics constant excluding the key experimental manipulations. In the first scenario, we offered the basic information about the criminal event and use this as our control scenario. In the second scenario, we added information about the perpetrator having an elevated blood-alcohol content level and history of drunk driving. In the final scenario, we omitted the alcohol related information, but instead informed the reader that the perpetrator was an immigrant who had entered the country illegally 10 years prior and was set to be deported[5]. Our goal in choosing these three experimental vignettes was to examine the effects of conduct (drunk driving) and denigrated group membership (immigrant unlawfully present) on blameworthiness and punishment.

Phase 2: Deployment on Amazon Mechanical Turk

We conducted our survey on Amazon Mechanical Turk, requiring the 191 participating Turkers to have above a 95% HIT rating and to be located in Minnesota[6]. We further confirmed their presence in the state of Minnesota by collecting the first three digits of each Turkers zip code at the end of the survey. While not a perfect proxy for residency, restricting the geography of participants makes it substantially more likely that participants would have been exposed to Minnesota crime media. We confirmed this by asking if participants had ever read news stories about crime in Minnesota, to which only 2.84% of respondents indicated that they never had (see **Table 1**). Participants were asked to read one of the three randomly assigned experimental vignettes and respond to questions about punishment, news consumption, and demographics.

Independent Variables

The key manipulated variable was the potential blameworthiness of the vehicular homicide offender. We used three scenarios to re-design the news vignettes: control, driving under the

influence (DUI), and immigration. In each scenario we altered only the blameworthiness information, holding all other facts about the incident constant. In the control vignette, we gave only basic information about the nature of the accident and the outcome. In the DUI condition, we included information about the elevated blood alcohol content (BAC) level of the offender. In the immigration vignette, we included information about the immigration history of the perpetrator, specifically that they immigrated to the United States illegally as a minor many years ago.

We measured a variety of demographic and related variables including gender, educational attainment, income, age, race, Hispanic ethnicity, and political views. Participants in our study were more likely to be male (56.02%) than female (43.43%). Nearly half had a bachelor's degree (46.59%) and 85.14% of them described themselves as white. Around 60% of the participants made between $35,000 and $100,000 per year and were between the ages of 25 and 44 (full descriptives can be found in **Supplementary Material**). Importantly, we also asked participants to indicate their political views using a sliding scale from 0 to 100, with 0 being very conservative and 100 being very liberal. The sample skewed slightly liberal with a mean response of 59.3, though the standard deviation was large (29.73).

Key Dependent Variable

The key dependent variable in this analysis is the extent of punishment assigned to the hypothetical offender. Each participant was shown a slider and asked to assign a number of years of punishment between 0 and 10. While the numbers may be conceptually meaningful, we also want to focus on the behavior inherent to the response pattern. That is, a selection of "10" means something beyond just 10 years of punishment, it means the maximum punishment allowable. We use duration of punishment as a measurable proxy for the idea of blameworthiness, that is, the idea that some perpetrators deserve more punishment than others even if the outcome of the criminal act is the same. In this study, we keep the outcome of the scenario constant, only varying factors that might affect the level of culpability on the part of the perpetrator.

We are reasonably confident in our assertion that we can interrogate perceptions of morality using years of suggested punishment due to internal validity checks undertaken in the study design. In addition to the punishment question described above, we also asked participants to indicate their perception of the moral character of the driver on a seven-point Likert scale. These morality assessments were 54.71% correlated with suggested years of punishment, suggesting substantial conceptual overlap. In a simple linear regression model predicting years of punishment using the morality assessment we found a strong significant relationship ($P < 0.00$) and an R^2 value of 0.30 again suggesting significant overlap between the two measures (see **Supplementary Material** for additional details and tabular representations).

We also included several other measures in the survey in order to collect additional information to contextualize the punishment responses. We asked participants about their news consumption, specifically how often they read news articles, watch the news on

[5]Note that this detail modeled on an actual case in Minnesota (see **Supplementary Material**). In this case, Jose O. Vasquez-Guillen was later deported, and a stream of mainstream and partisan media described Vasquez-Guillen in various ways that highlighted his immigration status including referring to him as "Salvadoran man," "undocumented," and an "illegal alien with deportation order" in news headlines. Interestingly, other headlines referred to him more generally as a "St. Paul resident."

[6]In order to ensure data quality, we included a short series of questions asking participants about their familiarity with a real-life case, then them to explain what happened in that case in words, and then asked them to evaluate the outcome as fair/unfair/not sure. Nine participants were removed from the final analysis because they provided incompatible or non-sensical responses.

TABLE 2 | Suggested punishment duration.

	N	Mean	Standard deviation
Control	55	5.37	3.47
DUI	62	9.19	1.52
Immigrant	75	7.54	3.06

television, and read Minnesota crime news articles specifically. We also surveyed participants about a recent police shooting case in Minnesota that dominated news headlines, both to give context to participants' understanding of the news and some of their opinions about fairness and justice[7].

RESULTS

Punishment duration varied greatly by conduct and characteristic (see **Table 2**). In the control vignette, which included information only about the event and not the driver, respondents chose a punishment duration of 5.37 years on a scale of 0–10. This regression to the mid-point makes sense, given the limited information. However, when exposed to the DUI vignette the respondents assigned the driver a more punitive 9.19 years of prison on average. Interestingly, participants assigned 7.54 years of prison in the illegal immigration condition, reflecting a judgment in between the control condition and the DUI condition.

We estimated separate linear regression models for each vignette type in order to understand how demographic factors and self-identified political views may impact punishment evaluations (**Table 3**). We found that none of the demographic factors predicted punishment duration in the control vignette, which is not particularly surprising given that the vignette contained very little information to potentially evoke differential responses. In the DUI vignette, respondent political views had some directional effects that approached significance, but none of the provided demographic variables significantly predicted punishment duration. This is consistent with literature suggesting the drunk driving is unanimously disparaged. Finally, in the immigration vignette, we found that only self-identified political views had a significant impact on punishment duration ($p < 0.01$). As self-identified political views became more conservative, suggested punishment duration increased.

In **Figure 1**, we plot the adjusted linear prediction of years of punishment by vignette type with a specific focus on political views, reversing the scale so that the left side of the x axis represents liberal identification and the right side represents conservative, for ease of visualization. We find that the slope of punishment across the control condition is flat across all ranges of political views. Consistent with our regression results, we see some effects of conservative political

---

[7]We also asked two alternative questions that measured similar concepts to the main punishment measure. We asked participants if they thought the driver should get more or less punishment than average (5-point Likert scale) and about the moral character of the driver (7-point Likert scale). We do not focus on these measures here, but find that they follow the same general patterns reported in **Table 2** (below).

TABLE 3 | Regression predicting years of punishment by vignette type.

Variables	Control	DUI	Immigrant
Political views	−0.01 (0.02)	−0.02+ (0.01)	−0.04** (0.02)
Income			
Less than 10,000	−0.13 (5.26)	−1.18 (1.21)	−1.63 (2.89)
200,000 or more	−0.79 (7.58)	−0.28 (1.43)	3.58 (2.68)
Education			
High school/GED	−0.03 (2.36)	1.68 (0.72)	3.41 (1.82)
Some college	−0.91 (1.32)	0.17 (0.50)	1.60 (1.03)
Gender			
Male	−1.03 (1.22)	−0.43 (0.41)	−0.56 (0.94)
Race			
Black	−4.07 (5.38)	2.09 (1.95)	4.01 (2.42)
White	−5.06 (4.52)	0.51 (1.36)	1.05 (1.92)
Ethnicity			
Hispanic	5.24 (3.82)	0.62 (1.61)	−0.42 (2.56)
Age			
20–24	−2.17 (3.64)	1.29 (1.26)	−1.48 (3.84)
60–64	−1.27 (5.14)	0.40 (2.20)	1.23 (4.63)
Constant	12.02 (7.82)	9.54 (1.67)	8.40 (3.20)
# of observations	54	60	73

$+p < 0.10$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.
Reported as regression coefficients with standard errors in parentheses. Insignificant values redacted for visual clarity, see **Supplementary Material**.

views on increased punishment in the DUI condition but find that suggested punishment in this condition is much higher all along the spectrum of self-identified political views. Also consistent with the regression results is the much larger positive slope in the immigration condition. In fact, at the furthest tail of self-identified conservative views predicted punishment duration scores in the immigration vignette and DUI vignette are not statistically different from each other. This means that the participants who self-identified as the most conservative perceived that an immigrant driver unlawfully present in the country who caused death deserved the same punishment enhancement as a drunk driver who caused death.

We report a tabular representation of these average marginal effects in **Table 4**, showing the ranges of confidence intervals for each vignette type.

We also plot the conditional marginal effects of political views on linear predictions of punishment duration with a 95% confidence interval, confirming the results above (**Figure 2**). In this visual depiction behavior at the tails of the distribution is shown to be highly differentiated, with self-identified liberal views assigning punishment in the control and immigration conditions very similarly, while respondents with self-identified conservative views seemed to assign punishment more similarly between the DUI and immigration conditions.

DISCUSSION

Our results show that news consumers assign blameworthiness differently for the same criminal incident depending on what they learn about the conduct of the perpetrator (here, drunk driving)
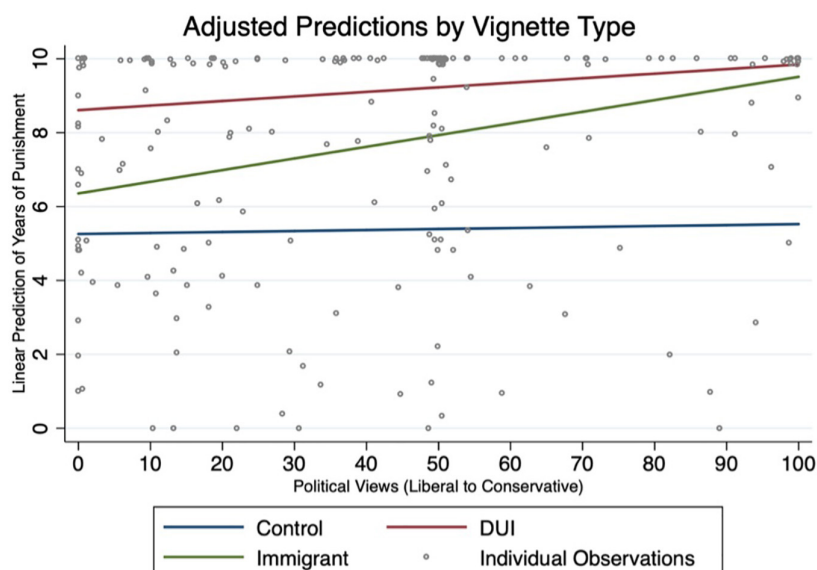
FIGURE 1 | Adjusted predictions by vignette type.

as well as the status of the perpetrator (here an immigrant present unlawfully). When we presented readers with the control vignette, which included no cues about immigration status or impaired driving, respondents selected punishment durations of a little over 5 years, reflecting moderately serious punishment. We argue that this relatively lower amount of punishment is reflective of a lack of moral cuing that was presented in the two other versions of the vignette. In the absence of any detail about circumstances, readers conceptualized the death as closer to an accident, because the perpetrator culpability is not specified by any moral characteristic of the person or the behavior. When we used predictive modeling, we found no significant demographic patterns in reader responses. This lack of influence of demographic characteristics suggests that we successfully retracted any moral cuing information from the control vignette that would prompt differential decision-making.

In contrast, in the DUI vignette, where we specify deviant behavior that has been entrenched as immoral (Schmidt, 2014) we see mean punishment substantially increased to more than 9 years of prison time. We want to stress that participants were not just choosing a particular number of years, rather they were selecting within a given range. That means that participants on average assigned close to the maximum amount

of punishment allowed in this scenario. Once again, we do not find that any particular demographic characteristic is predictive of recommended punishment. This second set of null findings again conforms to findings in the literature indicating that drunk driving gives rise to moral outrage, and this response has become culturally pervasive enough to nullify potential group differences.

In the immigration vignette, we see something different, where there is substantial variation across participants regarding punishment and moral blameworthiness. As we demonstrate in **Figure 1**, readers with more liberal political views (closer to 0) selected a punishment duration much closer to the control condition, where readers with conservative views (closer to 100) selected a punishment duration much closer to the DUI condition. There are several components that we think might help explain this difference in punishment assignment. First, the issue of illegal immigration in the United States is in many ways a partisan issue with research postulating that this political entrenchment has grown in recent years (Dionne et al., 2008). Therefore, differential assignment of punishment by political views on a polarizing political issue is not altogether surprising. What is more interesting is the particular context in which it occurs. Importantly, there was nothing different about the conduct of the driver in the control vignette and immigrant vignette, yet the proscribed punishments were very different[8]. This implies that the same offense committed by someone without legal immigration status is perceived as more blameworthy than the same crime committed by someone who is not identified as lacking legal immigration status. This

TABLE 4 | Average marginal effects by vignette type, political views.

	dy/dx	Standard error	$P > |t|$	95% confidence interval	
Vignette type					
Control	0.003	0.013	0.831	−0.022	0.027
DUI	0.012	0.013	0.323	−0.012	0.037
Immigrant	0.032	0.011	0.003	0.011	0.053

N = 190.

[8]It is feasible that participants were concluding that someone without legal immigration status would not have a driver's license, making their criminal circumstances worse. However, we feel it is unlikely that this consideration explains the large amount of increased punishment assigned primarily by self-identified conservatives.
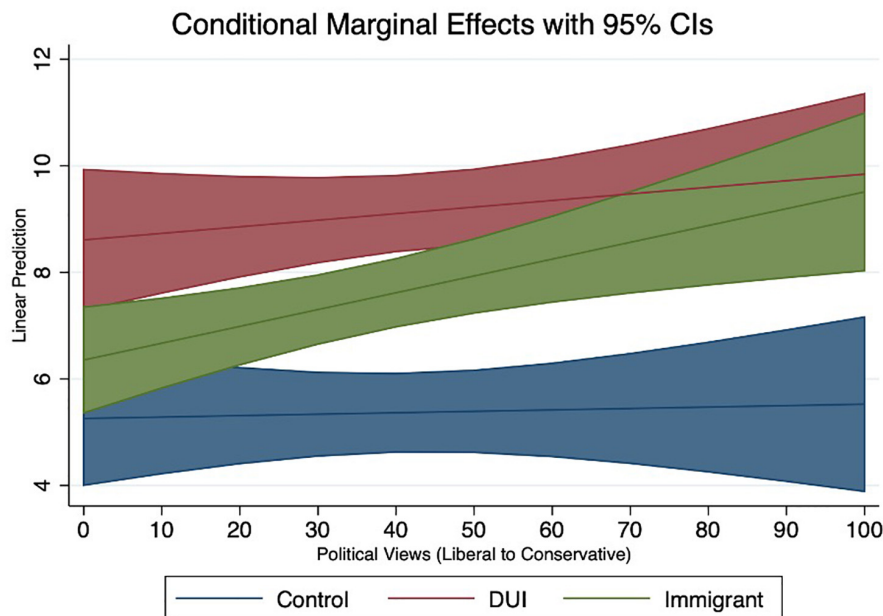
FIGURE 2 | Conditional marginal effects of political views relative to control.

difference represents a very tangible consequence to differing interpretations of morality. This finding in particular merits future study to understand how political views may impact ultimate consequences for defendants in the criminal justice system, especially lawyers, judges, and laypeople involved in the justice system (i.e., juries) may bring their political ideologies into the courtroom.

Importantly, we did not assign an ethnicity to the driver, but rather only noted that he immigrated illegally as a minor many years ago. This likely presents a race cue of some kind, so the immigration could be proxying for racial resentment which has been shown to impact beliefs about illegal immigration (Hainmueller and Hopkins, 2014). Another possibility is that the difference in punishment is measuring the distinct but related concept of xenophobia.

These possibilities are especially salient in the Minnesotan context. The largest two immigrant communities in Minnesota are from Mexico (about 64,500 foreign-born Minnesotans) and Somalia (about 33,500 foreign-born Minnesotans) (Minnesota State Demographics Center [MSDC], 2018). So, the blameworthiness differences we observe might result from anti-Mexican racism and/or a version of anti-Black racism. In our study we collected that our Minnesota participants were conscious of race and national origin around the time they participated in this survey.

To get a sense of how participants understood crime and culpability in their community, after responding to the experiment vignette we asked them if they were familiar with the recent case of Mohamed Noor and Justine Damond. This case made headlines when Noor, an immigrant Somali police officer, mistakenly shot the unarmed Australian native Justine Damond who had called 911 to report a suspected sexual assault. Noor

was found guilty of third-degree murder and manslaughter and sentenced to 12.5 years in prison, a marked difference in criminal justice outcomes compared to other police officers who killed civilians (Jackson, 2019). Notably, 1 year earlier, Minneapolis officer Jeronimo Yanez was acquitted of the Killing of Philando Castile (Jackson, 2019). When asked if they were familiar with the Noor case, 58.12% of participants said they were at least a little familiar. When asked about whether or not the verdict was fair participants were divided (34.74% believed it was fair, 12.11% believed it was not fair, and 53.16% were not sure) and themselves brought up the issues of race and immigration status. One respondent wrote:

"The facts in that case were not significantly different than other cop involved shootings in which the cop was exonerated. There was a feeling of racial undertones to the conviction."

This represents a common theme among respondents: not necessarily a belief that Noor was innocent, but rather than inequality in the criminal justice based on race led to an unfair overall outcome. Participants struggled to choose a dichotomous marker of "fair" but were able to articulate agreement with a guilty verdict – without endorsing the broader system of punishment.

Another respondent compared the Damond case directly to the case of Castile saying:

"I think he should do SOME time, but not that much. Yes, he killed her. He didn't listen to her. He didn't follow training or protocol. However, other cops in the TCs (Twin Cities) have shot black, Hmong, Indian people, etc., and were not sentenced. If this cop is getting 12.5, the one that shot Philando Castile should have gotten 25."

This respondent carefully articulates a disparity in blameworthiness relative to other cases that they conceptualize as similar. That is not to say that respondents were all in agreement.

Many focused-on Noor as "trigger-happy" or articulated a belief that police officers should be held to a higher standard. Specific mentions of race or immigration status were generally avoided by participants who positively endorsed the outcome of the case, excluding one participant who suggested that:

"In my opinion he should have been deported back to his country with no chance of reentry."

These responses demonstrate patterns in assessing blameworthiness mentally – but also in articulating blameworthiness around race. Further testing with a similar vignette design could more directly test these possibilities.

This study is limited in its generalizability given our focus on to vehicular homicides in the state of Minnesota. Future research should expand crime types and social contexts to examine whether these patterns are stable. Additionally, this analysis also only makes use of varying information about the offender (driver). Future work should consider varying the victim characteristics to more effectively measure the dyadic bias potentials between victim and offender.

This study advances knowledge about the role of news media in constructing popular perceptions moral guilt. All the scenarios we presented here were derivative of the same set of base facts. Moreover, both factors tested might have been present, simultaneously, about the actual incident, and the decision about whether and how to include either aspect in the story would be in the discretion of the writer. In other words, just because a driver had an elevated BAC level does not guarantee a news article reports on it, which may change the guilt perception of the perpetrator in that case. Evoking Schudson (2011), we do not mean to suggest that intentional misrepresentation by news writers causes distorted perceptions. Rather, a different portrayal of the truth for any number of reasons (unknown facts, facts perceived to be uninteresting or not newsworthy, limits on length, etc.) can change the contents of news unbeknownst to news readers. In the case of our sample, nearly all had read crime news before and a vast majority in the specific context of Minnesota. This ubiquity further explains the amplified importance of context in crime news. Even if news readers are not called to make direct decisions about a particular crime they read about in the news, the cumulative consequences of news can lead to racial stereotyping, fostering inaccurate fear of crime, and reifying mis-perceptions of who commits crime do affect everyone in society (Barlow et al., 1995; Gilliam et al., 1996; Sorenson et al., 1998; Thorson, 2001; Boulahanis and Heltsley, 2004).

CONCLUSION

The construction of news stories can substantially influence readers' judgments about blame and punishment for vehicular homicide offenders. By varying moral cues from neutral to negative in the same scenario, we demonstrate that readers select punishments around the mid-point when they lack information and select higher levels of punishment for universally condemnable moral behavior like drinking and driving. When faced with a morally controversial piece of information, like immigration status, we find that readers with differing political views assign different amounts of punishments. This finding underscores the importance of how news writing and presentation matters and how its influence can vary sharply according to pre-existing moral and political commitments of the reader.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Northwestern University IRB. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

KA: conceptualization, design, data collection, data analysis, writing, and revising. JN: conceptualization, data analysis, writing, and revising. Both authors contributed to the article and approved the submitted version.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.784428/full#supplementary-material

REFERENCES

Alicke, M. (2014). Evaluating blame hypotheses. *Psychol. Inq.* 25, 187–192. doi: 10.1080/1047840X.2014.902723

Alicke, M. D. (1992). Culpable causation. *J. Pers. Soc. Psychol.* 63, 368–378. doi: 10.1037/0022-3514.63.3.368

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychol. Bull.* 126, 556–574. doi: 10.1037/0033-2909.126.4.556

Alicke, M. D., and Davis, T. L. (1989). The role of a posteriori victim information in judgments of blame and sanction. *J. Exp. Soc. Psychol.* 25, 362–377. doi: 10.1016/0022-1031(89)90028-0

Alicke, M. D., Davis, T. L., and Pezzo, M. V. (1994). A posteriori adjustment of a priori decision criteria. *Soc. Cogn.* 12, 281–308. doi: 10.1521/soco.1994.12.4.281

Barlow, M. H., Barlow, D. E., and Chiricos, T. G. (1995). Economic conditions and ideologies of crime in the media: a content analysis of crime news. *Crime Delinq.* 41, 3–19.

Bennett, C., and Brownlee, K. (2020). "Punishment," " in *The Cambridge Companion to the Philosophy of Law*, ed. J. Tasioulas (Cambridge: Cambridge University Press), 253–270. doi: 10.1017/9781316104439.014

Berkowitz, D. A. (1997). *Social Meanings of News: A Text-Reader*. Thousand Oaks, CA: SAGE Publications.

Bilz, K. (2016). Testing the expressive theory of punishment. *J. Empir. Legal Stud.* 13, 358–392. doi: 10.1111/jels.12118

Boulahanis, J. G., and Heltsley, M. J. (2004). Perceived fears: the reporting patterns of juvenile homicide in chicago newspapers. *Crim. Justice Policy Rev.* 15, 132–160. doi: 10.1177/0887403404263624

Branton, R. P., and Dunaway, J. (2009). Spatial proximity to the u.s.—mexico border and newspaper coverage of immigration issues. *Polit. Res. Q.* 62, 289–302. doi: 10.1177/1065912908319252

Carlsmith, K. M., and Darley, J. M. (2008). "Psychological aspects of retributive justice," in *Advances in Experimental Social Psychology*, ed. M. P. Zanna (Cambridge, MA: Academic Press), 193–236. doi: 10.1016/s0065-2601(07)00004-4

Carlsmith, K. M., Darley, J. M., and Robinson, P. H. (2002). Why do we punish?: deterrence and just deserts as motives for punishment. *J. Pers. Soc. Psychol.* 83, 284–299. doi: 10.1037/0022-3514.83.2.284

Chermak, S. M. (1995). *Victims in the News: Crime and the American News Media*. Boulder, CO: Westview Press.

Coates, D. J., and Tognazzini, N. A. (2012). The nature and ethics of blame. *Philos. Compass* 7, 197–207. doi: 10.1111/j.1747-9991.2012.00477.x

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. (2014). The scope of blame. *Psychol. Inq.* 25, 201–205. doi: 10.1080/1047840X.2014.904692

Dionne, E. J. Jr., and Suro, R. (2008). *Migrating Attitudes, Shifting Opinions: The Role of Public Opinion in the Immigration Debate. A Report on the Media and the Immigration Debate*. Washington, DC: Governance Studies at Brookings.

Dixon, T. L. (2006). Psychological reactions to crime news portrayals of black criminals: understand the moderating roles of prior news viewing and stereotype endorsement. *Commun. Monogr.* 73, 162–187. doi: 10.1080/03637750600690643

Dixon, T. L. (2008). Crime news and racialized beliefs: understanding the relationship between local news viewing and perceptions of african americans and crime. *J. Commun.* 58, 106–125. doi: 10.1111/j.1460-2466.2007.00376.x

Dorfman, L., Thorson, E., and Stevens, J. E. (2001). Reporting on violence: bringing a public health perspective into the newsroom. *Health Educ. Behav.* 28, 402–419. doi: 10.1177/109019810102800402

Fishman, M. (1988). *Manufacturing the News*. Austin, TX: University of Texas Press.

Fiske, S. T. (2018). Stereotype content: warmth and competence endure. *Curr. Direct. Psychol. Sci.* 27, 67–73. doi: 10.1177/0963721417738825

Gardner, M. R. (1993). The Mens Rea Enigma: Observations on the Role of Motive in the Criminal Law Past and Present. Utah Law Review. 635.

Gieber, W. (1964). "News is what newspapermen make it," in *People, Society and mass Communication*, eds D. M. White and L. A. Dexter (New York, NY: Free Press), 173–192.

Gilliam, F. D. Jr., Iyengar, S., Simon, A., and Wright, O. (1996). Crime in black and white: the violent, scary world of local news. *Harvard Int. J. Press Polit.* 1, 6–23. doi: 10.1177/1081180x96001003003

Graber, D. A. (1979). Is crime news coverage excessive? *J. Commun.* 29, 81–92. doi: 10.1111/j.1365-2850.2006.00953.x

Hainmueller, J., and Hopkins, D. J. (2014). Public attitudes toward immigration. *Annu. Rev. Polit. Sci.* 17, 225–249.

Hieronymi, P. (2004). The force and fairness of blame. *Philos. Perspect.* 18, 115–148. doi: 10.1111/j.1520-8583.2004.00023.x

Humphries, D. (1981). Serious crime, news coverage, and ideology: a content analysis of crime coverage in a metropolitan paper. *Crime Delinq.* 27, 191–205. doi: 10.1177/001112878102700202

Jackson, J. (2019). *Mohamed Noor's Sentence Raises Uncomfortable Questions About Race*. Atlanta, GA: CNN.

Katz, J. (1987). What makes crime news? *Media Cult. Soc.* 9, 47–75. doi: 10.1177/016344387009001004

Kinder, D. R., and Mendelberg, T. (2000). "Individualism reconsidered: principles and prejudice in contemporary american opinion," in *Racialized Politics: The Debate about Racism in America*, eds D. O. Sears, J. Sidanius, and L. Bobo (Chicago, IL: University of Chicago Press), 44–74.

Kinder, D. R., Sanders, L. M., and Sanders, L. M. (1996). *Divided by Color: Racial Politics and Democratic Ideals*. Chicago, IL: University of Chicago Press.

Kneer, M., and Machery, E. (2019). No luck for moral luck. *Cognition* 182, 331–348. doi: 10.1016/j.cognition.2018.09.003

Lattimore, P. K. (1997). *Homicide in Eight US Cities: Trends, Context, and Policy Implications: An Intramural Research Project*. Washington, DC: US Department of Justice, Office of Justice Programs, National Institute of Justice.

Lu, S. (2012). Toward an integrated model: the influence of presumed media influence in news production. *Paper Presented at the 65th Annual Conference of World Association for Public Opinion Research*, Hong Kong.

McCarthy, J. D., and Wolfson, M. (1996). Resource mobilization by local social movement organizations: agency, strategy, and organization in the movement against drinking and driving. *Am. Sociol. Rev.* 61, 1070–1088. doi: 10.2307/2096309

Merolla, J., Ramakrishnan, S. K., and Haynes, C. (2013). "Illegal," "undocumented," or "unauthorized": equivalency frames, issue frames, and public opinion on immigration. *Perspect. Polit.* 11, 789–807.

Meyers, M. (2004). African american women and violence: gender, race, and class in the news. *Crit. Stud. Media Commun.* 21, 95–118. doi: 10.1080/07393180410001688029

Mikhail, J. (2007). Universal moral grammar: theory, evidence and the future. *Trends Cogn. Sci.* 11, 143–152. doi: 10.1016/j.tics.2006.12.007

Minnesota Sentencing Guidelines Commission [MSGC] (2016/2017). *CVH/I Offenses: Recap Report*. Saint Paul, MN: Minnesota Sentencing Guidelines Commission.

Minnesota State Demographics Center [MSDC] (2018). *Immigration and Language: Key Facts Refers to 2018 American Community Survey*. Saint Paul, MN: Minnesota State Demographics Center.

Nadler, J. (2012). Blaming as a social process: the influence of character and moral emotion on blame. *Law Contemp. Problems* 75, 1–31.

Nadler, J. (2022). *Criminal Law & Psychological Blame*. Available online at: https://ssrn.com/abtract=4007663 (accessed January 14, 2022).

Nadler, J., and McDonnell, M. H. (2012). Moral character, motive, and the psychology of blame. *Cornell Law Rev.* 97, 255–304.

Norman, J. (2018). *American's Concerns About National Crime Abating*. Washington, DC: Gallup.

Ostfeld, M. (2017). The backyard politics of attitudes toward immigration. *Polit. Psychol.* 38, 21–37. doi: 10.1111/pops.12314

Pan, Z., and Kosicki, G. M. (1993). Framing analysis: an approach to news discourse. *Polit. Commun.* 10, 55–75. doi: 10.1080/10584609.1993.9962963

Pizarro, D. A., and Tannenbaum, D. (2012). "Bringing character back: How the motivation to evaluate character influences judgments of moral blame," in *The social psychology of morality: Exploring the causes of good and evil*, eds M. Mikulincer and P. R. Shaver (Washington, DC: American Psychological Association), 91–108. doi: 10.1037/13091-005

Purcell, K., Rainie, L., Mitchell, A., Rosenstiel, T., and Olmstead, K. (2010). Understanding the participatory news consumer. *Pew Internet Am. Life Project* 1, 19–21. doi: 10.1371/journal.pone.0206825

Robinson, P. (2002). *Mens Rea. Faculty Scholarship at Penn Law*.

Ramakrishnan, S. K., Esterling, K. M., Neblo, M., and Lazer, D. (2010). Illegality, national origin cues, and public opinion on immigration. *Paper Presented at the Annual Meeting of The Midwest Political Science Association*, Chicago, IL.

Risse, M. (2008). On the morality of immigration. *Ethics Int. Affairs* 22, 25–33.

Robbennolt, J. K. (2000). Outcome severity and judgments of 'responsibility: a meta-analytic review1. *J. Appl. Soc. Psychol.* 30, 2575–2609. doi: 10.1111/j.1559-1816.2000.tb02451.x

Robinson, P. H., and Darley, J. M. (1995). *Justice, Liability, and Blame: Community Views and the Criminal Law.* Boulder, CO: Westview Press.

Schmidt, I. B. (2014). Perpetual trauma and its organizations: mothers against drunk driving and drunk driving revisited. *Memory Stud.* 7, 239–253. doi: 10.1177/1750698012470836

Schudson, M. (2011). *The Sociology of News.* New York, NY: WW Norton & Company.

Shoemaker, P. J. (2006). News and newsworthiness: a commentary. *Communications* 31, 105–111.

Siegel, J. Z., Crockett, M. J., and Dolan, R. J. (2017). Inferences about moral character moderate the impact of consequences on blame and praise. *Cognition* 167, 201–211. doi: 10.1016/j.cognition.2017.05.004

Sorenson, S. B., Manz, J. G., and Berk, R. A. (1998). News media coverage and the epidemiology of homicide. *Am. J. Public Health* 88, 1510–1514. doi: 10.2105/ajph.88.10.1510

Surette, R. (1998). *Media, Crime, and Criminal Justice: Images and Realities.* New York, NY: Wadsworth Publishing.

Tannenbaum, D., Uhlmann, E. L., and Diermeier, D. (2011). Moral signals, public outrage, and immaterial harms. *J. Exp. Soc. Psychol.* 47, 1249–1254. doi: 10.1016/j.jesp.2011.05.010

Taylor, M. R. (2008). Illegal immigration and moral obligation. *Public Affairs Q.* 22, 29–41.

Thorson, E. (2001). The reporting of crime and violence in the los angeles times: is there a public health perspective? *J. Health Commun.* 6, 169–182. doi: 10.1080/10810730120636

Uhlmann, E. L., Pizarro, D. A., and Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspect. Psychol. Sci.* 10, 72–81. doi: 10.1177/1745691614556679

frontiers
in Psychology

Judgment and Embodied Cognition of Lawyers. Moral Decision-Making and Interoceptive Physiology in the Legal Field

Laura Angioletti[1,2]*, Federico Tormen[1,2] and Michela Balconi[1,2]

[1]International Research Center for Cognitive Applied Neuroscience (IrcCAN), Università Cattolica del Sacro Cuore, Milan, Italy, [2]Research Unit in Affective and Social Neuroscience, Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy

Past research showed that the ability to focus on one's internal states (e.g., interoceptive ability) positively correlates with the self-regulation of behavior in situations that are accompanied by somatic and/or physiological changes, such as emotions, physical workload, and decision-making. The analysis of moral oriented decision-making can be the first step for better understanding the legal reasoning carried on by the main players in the field, as lawyers are. For this reason, this study investigated the influence of the decision context and interoceptive manipulation on the moral decision-making process in the legal field gathering the responses of two groups of lawyers. A total of 20 lawyers were randomly divided into an experimental group (EXP), which was explicitly required to focus the attention on its interoceptive correlates, and a control group (CON), which only received the general instruction to perform the task. Both groups underwent a modified version of the Ultimatum Game (UG), where are presented three different moral conditions (professional, company, and social) and three different offers (fair, unfair, and equal). Results highlighted a significant increase of Acceptance Rate (AR) in those offers that should be considered more equal than fair or unfair ones, associated with a general increase of Reaction Times (RTs) in the equal offers. Furthermore, the interoceptive manipulation oriented the Lawyers toward a more self-centered decision. This study shows how individual, situational, contextual, and interoceptive factors may influence the moral decision-making of lawyers. Future research in the so-called Neurolaw field is needed to replicate and expand current findings.

Keywords: legal reasoning, moral decision-making, lawyers, ultimatum game, interoception

INTRODUCTION

While moral choice behavior has received much attention in economics and psychology, it is rarely considered in the decision-making process applied to law. However, investigations into moral, regulatory, and decision-making judgments regarding persons involved in judicial proceedings are becoming more and more common in literature, while there are not many

investigations regarding the main actors of justice and law, such as judges and lawyers (Goodenough and Zeki, 2006; Danziger et al., 2011; Tormen, 2020).

The analysis of the cognitive processes that lead to regulatory and legal reasoning by legal professionals has fundamental importance. And it is precisely the analysis of moral-oriented decision-making process that is the first step to better understand the legal reasoning carried out by the main players in the field, as lawyers are. Recently, a consensus view has emerged, which recognizes important roles for emotion and intuition, and which suggests that normative judgment is a distributed process in the brain (Goodenough and Prehn, 2004).

It was previously assumed that decision-making informed by the legal norm, as an expression of normative morality within a given culture, must necessarily be informed by cognitive processes strongly influenced by emotional components (Haidt, 2001). To better comprehend the moral decision-making process, the context of the Ultimatum Game (UG) permits us to investigate what is considered one of the pillars of human morality: fairness. Fairness is chiefly investigated in the context of the UG, an extensively studied game in psychology, neuroscience, philosophy, and behavioral economics (Nobandegani et al., 2020). The UG has a simple design: two players, the proposer and the responder, have to agree on how to split a sum of money. Proposer makes an offer. If the responder accepts, the deal goes ahead; if the responder rejects, neither player gets anything. In both cases, the game is over.

Current findings on embodied cognition support the view that the body and the mind are inextricably linked in the production of cognitions (Häfner, 2013). According to embodied cognition theories, higher cognitive processes entail reactivations of sensory-motor states that occur during the experience with the world (Pace-Schott et al., 2019). Similarly, emotional experience and cognitive functioning are strongly linked to the activation of interoceptive representations and meta-representations of body signals that promote interoceptive awareness (Herbert and Pollatos, 2012; Angioletti and Balconi, 2020; Balconi and Angioletti, 2022). Also, the neuroanatomic basis of interoception constitute the correlate of the body in the mind and the mechanisms that allow affective and cognitive activities to be embodied (Craig, 2002).

To investigate the deeper dynamics of moral decisions, this study highlights the value of interoception as a factor that influences the decision-making process also in the legal field. The literature suggests a relation between the rationality of decision making and the interoception construct (Sugawara et al., 2020), conceived as the perception of afferent information that arises from any point within the body, and which is transmitted to the brain (Craig, 2002). Specifically, individual differences in the accuracy of perceiving bodily interoceptive signals have been associated with affective and decision-making processing (Sugawara et al., 2020).

The relationship between decision-making processes and body correlates has been studied before. For instance, Somatic Marker framework of Damasio (1996) illustrated that increased skin conductance, reflecting sympathetic nervous activity (a main interoception pathway) preceded rational decision-making

processes. Notably, participants with interoceptive dysfunction tend to select the disadvantageous option in a classical decision-making paradigm (Werner et al., 2009); in contrast, participants with increased interoceptive accuracy were likely to exhibit adaptive intuitive decision making (Dunn et al., 2010). Regarding moral decision-making in the UG, a previous study showed that experimental exposure to interoceptive signals influences participants' behavior at the task. It was found that listening to one's heart sound, compared to the other bodily sounds: (1) increased subjective feelings of unfairness, but not rejection behavior, in response to unfair offers and (2) increased the unfair offers while playing in the proposer role (Lenggenhager et al., 2013).

However, what is interesting to note is that Interoceptive Attentiveness (IA), i.e., attention focused on a particular interoceptive signal for a certain time interval (Schulz, 2016; Tsakiris and De Preester, 2018), is not a static dimension, but rather it can be manipulated and trained (Farb et al., 2013).

In the UG, usually, the rejection of asymmetric rewards is often seen as an important way for enforcing social norms and encouraging cooperative behavior (Fehr and Gächter, 2002). Past research showed that the ability to focus on ones' internal states positively correlates with the self-regulation of behavior in situations that are accompanied by somatic and/or physiological changes, such as emotions, physical workload, and decision-making (Herbert et al., 2007; Werner et al., 2009; Dunn et al., 2010).

Another perspective demonstrated that interoception could render individuals more empathic toward others with greater emotional arousal and affect sharing (Grynberg and Pollatos, 2015). In this regard, a previous study demonstrated that specific categories of individuals, such as meditators, seems not to experience the acceptance of unfair offers as social norm violations, as suggested by their higher acceptance rates (ARs) for asymmetric offers at the UG, but more as an acceptance of the interoceptive qualities that accompany any reward (small or large) compared to nothing (Kirk et al., 2011). Indeed, in the study meditators were better able to maintain the focus on their internal bodily states and uncouple negative emotional responses while confronted with asymmetric offers, with a related activation of brain areas involved in attention to the present moment and interoception (Kirk et al., 2011).

Concerning the fairness of the moral decisions, previous findings suggest an increase in Reaction Times (RTs) for fair and unfair offers compared with equal ones in the company and prosocial conditions (Balconi and Fronda, 2020). This result was interpreted according to the social context and attributed to the indirect involvement of individuals' interest in the company and prosocial conditions, for which equal offers (i.e., the offers in perfect balance, without concessions to the other) appear to be the most immediately acceptable options compared with fair and unfair ones, because they maintain an advantageous equilibrium, without gains or losses for anyone (Balconi and Fronda, 2020).

Given these premises, this exploratory study aims to explore if the modulation of the attention to internal states, namely IA, could influence the moral decision-making of lawyers at the UG. The experimental group of lawyers was explicitly required to focus the attention on their interoceptive correlates

while performing the task, compared to the control group of lawyers, which were instructed to perform the task only.

The behavioral effects related to RTs and the total number of accepted responses found in the previous study of Balconi and Fronda (2020) are expected in the sample of lawyers involved in the present study. Specifically, higher RTs are expected for offers implying higher cognitive control and cognitive dissonance, and in conditions requiring the evaluation of self-interest. Moreover, higher acceptance of the fair altruistic offers (i.e., the offers where I give up something for the other) is expected for social and professional conditions, displaying empathic responses toward the others, compared to company condition, in which the self-interest dimension could emerge instead, as in the previous study (Balconi and Fronda, 2020). Finally, regarding the manipulation of IA, it is supposed that the experimental group of lawyers will display a "gain effect" by accepting more the rewarding offers, according to Kirk et al. (2011) evidence on meditators, compared to the controls. To the best of our knowledge, this is the first time IA manipulation has been applied to a specific sample of legal professionals, while performing a moral decision-making task.

MATERIALS AND METHODS

Participants

Twenty lawyers with an age range between 25 and 54 years old took part in the present study. Participants were recruited on a voluntary basis; they were all physically health, Caucasian lawyers, mainly senior associates of a law firm. The following inclusion criteria were used for all participants: normal or corrected to normal visual acuity and absence of neurological or psychiatric pathologies. Exclusion criteria were age < 25 years old; less than 1 year of legal practice; the presence of neurocognitive deficits and a clinical history of neurological or psychiatric disorders. They were randomly assigned to experimental (EXP) and control (CON) group conditions; groups were matched for gender and age. All participants signed the informed consent and did not receive any compensation for their participation in the study. The research was conducted following the principles and guidelines of the Helsinki Declaration and was approved by the local ethics committee of the Department of Psychology of Catholic University of the Sacred Heart, Milan, Italy.

Experimental IA Manipulation

Before the task, the CON group received the general instruction to perform the task, without the IA manipulation, while the EXP group was explicitly asked to concentrate on their interoceptive correlates, while observing the stimuli and received the following instruction: "*During this task, we ask you to focus your attention on your bodily sensations (such as on your breath). Try to observe how you feel and if there are any variations in your body as you perform the task*" (Balconi and Angioletti, 2021a,b).

Moral Decision-Making Task

Participants were asked to perform a modified version of the task adopted by Balconi and Fronda (2019, 2020), implemented using the Qualtrics XM platform (Qualtrics LLC, Provo, UT, United States). The task, which consists of a modified version of the UG (Balconi and Fronda, 2019, 2020), proposed three different randomized moral conditions of choice (professional fit, company fit, and social fit) adapted to the legal context.

In particular, the task required two players: the proponent (different according to the context of choice) and the respondent (the individual who performed the task) to distribute a sum of money. The proposer decided how to distribute the sum of money, and the respondent could decide whether to accept or reject the proposed offer. If the respondent decided to refuse the offer, no player would take the money.

In the professional fit condition, the decision on how to divide the amount of money relates to one's profession (e.g., lawyers were required to accept or reject the proposal of a colleague for a work done together). In the company fit condition, the decision on how to divide the sum of money concerns the effects on the business organization (e.g., lawyers were asked to accept or reject the proposal of the law firm for the realization of certain common benefits, such as the laundry service, in the Firm). In the social fit condition, the decision on how to divide the sum of money concerns the social context (e.g., lawyers were required to accept or reject the proposal of the law firm for making a financial contribution to a colleague's relative with health problems).

For each condition (professional fit, company fit, and social fit), 10 coherent scenarios were presented. The different choice conditions were presented in three blocks of three randomized scenarios and each block lasted approximately 15 min. At the end of each scenario presentation, three different offers of distribution of money were proposed: equal (50% to the proposer and 50% to the responder), fair (60% to the proposer and 40% to the responder), and unfair (40% to the proposer and 60% to the responder).

Participants could accept or reject the proposed offer by clicking the "Accept" or "Reject" button on the survey. Specifically, the three offers (fair, unfair, and equal) were presented separately on the screen until the participant decided whether to accept or reject the offer proposed to record the response times. Moreover, participants were not given a defined time interval to decide whether to accept or reject the proposed offer.

After each block of three randomized scenarios, participants evaluated how much attention (i.e., their Attention Focus, AF) they paid to the situation, the self, or other on a Visual Analog Scale (VAS) ranging from 0 (no attention) to 10 (complete attention). This way, fluctuations of attention over the three blocks were assessed. For a graphical representation of the procedural steps and examples of scenarios, see **Figure 1**.

Data Analysis

Visual Analog Scale scores, AR (total number of accepted proposed offers), and RTs for each condition related to participants' choices were obtained. A first mixed repeated measure ANOVA with independent factors *Block* (3: First, Second, and Third) × *AF* (3: Situation, Self, and Other), and as between factor, the *Group* (2: EXP, CON) was applied to VAS scores. Then, two mixed repeated measures ANOVA with independent within factors
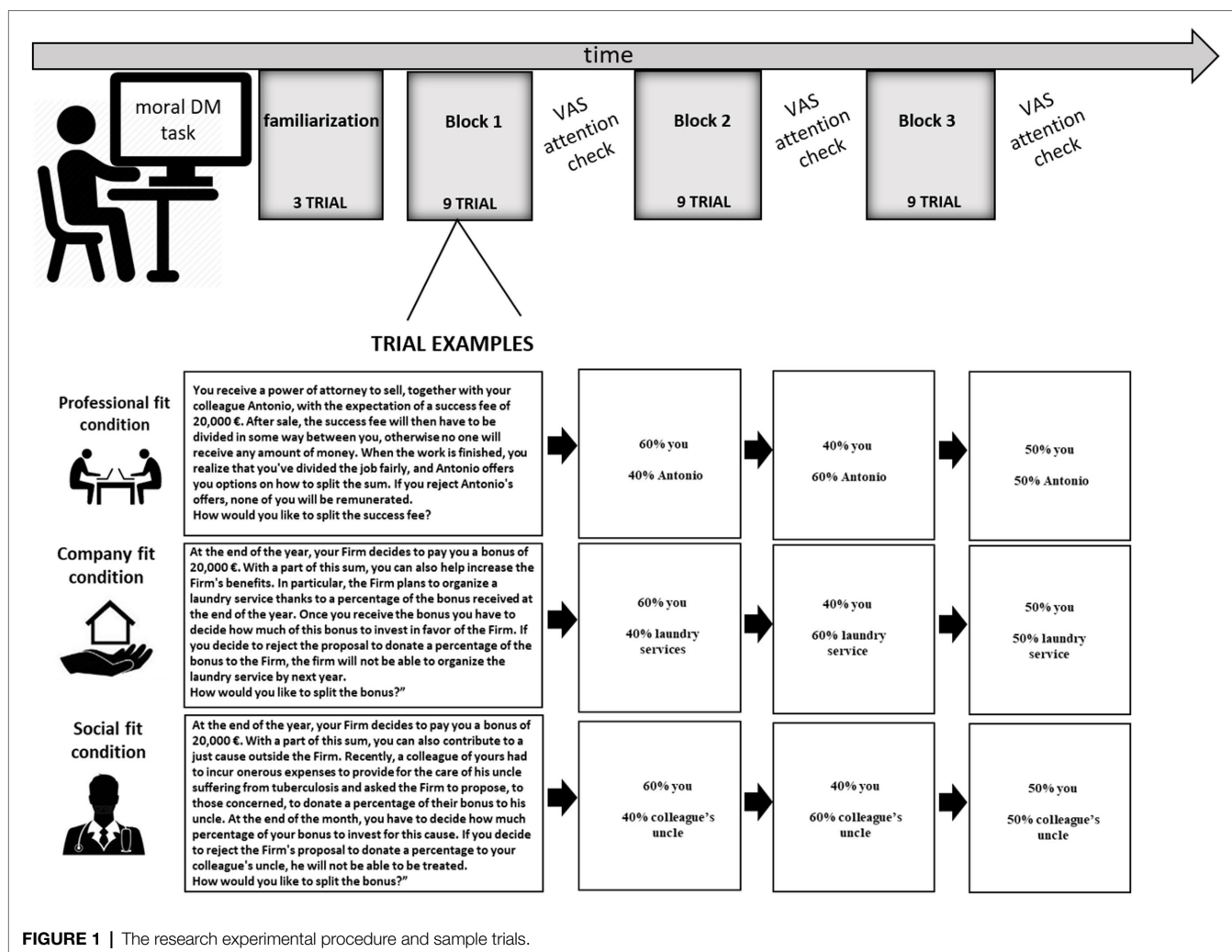
FIGURE 1 | The research experimental procedure and sample trials.

Condition (3: Professional, Company, and Social) × *Offer Type* (3: Fair, Unfair, and Equal), and as between factor the *Group* (2: EXP, CON) were applied to behavioral measures, i.e., RTs and AR. For all the ANOVA tests, the degrees of freedom have been corrected using Greenhouse–Geisser epsilon where appropriate. The threshold for statistical significance was set to $\alpha = 0.05$. Pairwise comparisons were applied to the data in case of significant effects. Simple effects for significant interactions were further checked *via* pairwise comparisons, and Bonferroni correction was used to reduce multiple comparisons' potential biases. Furthermore, the normality of the data distribution was preliminarily assessed by checking kurtosis and asymmetry indices. The size of statistically significant effects has been estimated by computing partial eta squared (η^2) indices.

RESULTS

Visual Analog Scale Score

ANOVA showed a significant main effect for *Group* [$F(1,18) = 6,675$, $p = 0.019$, $\eta^2 = 0.271$]. Pairwise comparisons revealed higher mean scores for the EXP group compared to the CON group ($p = 0.019$; **Figure 2A**).

Secondly, a significant main effect for *AF* was detected [$F(2,36) = 4,232$, $p = 0.041$, $\eta^2 = 0.190$]. In particular, pairwise comparisons showed significant higher mean scores for the situation compared to the self ($p = 0.021$; **Figure 2B**).

Acceptance Rate

For individuals' options of response, ANOVA revealed a significant main effect for Offer Type [$F(2,36) = 4.977$, $p = 0.014$, $\eta^2 = 0.217$]. Pairwise comparisons showed a significant increase of accepted responses for equal offers compared to fair offers ($p = 0.037$; **Figure 3A**).

Secondly, ANOVA showed a significant interaction effect *Condition × Offer Type* [$F(4,72) = 13.809$, $p < 0.001$, $\eta^2 = 0.434$]. Specifically, simple effect analysis revealed that in the professional fit condition there is an increase of accepted responses for the equal offers compared to unfair ones ($p < 0.001$) and, also, compared to fair offers ($p < 0.001$). Instead, in the company fit condition, higher values of accepted response were found for the unfair offers compared to the fair offers ($p = 0.029$).

In addition, simple effect analysis revealed that fair offer type is more accepted in the social fit condition than in

FIGURE 2 | **(A,B)** Visual Analog Scale (VAS) score. Bar chart shows VAS scores in terms of **(A)** higher mean values of attention scores for the experimental compared to control group; **(B)** higher mean values of self-reported attention for the situation compared to the self. For all charts, bars represent ±1 SE; all asterisks mark statistically significant differences, with $p \leq 0.05$.
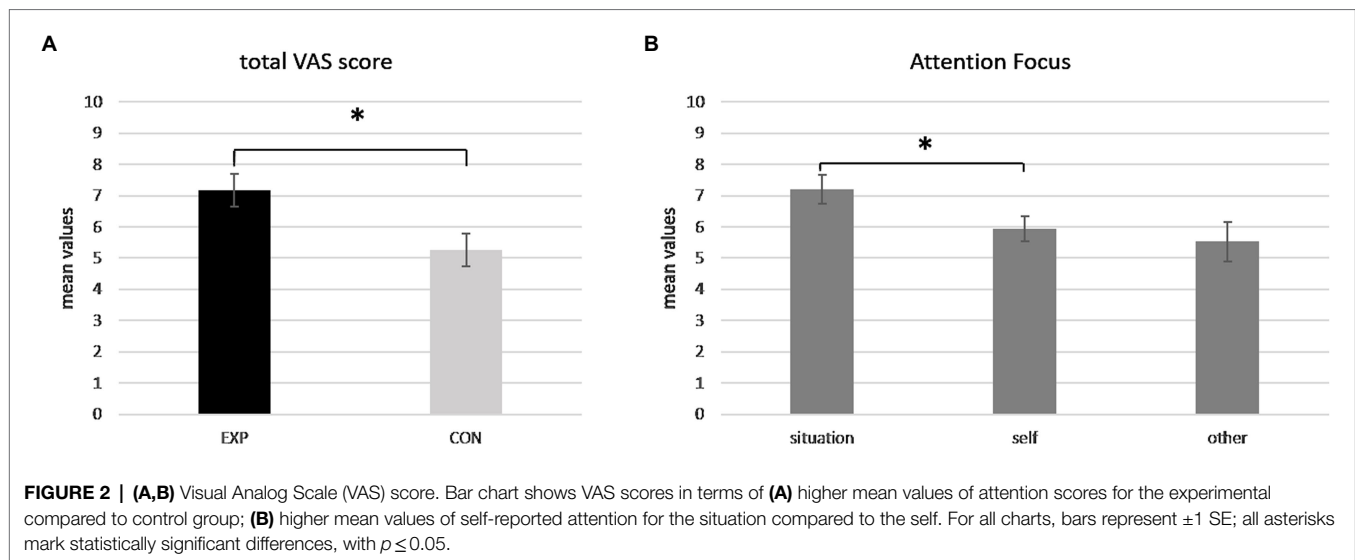
the professional fit ($p = 0.045$) and in the company fit condition ($p = 0.003$). Moreover, equal offers were more accepted in the professional fit condition compared to the company fit condition ($p < 0.001$) and social fit condition ($p < 0.001$; **Figure 3B**).

Thirdly, a significant interaction effect *Condition × Group* was detected [$F(4,72) = 6.342$, $p = 0.013$, $\eta^2 = 0.261$]. As revealed by pairwise comparisons, an increase of accepted responses in the professional fit condition was found for the EXP group compared to the CON group ($p = 0.030$; **Figure 3C**).

Lastly, a significant interaction effect *Offer Type × Group* was found [$F(4,72) = 3.546$, $p = 0.043$, $\eta^2 = 0.165$]. In particular, the pairwise comparison showed that for the EXP group significantly higher accepted responses were obtained for equal compared to fair offers ($p = 0.020$; **Figure 3D**).

Reaction Times

For the RTs, a significant main effect for *Offer Type* was found [$F(2,36) = 17.221$, $p < 0.001$, $\eta^2 = 0.489$]. The pairwise comparison revealed significantly lower RTs for unfair offers compared to equal offers ($p < 0.001$). Also, the pairwise comparison showed significantly lower RTs for fair offers compared to equal offers ($p = 0.002$; **Figure 4**).

DISCUSSION

This exploratory study investigated lawyers' moral choices concerning different decision-making conditions and offers within a law firm context. The manipulation of the IA during the presentation of different moral scenarios and offers made it possible to investigate behavioral responses in relation to three moral decision-making conditions (professional fit, company fit, and social fit) and offers (fair, unfair, and equal).
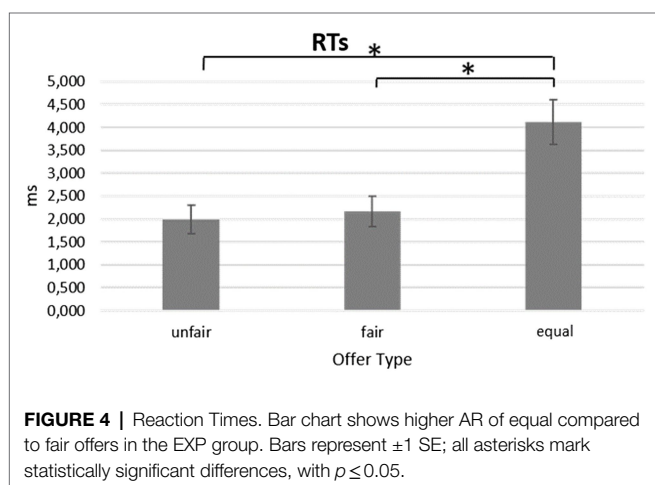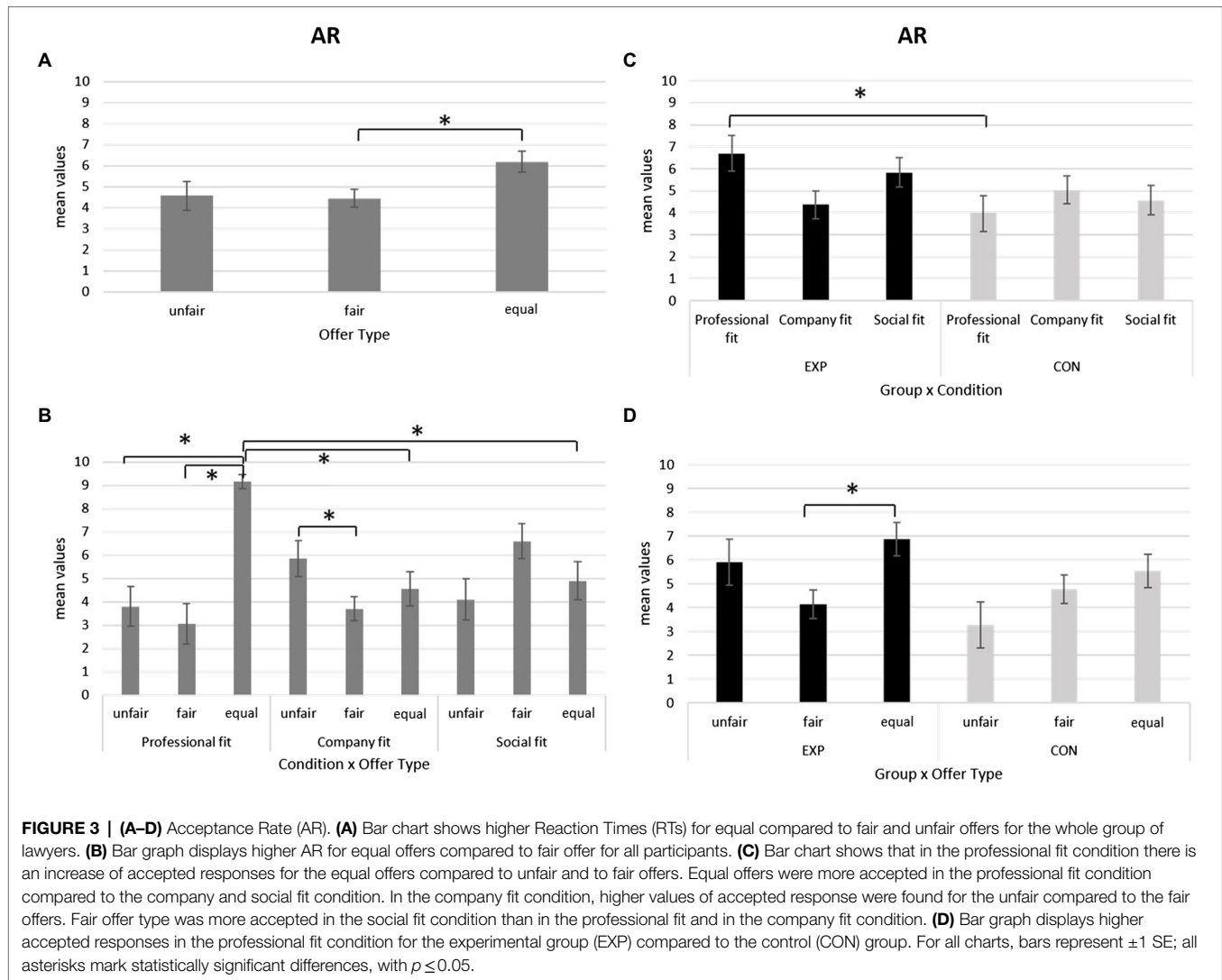
Firstly, we have observed a general increase of AR for equal offers than fair or unfair ones. This general evidence

shows how lawyers are more willing to offer benefits equally, despite this choice having a noticeable cost in terms of RTs. As highlighted in the results, it seems that being quicker in selecting both fair and unfair options compared to equal options, independently from the condition, constitutes easier choices for lawyers. Moreover, a general increase of RTs in the equal decisions was observed: all the lawyers spent more time for accepting an equal offer compared to the other offers proposed.

Previous literature showed that cognitively complex processes and conflicts restrain economically self-interested responses (Halali et al., 2011) require higher cognitive cost and resources which are associated with higher RTs; on the contrary, cognitively fewer complex processes are associated with faster RTs, as they require less information processing (Klapp, 2010).

Therefore, in the present study, regardless of the condition, a possible interpretation could be that lawyers were more likely to make immediate fair and unfair choices (significant reduction of RTs for both fair and unfair offers), perhaps due to less complex cognitive processes and because of the higher direct engagement, which supports a more immediate ability to produce the moral decision. On the other hand, it may be possible that equal choices required a greater cognitive decision-making effort for lawyers, perhaps because of the greater degree of uncertainty in the choice, due to the assessment that does not directly concern one's interests.

This result is partially in contrast with a previous study showing an increase of RTs for fair and unfair offers compared with equal ones, specifically in the company and social fit conditions (Balconi and Fronda, 2020). However, two main aspects distinguish our result on RTs from the evidence of this previous research and are that (i) the variation in RTs was interpreted in relation to the condition in which the offers were significantly accepted (company and social fit condition), and (ii) the study referred to a sample of managers, not including lawyers. The lawyers'

FIGURE 3 | **(A–D)** Acceptance Rate (AR). **(A)** Bar chart shows higher Reaction Times (RTs) for equal compared to fair and unfair offers for the whole group of lawyers. **(B)** Bar graph displays higher AR for equal offers compared to fair offer for all participants. **(C)** Bar chart shows that in the professional fit condition there is an increase of accepted responses for the equal offers compared to unfair and to fair offers. Equal offers were more accepted in the professional fit condition compared to the company and social fit condition. In the company fit condition, higher values of accepted response were found for the unfair compared to the fair offers. Fair offer type was more accepted in the social fit condition than in the professional fit and in the company fit condition. **(D)** Bar graph displays higher accepted responses in the professional fit condition for the experimental group (EXP) compared to the control (CON) group. For all charts, bars represent ±1 SE; all asterisks mark statistically significant differences, with $p \leq 0.05$.

FIGURE 4 | Reaction Times. Bar chart shows higher AR of equal compared to fair offers in the EXP group. Bars represent ±1 SE; all asterisks mark statistically significant differences, with $p \leq 0.05$.

category may differ, in terms of moral decision-making behavior, from other professional categories, and therefore it would be interesting for future studies to test the differences

between different professional groups, even for clarifying the present results.

Secondly, the AR effect for equal offers was mainly found in the professional condition, in which the whole sample of lawyers tend to accept equal than fair and unfair offers more frequently. While in the company fit condition, lawyers accept more unfair offers compared to fair offers. Specifically, it is plausible that a rational responder motivated purely by self-interest prefers to accept the equal amount of money offered by the proposer (i.e., the colleague, in the professional condition), as this offer will represent a fair gain for a work task equally done together. Instead, lawyers were more willing to accept unfair offers for deriving personal benefits when faced with splitting the sum of money with the law firm, in the company condition. In line with the previous study (Balconi and Fronda, 2020), in which the abovementioned sample of managers tend to accept more personally advantageous offers when in the company condition, a possible explanation could be that, even for the present group of lawyers, the company condition has

been conceived as a distant or "external" situation, not personified, and therefore less relevant on a personal level (where individuals can gain some advantages without excessive "moral costs"). As previously suggested by Balconi and Fronda (2020), when moral decision-making dynamics are framed by the company context, the decision process seems to be strongly influenced by the subjective understanding of the inherent benefits to themselves, more than by the disadvantages for the others, intended as the company or, in this case, as the law firm.

Thirdly, the evaluation linked to the benefits and advantages of others, also compared to one's advantages, varies depending on the decision context (professional, company, or social fit conditions). As highlighted above, in the company condition lawyers preferred the unfair compared to fair offers, whereas in the social condition fair offers are more accepted compared to the professional and company condition. Furthermore, in the professional condition, the equal offer is significantly more accepted than in the company and social condition. Therefore, the condition in which offers are proposed may modulate the lawyers' moral decision-making process, since they seem to be more unfair in a company condition, fairer in a social condition, and, most of all, equal in a condition of professional engagement with a colleague.

A possible explanation for these outcomes could be that lawyers could prefer to accept offers that are inherently more advantageous and promote their self-interest in decisions involving the law firm, perhaps displaying less empathic and altruistic behavior. In contrast, lawyers showed more empathic behavior in the conditions, where other individuals such as the relative of a colleague (the social condition) are involved in the attribution of the sum of money. Indeed, as shown by some previous research, empathic behavior can facilitate a greater understanding of the potential effects, consequences, and obligations of actions about the well-being of others and allows to better evaluate the cost of personal choices and the extent of social benefits (Mencl and May, 2009; Dietz and Kleinlogel, 2014). As supposed in our hypotheses, it is possible that in the social fit condition, lawyers experienced the highest levels of empathy with respect to the company and professional fit condition and they tend to overshadow their self-interest for pursuing a more important social cause.

Fourthly, as highlighted in the results of VAS scores, the manipulation of the attention to the internal state induced in the lawyers increased subjective attention to the situation concerning the self or other conditions. Moreover, the EXP group, which was explicitly required to focus the attention on their interoceptive correlates while performing the task, significantly expressed higher general levels of attention during the task compared to the CON group. The present finding could be considered the first evidence of the successful manipulation of the interoception in the EXP group of lawyers. Despite this interesting preliminary result and the use of VAS in previous similar studies (Lenggenhager et al., 2013), we are aware it consists of a subjective self-report measure of attention and further

studies would benefit from the integration of objective measures, applied both before (e.g., heartbeat detection task, to control individual differences in interoceptive ability) and during the task (e.g., eye-tracking technique, as an indirect measure of the attentional focus; cortical frequency bands recording through electroencephalogram).

Regarding the different moral conditions in the task, the EXP group showed higher AR for offers displayed in the professional fit situation than the CON group. This result could be related to the fact that the interoceptive manipulation oriented the lawyers' focus toward a more self-centered decision process, which aims to obtain a direct greater profit for themselves. It is worth noting the professional condition is the unique condition in which participants were explicitly involved in the first person compared to the other conditions (since they were required to split a sum of money about work they conducted together with a colleague). In line with this, previous research found that receiving interoceptive feedback might enhance self-centered perspective taking and "first-person perspective" (Kirk et al., 2011; Lenggenhager et al., 2013).

Regarding the type of offer, it seems that regardless of the condition, the EXP group preferred more equal offers than fair and unfair ones, at the expense of personal and others' advantage. Perhaps, a possible explanation for this second effect is that IA in this group might enhance a more equal attitude, with a preference for equal and more balanced choices. This result is partially in contrast with study of Kirk et al. (2011), for which specific groups that are trained to modulate their IA (i.e., meditators) tend to show higher acceptance rates for asymmetric offers at the UG, but also with Lenggenhager et al. (2013) and with Piech et al. (2017) research for which only interoceptive sensitivity, conceived as a trait component, predict altruistic behavior in the dictator game. In the literature, the ability to accurately detect one's internal body signals has been also associated with cognitive and emotional components of empathy, in terms of greater emotional arousal and affect sharing toward others (Grynberg and Pollatos, 2015). Therefore, future studies are needed to better deepen this "equity effect" in lawyers and the basic relation between IA and moral decision-making in the UG.

To the best of our knowledge, there is still no evidence in the literature about this specific phenomenon, and no previous data or analysis regarding the lawyers were reported in similar studies. Indeed, despite the innovativeness of the paradigm, the present exploratory study has some limitations: the sample size should be augmented and a multimethodological approach should be adopted to collect also the neural and psychophysiological correlates underlying the cognitive and emotional moral decision-making processes (Balconi and Molteni, 2016). Therefore, future research is needed to replicate and expand current findings.

DATA AVAILABILITY STATEMENT

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Department of Psychology, Catholic University of the Sacred Heart, Milan, Italy. The patients/participants provided their written informed consent to participate in this study.

REFERENCES

Angioletti, L., and Balconi, M. (2020). Interoceptive empathy and emotion regulation: the contribution of neuroscience. *Neuropsychol. Trends* 27, 85–100. doi: 10.7358/neur-2020-027-ang2

Balconi, M., and Angioletti, L. (2021a). Interoception as a social alarm amplification system. What multimethod (EEG-fNIRS) integrated measures can tell us about interoception and empathy for pain? *Neuropsychol. Trends.* 39–64. doi: 10.7358/neur-2021-029-bal1

Balconi, M., and Angioletti, L. (2021b). One's interoception affects the representation of seeing others' pain: a randomized controlled qEEG study. *Pain Res. Manag.* 2021, 1–15. doi: 10.1155/2021/5585060

Balconi, M., and Angioletti, L. (2022). Aching face and hand: the interoceptive attentiveness and social context in relation to empathy for pain. *J. Integr. Neurosci.* 21:34. doi: 10.31083/j.jin2101034

Balconi, M., and Fronda, G. (2019). Physiological correlates of moral decision-making in the professional domain. *Brain Sci.* 9:229. doi: 10.3390/brainsci9090229

Balconi, M., and Fronda, G. (2020). Morality and management: an oxymoron? fNIRS and neuromanagement perspective explain us why things are not like this. *Cogn. Affect. Behav. Neurosci.* 20, 1336–1348. doi: 10.3758/s13415-020-00841-1

Balconi, M., and Molteni, E. (2016). Past and future of near-infrared spectroscopy in studies of emotion and social neuroscience. *J. Cog. Psychol.* 28, 129–146.

Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nat. Rev. Neurosci.* 3, 655–666. doi: 10.1038/nrn894

Damasio, A. R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 351, 1413–1420. doi: 10.1098/rstb.1996.0125

Danziger, S., Levav, J., and Avnaim-Pessoa, L. (2011). Extraneous factors in judicial decisions. *Proc. Natl. Acad. Sci. U. S. A.* 108, 6889–6892. doi: 10.1073/pnas.1018033108

Dietz, J., and Kleinlogel, E. P. (2014). Wage cuts and managers' empathy: how a positive emotion can contribute to positive organizational ethics in difficult times. *J. Bus. Ethics* 119, 461–472. doi: 10.1007/s10551-013-1836-6

Dunn, B. D., Galton, H. C., Morgan, R., Evans, D., Oliver, C., Meyer, M., et al. (2010). Listening to your heart. How interoception shapes emotion experience and intuitive decision making. *Psychol. Sci.* 21, 1835–1844. doi: 10.1177/0956797610389191

Farb, N. A. S., Segal, Z. V., and Anderson, A. K. (2013). Mindfulness meditation training alters cortical representations of interoceptive attention. *Soc. Cogn. Affect. Neurosci.* 8, 15–26. doi: 10.1093/scan/nss066

Fehr, E., and Gächter, S. (2002). Altruistic punishment in humans. *Nature* 415, 137–140. doi: 10.1038/415137a

Goodenough, O., and Prehn, K. (2004). A neuroscientific approach to normative judgment in law and justice. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 359, 1709–1726. doi: 10.1098/rstb.2004.1552

Goodenough, O., and Zeki, S. (2006). *Law and the Brain*. Oxford, UK: Oxford University Press.

Grynberg, D., and Pollatos, O. (2015). Perceiving one's body shapes empathy. *Physiol. Behav.* 140, 54–60. doi: 10.1016/j.physbeh.2014.12.026

Häfner, M. (2013). When body and mind are talking. Interoception moderates embodied cognition. *Exp. Psychol.* 60, 255–259. doi: 10.1027/1618-3169/a000194

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

Halali, E., Bereby-Meyer, Y., and Meiran, N. (2011). When rationality and fairness conflict: the role of cognitive-control in the ultimatum game. *SSRN Electron. J.*, 1–22. doi: 10.2139/ssrn.1868852

Herbert, B. M., and Pollatos, O. (2012). The body in the mind: on the relationship between interoception and embodiment. *Top. Cogn. Sci.* 4, 692–704. doi: 10.1111/j.1756-8765.2012.01189.x

Herbert, B. M., Ulbrich, P., and Schandry, R. (2007). Interoceptive sensitivity and physical effort: implications for the self-control of physical load in everyday life. *Psychophysiology* 44, 194–202. doi: 10.1111/j.1469-8986.2007.00493.x

Kirk, U., Downar, J., and Montague, P. R. (2011). Interoception drives increased rational decision-making in meditators playing the ultimatum game. *Front. Neurosci.* 5:49. doi: 10.3389/fnins.2011.00049

Klapp, S. T. (2010). Comments on the classic henry and rogers (1960) paper on its 50th anniversary: resolving the issue of simple versus choice reaction time. *Res. Q. Exerc. Sport* 81, 108–112. doi: 10.1080/02701367.2010.10599634

Lenggenhager, B., Azevedo, R. T., Mancini, A., and Aglioti, S. M. (2013). Listening to your heart and feeling yourself: effects of exposure to interoceptive signals during the ultimatum game. *Exp. Brain Res.* 230, 233–241. doi: 10.1007/s00221-013-3647-5

Mencl, J., and May, D. R. (2009). The effects of proximity and empathy on ethical decision-making: an exploratory investigation. *J. Bus. Ethics* 85, 201–226. doi: 10.1007/s10551-008-9765-5

Nobandegani, A., Destais, C., and Shultz, T. R. (2020). A resource-rational process model of fairness in the ultimatum game. in *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*. July 29–August 1; 3205–3211.

Pace-Schott, E. F., Amole, M. C., Aue, T., Balconi, M., Bylsma, L. M., Critchley, H., et al. (2019). Physiological feelings. *Neurosci. Biobehav. Rev.* 103, 267–304. doi: 10.1016/j.neubiorev.2019.05.002

Piech, R. M., Strelchuk, D., Knights, J., Hjälmheden, J. V., Olofsson, J. K., and Aspell, J. E. (2017). People with higher interoceptive sensitivity are more altruistic, but improving interoception does not increase altruism. *Sci. Rep.* 7, 15652–15655. doi: 10.1038/s41598-017-14318-8

Schulz, S. M. (2016). Neural correlates of heart-focused interoception: a functional magnetic resonance imaging meta-analysis. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 371:20160018. doi: 10.1098/rstb.2016.0018

Sugawara, A., Terasawa, Y., Katsunuma, R., and Sekiguchi, A. (2020). Effects of interoceptive training on decision making, anxiety, and somatic symptoms. *Biopsychosoc. Med.* 14, 1–8. doi: 10.1186/s13030-020-00179-7

Tormen, F. (2020). Cognitive neuroscience applied to law. A neurolaw introduction. *Neuropsychol. Trends* 28, 83–91. doi: 10.7358/neur-2020-028-torm

Tsakiris, M., and De Preester, H. (eds.) (2018). The interoceptive mind: from homeostasis to awareness. Oxford University Press.

Werner, N. S., Jung, K., Duschek, S., and Schandry, R. (2009). Enhanced cardiac perception is associated with benefits in decision-making. *Psychophysiology* 46, 1123–1129. doi: 10.1111/j.1469-8986.2009.00855.x

AUTHOR CONTRIBUTIONS

LA and FT wrote the first draft and each section of the manuscript. LA, FT, and MB contributed to the manuscript final writing and revision, read, and approved the submitted version.

Check for updates

A Positive Versus Negative Interaction Memory Affects Parole Officers' Implicit Associations Between the Self-Concept and the Group Parolees

Marina K. Saad[1,2]*, Luis M. Rivera[3]* and Bonita M. Veysey[2]

[1]*Department of Criminology and Criminal Justice, Fairleigh Dickinson University, Hackensack, NJ, United States,* [2]*School of Criminal Justice, Rutgers University-Newark, Newark, NJ, United States,* [3]*Department of Psychology, Rutgers University-Newark, Newark, NJ, United States*

Background: Parole officers are one of many actors in the legal system charged with interpreting and enforcing the law. Officers not only assure that parolees under their supervision comply with the terms of their release, but also monitor and control parolees' criminal behavior. They conduct their jobs through their understanding of their official mandate and make considered and deliberate choices while executing that mandate. However, their experiences as legal actors may impact their implicit cognitions about parolees. This experiment is the first of its kind to examine implicit (i.e., automatic) associations between the self and parolees among actors of the legal system.

Objective: The present study examines the implicit cognitive consequences of the quality of the parole officer-parolee relationship from the perspective of the parole officer; specifically, whether parole officers who are reminded of positive experiences with parolees implicitly associate more with the group parolees than those reminded of a negative experience. In addition, we explore the moderating effects of parole officers' subjective professional orientation and identification.

Method: Eighty-four New Jersey parole officers participated in the study. First, an experimental manipulation of either a past positive or negative experience was administered *via* a writing task. Participants then completed an Implicit Association Test (IAT) to measure associations between the self-concept of parole officers with parolees who are part of the group criminal, followed by measures of professional orientation and identification.

Results: Participants who were reminded of a positive experience with a parolee exhibited stronger associations between self and the group parolee when compared to those who were reminded of a negative experience. Neither professional orientation nor parole officer group identification were related to implicit associations and did not moderate the effect of the past experience reminder on implicit associations.

Conclusion and Implications: Implicit cognitions of parole officers may influence their behaviors and interactions with those whom they supervise. Positive reminders affect implicit self-associations with parolees presumably *via* empathy, which is known to affect the quality of therapeutic and supervision relationships; thus, theoretically, leading to improved outcomes for both officers and parolees.

INTRODUCTION

We say, "justice is blind," yet we know that extra-legal factors (e.g., race, gender) are often associated with various law-related decisions, including arrests (e.g., Fielding-Miller et al., 2020), verdicts and sentencing (e.g., Cohen and Yang, 2019), and parole release decisions (e.g., Huebner and Bynum, 2008). Much of the extant research assumes that decisions are arrived at in a thoughtful considered way, and, despite this, that bias often influences these decisions. Bias, of course, can be a conscious cognitive process, and, importantly, it also can operate automatically outside of conscious awareness (for reviews, see Gawronski and Payne, 2011). Automatic or implicit cognitions, such as bias, also may affect interactions between actors in the justice system (e.g., between citizens and police officers, parolees and parole officers). Only recently has implicit social cognition been studied in samples of justice-involved people (Rivera and Veysey, 2015, 2018; Veysey and Rivera, 2017). The present study is the first to our knowledge to examine implicit self-cognitions among actors in the justice system. Using a sample of parole officers, we investigate the implicit association of the self with the group parolee following a positive (versus negative) interaction with a parolee. We suggest that this effect occurs *via* self-expansion, a mechanism that supports positive downstream outcomes for both parole officers and parolees.

Parole in the United States

At year-end 2019, there were 878,900 individuals under parole supervision (Oudekerk and Kaeble, 2021). Historically, and to a large degree today, the role of the parole officer is to assist in the successful reentry of individuals who have been released from prison while monitoring their behavior and the terms of their release (e.g., maintaining employment and stable housing, abstaining from alcohol and drugs); in sum, responding to their criminogenic and non-criminogenic needs to ensure community safety (Seiter, 2002; National Research Council, 2008).

In the parole officer-parolee relationship, parole officers may function as agents of change who encourage parolees to comply with the conditions of parole, engage in specialized programs, and promote pro-social behaviors and identities (Gibbons and Rosecrance, 2005; Abadinsky, 2009) or alternatively, parole officers may foster an adversarial relationships with their parolees (Ireland and Berg, 2007; Morash et al., 2014; Chamberlain et al., 2017). The effectiveness of the parole officer-parolee relationship is dependent on the parole officer's ability to create and maintain positive relationships with parolees on their caseloads (Dowden and Andrews, 2004; Landenberger and Lipsey, 2005; Jolliffe and Farrington, 2007; Morash et al., 2014). Evidence-based practices in community corrections highlight the importance of practitioner-client relationships that are characterized by warmth, empathy, respect, and support (Dowden and Andrews, 2004; Ireland and Berg, 2007; Andrews, 2011). In these relationships, parole officers connect and collaborate with offenders, model pro-social behavior, communicate

effectively, and apply motivational techniques (Walters et al., 2007). Positive parole officer-parolee relationships foster bonding, closeness, and trust between both parole officers and parolees (Ireland and Berg, 2007; Ross et al., 2008). These factors provide the necessary context in which change can happen. For example, interviews with parolees consistently demonstrate that positive relationships with their parole officers facilitate pro-social cognitive changes (e.g., identity shifts; Giordano et al., 2007; Bui and Morash, 2010; Morash et al., 2014; Stone et al., 2016) that can lead to stable, long-term success.

In contrast, negative parole officer-parolee relationships are characterized as authoritarian, unsupportive, inflexible, and controlling (Stone et al., 2016). Negative parole officer-parolee relationships have been found to be related to confrontation and noncompliance with the terms of supervision (Ireland and Berg, 2007; Morash et al., 2014; Chamberlain et al., 2017). Importantly, negative parole officer-parolee relationships are not conducive to pro-social cognitive changes (Morash et al., 2014; Stone et al., 2016). Taken together, this research suggests that the quality of the parole officer-parolee relationship plays a role in parolees' cognitions and self-perception with implications for parolees' overall success.

The extant research also begs the question, if experiences in the parole officer-parolee relationship has an impact on parolees' cognitions and self-perception, can these experiences also impact the parole officers' cognitions; especially their self-perception? Specifically, how do negative or positive parole officer-parolee experiences impact the way parole officers think about themselves in relation to those whom they supervise? The present study is the first to our knowledge to address this question by targeting implicit cognitions about the self in relation to parolees.

Self-Expansion Theory and Implicit Social Cognition

Self-expansion theory posits that frequent and positive experiences with close others (e.g., spouses, close friends) can lead individuals to assume attributes, cognitions, and behaviors of those close others (Aron and Aron, 1986; Aron et al., 1991). Put differently, when the conditions of self-expansion are met, that is, frequent and positive experiences with a close other, an individual is more likely to associate their self-concept (i.e., identity and self-perception) with aspects of another person's identity.

Self-expansion is based on individuals' desire to enhance personal growth, progress, and self-efficacy (Aron and Aron, 1996). In close relationships, both individuals mutually include some or all aspects of the other into their self-concept. As a result, a mental overlap occurs between the self and the close other which allows both individuals to: (1) vicariously take on the resources (i.e., physical and social capital), (2) cultivate new perspectives, and (3) acquire new characteristics or identities related to the other and incorporate them into their self-concept (Aron and Aron, 1986, 1996, 1997; Aron et al., 1991, 1992). For instance, when a close other is perceived as a part of the

self, the allocation of resources is shared (Clark and Mills, 1979), perspective differences are decreased (Brenner, 1973), and characteristics of others are perceived as one's own (Tesser et al., 1988).

Self-expansion has been applied to understand the effects of intergroup relationships, such as cross-group friendships (e.g., Latino and White friendships; Page-Gould et al., 2010). This line of research suggests that people are motivated not only to expand their self-concept between close others as individuals but also the groups to which a close other belongs. Self-expansion in intergroup relationships requires that an individual has a relationship with a member of a different social group and that they engage in close, frequent, and positive experiences with each other, that then results in the motivation to associate with the group as a whole and its related traits (Smith and Henry, 1996; Coats et al., 2000; Aron and McLaughlin-Volpe, 2001).

Studies have demonstrated the self-expansion phenomenon using "explicit" or direct self-report measures (Aron et al., 1991, 1992; Page-Gould et al., 2010) and "implicit" or indirect (i.e., reaction time) measures (Aron et al., 1991; Aron and McLaughlin-Volpe, 2001; Page-Gould et al., 2010). For example, married couples exhibit explicit associations between the self and their spouse using the Inclusion of the Other in the Self (IOS) scale, a self-report measure of self-expansion on which participants indicate how close they perceive another person by selecting one of seven pairs of circles that vary in their between-circle distances to represent different degrees of cognitive overlap between the self and the other (Aron et al., 1992, 1995). Married individuals also demonstrate self-expansion on "implicit" measures (Aron et al., 1991; Aron and McLaughlin-Volpe, 2001). On "me/not-me" tasks in which reaction time is used to measure similarity, married participants are quicker (i.e., press a button labeled "me" versus one labeled "not me") to categorize traits related to their spouses as self-descriptive than traits unrelated to their spouses. Similarly, on a "yes/no" task, participants in a romantic relationship respond faster (i.e., press a button labeled "yes" versus one labeled "no") to traits relevant to both the participants and the partner than to traits that were different between themselves and their partner (Smith et al., 1999). Taken together, this past research provides evidence for explicit and implicit self-expansion.

Professional Self-Expansion

Most of the self-expansion research has focused on self-expansion in personal relationships. To our knowledge, one study has examined self-expansion within the context of people's occupations (McIntyre et al., 2014). McIntyre et al. (2014) employed a 14-item self-expansion questionnaire to measure the extent to which people exhibited self-expansion with their occupation as a whole. The study demonstrated that people can indeed self-expand with their occupation. Although McIntyre et al. (2014) focused on self-expansion with people's occupations as a whole, they suggest that people can self-expand with others with whom they interact in the workplace if the interactions meet the pre-requisites of self-expansion. We extend their work and examine the extent to which parole officers self-expand

with the group parolee, a group with which parole officers frequently interact in the workplace by measuring implicit criminal-self associations; one potential consequence of self-expansion with parolees.

Implicit Self-Expansion in Parole Officers

The present study adopts an implicit social cognition approach utilizing an Implicit Association Test (IAT), to measure associations between the self-concept of parole officers with the social identity group criminal. Like other social identity groups, "criminals," including parolees, self-categorize (Krueger, 2001) as criminal, and, as such, perceive themselves as sharing attributes, cognitions and experiences with the group (Boduszek et al., 2013). Further, past criminal experiences are sufficient for individuals to identify the self both explicitly (Asencio, 2011; Boduszek et al., 2013) and implicitly (Rivera and Veysey, 2015, 2018; Veysey and Rivera, 2017) with the group criminal. Indeed, repeated studies have established the IAT as a reliable and valid measure of implicit self-criminal associations (Rivera and Veysey, 2015, 2018; Veysey and Rivera, 2017). The present research, measuring the extent to which parole officers implicitly associate themselves with their parolees, extends past work on direct personal experience as a criminal to indirect personal experience with a member of the group criminal (i.e., parolee).

Our *a priori* hypothesis is that parole officers who have positive (versus negative) experiences with parolees are likely to exhibit relatively strong implicit associations with the group criminal. The logic underlying this hypothesis follows previous examinations of self-expansion. Parole officers may experience self-expansion through frequent and positive experiences with parolees. Indeed, the occupational role of parole officers requires them to frequently and directly meet with multiple parolees on a daily basis (at roughly 76 h per month; DeMichele, 2007). As such, these meetings provide opportunities for positive interactions that may foster a sense of closeness between parole officers and their parolees. Moreover, parole officers may view the success or shortcomings of their parolees as their own, reflecting a sense of interconnectedness. Finally, parole officers have the potential to engage in positive experiences with their parolees as they support parolees' reintegration into society. For parole officers, therefore, it is not an experience within the criminal justice system that contributes to self-expansion and its effects on their mental associations with the group criminal, but rather their occupational experiences with others only who have had criminal justice experiences.

Parole Officer's Subjective Professional Identity and Orientation

The implicit cognitive consequences of the parole officer-parolee relationship to some degree may be dependent on parole officers' professional characteristics, particularly the importance of, and their basic orientation toward, their role (Seiter and West, 2003; Walters et al., 2007). Consistent with role identity theory, roles that individuals take on, such as occupation, theoretically should affect officers' self-concept and cognitions (Stryker and Burke, 2000), especially on the job where the role is most likely to

be highly salient. Two aspects of parole officers' role are professional orientation and subjective professional identification.

Parole officer orientation has been measured along a continuum ranging from surveillance and strict law enforcement to therapeutic support (Sigler and Mcgraw, 1984; Seiter and West, 2003). Research suggests that professional orientation affects the quality of experiences within the parole officer-parolee relationship (Skeem et al., 2003, 2007; Kennealy et al., 2012; Blasko et al., 2015). Surveillance restricts and controls parolee behavior to ensure that individuals fulfill the responsibilities and conditions of parole (Fulton et al., 1997; Seiter, 2002; Skeem et al., 2003). Also, surveillance is related to low levels of trust and cooperation, which may be related to negative experiences within the parole officer-parolee relationship (Fulton et al., 1997; Seiter, 2002; Skeem et al., 2003). Conversely, parole officers who take on a therapeutic role may be more likely to have positive experiences with their parolees as this orientation requires the parole officer to deeply engage with parolees, and to aid them in addressing criminogenic obstacles, such as mental health problems, substance abuse, physical health conditions, inadequate educational and employment skills, and lack of stable housing (Seiter, 2002; Petersilia, 2003; Travis, 2005; National Research Council, 2008; Tatman and Love, 2010; Blasko et al., 2015). Therefore, professional orientation may moderate the effect of parole officers' relationship experiences with parolees on their implicit associations with parolees. Specifically, among parole officers who take on a therapeutic orientation, a positive experience with a parolee may yield stronger implicit associations with parolees in comparison to those who do not take on a therapeutic orientation.

In addition, parole officers vary in the extent to which they identify with their occupational group. According to social identity theory, people often identify with the social groups to which they belong, including occupational groups (Hogg and Turner, 1987). The extent to which people identify with their occupational group may reflect commitment (Ellemers and Rink, 2005) and is related to positive workplace behavior such as job performance (Meyer et al., 2002; Becker and Kernan, 2003). By extension, the extent to which people identify as a parole officer may influence the types of experiences parole officers have with parolees. For example, those who strongly identify as a parole officer may be more committed to assisting parolees in successful reentry, in turn, impacting their experiences with parolees. Therefore, the extent to which they identify with their occupational group may moderate the effect of parole officers' relationship experiences with parolees on their implicit associations with parolees. Specifically, among parole officers who strongly identify with their occupational group, a positive experience with a parolee may yield stronger implicit associations with parolees in comparison to those who do not strongly identify with their occupational group.

The Present Study

This experiment is the first of its kind to examine implicit associations between the self-perception of criminal justice practitioners, specifically parole officers, and individuals in the criminal justice system. We experimentally manipulated parole officers' memory of either a positive or negative experience with a parolee, then utilized an IAT to measure implicit self-criminal associations. We tested the *a priori* main hypothesis that officers who are reminded of a positive experience with a parolee will exhibit stronger implicit associations with the group criminal in comparison to parole officers who are reminded of a negative experience. Finally, we explored the moderating effect of parole officer orientation and parole officer group identification on the relation between the manipulation (i.e., positive or negative memory task) and implicit self-criminal association.

MATERIALS AND METHODS

Participants and Design

The population of interest were the 258 parole officers in the New Jersey State Parole Board Division of Parole. At the time of the experiment, the parole officers averaged caseloads of approximately 50 parolees per officer. Officers were expected to conduct three face-to-face interactions per month, including one home visit, per individual. However, it is important to note that these were the minimum standards per parolee, and officers had the discretion to meet with parolees as frequently as deemed fit given the circumstances. Officers had the ability to interact with parolees in a variety of contexts ranging from the parole office to counseling locations to transporting parolees to important appointments.

We invited all New Jersey State parole officers to participate; first *via* an on-line platform and then through face-to-face invitations to officers on duty at district offices. Eighty-seven ($n = 18$ online; $n = 69$ district office) active parole officers completed the experiment. All data were collected anonymously, and participants volunteered to complete the study without any incentive. Three participants' data were excluded from analyses due to reaction time error rates on the IAT that were greater than 30% overall or over 40% on any given block as recommended by Greenwald et al. (2003).

The final sample consisted of 84 parole officers (28 females, 55 males, 1 other,[1] $M_{age} = 37.10$, $SD_{age} = 7.39$, age range: 25–54 years). **Table 1** lists the demographics and characteristics of the final sample. Approximately 49% percent of officers identified as White, 26% were Hispanic, 13% were Black, 11% were another ethnicity not listed, and 1% identified as Asian or Pacific Islander. On average, parole officers had been working in their position for nearly nine years ($M_{yearsparole} = 8.64$, $SD_{yearsparole} = 6.634$, range: 1–25 years). Approximately 51% of officers were from the sex offender management unit, 35% of parole officers were from a traditional unit, and 14% were from other units. Approximately 18% of parole officers had a criminal history (i.e., arrest, conviction, and/or incarceration). The experiment employed a one-factor two-level (parole officer-parolee experience condition: positive versus negative) between-participants design.

---

[1] One participant did not indicate their gender.

TABLE 1 | Demographics and characteristics of sample participants (*N* = 84).

Variable	%	*M* (SD)
Gender		
Male	65.5	–
Female	33.3	–
Age	–	37.10 (7.39)
Race/ethnicity		
White	48.8	–
Hispanic/Latino	26.2	–
African American	13.1	
Other	10.7	–
Asian or Pacific Islander	1.2	–
Years as PO	–	8.6 (6.63)
Caseload type		
Sex Offender Management Unit	51.2	–
Traditional	34.5	–
Other	14.3	–
Criminal history	17.9	–

Manipulated Variable

Positive Versus Negative Memory Writing Task

Rooted in the basic principles of self-expansion theory, the purpose of the manipulation was to make salient an experience parole officers have had with a parolee. Parole officers were randomly assigned to vividly re-experience either a positive or a negative experience with a parolee. Participants in the positive experience condition were given the following prompt: "Please imagine a positive interaction you have had with a parolee. Please describe the positive experience as well as your thoughts and feelings during this positive interaction. Please provide details and write freely." Participants in the negative experience condition were given the same prompt, but were asked to focus on a negative interaction. Fifty-two percent of parole officers were randomly assigned to the positive experience condition and 48% to the negative experience condition.

A review of the writing responses confirmed that participants responded to the prompt with positive responses (i.e., used positive words) in the positive experience condition and negative responses (i.e., used negative words) in the negative experience condition, with two exceptions. Two participants (2%) provided a mixed response, one in each condition. Also, we conducted an ANOVA to test if the length of responses (i.e., the number of words) between conditions were different. There was no difference, on average, between the positive (*M* = 66.65, SD = 64.11) and negative (*M* = 65.50, SD = 50.44) conditions, $F(1, 82) = 0.008$, $p = 0.928$.

Measured Variables

Implicit Criminal-Self Associations

The current study used a Single-Category Implicit Association Test (SC-IAT; Greenwald et al., 1998; Karpinski and Steinman, 2006), which uses reaction times to operationalize the strength of implicit mental associations between the self and criminal (Rivera and Veysey, 2015, 2018; Veysey and Rivera, 2017). Parolees are members of the group criminal. Prior to becoming a parolee, a person must first be convicted of a crime and incarcerated, and, therefore, is formally labeled a criminal. Also, during a

first introduction, parole officers are made aware of a parolee's criminal history, thus affirming the parolee's identity as a criminal.

The SC-IAT was administered on a computer and participants were asked to complete four blocks of reaction time trials that were preceded by a set of instructions. Semantic stimuli that represent self, other, and criminal randomly appeared one after the other in the center of the screen. The self-related words were *me, my, mine, I,* and *myself.* The other-related words were *they, them, their, theirs,* and *other* ("self" and "other" words were used in prior studies, e.g., Rivera and Veysey, 2015; Veysey and Rivera, 2017). The criminal-related words were *criminal, felon, lawbreaker, offender, convict, delinquent,* and *prisoner.* The criminal words were pre-tested with a separate adult sample (for a full description, see Rivera and Veysey, 2018).

As each word appeared on the screen, category labels were positioned on the top left and top right of the screen. For half of the task, participants used the "A" key to classify the words that belong to either the "self" or "criminal" category (labels on the top left) and the "K" key to classify the words that belong to the "other" category (label on the top right). The second half of the task was reversed; participants used the "A" key to classify the words that belong to the "self" category (label on the top left) and the "K" key to classify the words that belong to the "other" or "criminal" category (labels on the top right). These tasks were counterbalanced between participants. For each task, participants first read the instructions then completed 17 practice trials, followed by 51 critical trials. For each trial, the target word remained on the screen until participants classified it to one of the three categories on the monitor ("self," "other," or "criminal"). If the participant responded correctly, a new target word appeared. If the participant responded incorrectly, the message "ERROR" appeared on the screen in place of the target word and remained until the participant pressed the correct key.

The SC-IAT was scored in accordance with the procedures outlined by Greenwald et al. (2003) and Karpinski and Steinman (2006). The score is the difference between the reaction times between the self and criminal trials and the other and criminal trials. Relatively higher SC-IAT scores indicate faster reaction times when self-related words and criminal-related words are paired together than when other-related words are paired with criminal words. Thus, a higher IAT score indicates stronger associations between self and criminal.

Subjective Identification With Parole Officers

Participants indicated the extent to which they identified with their professional group. Parole officers were asked to think about their identification with other parole officers and respond to two items ("Being a parole officer is an important part of who I am" and "Being a parole officer is important to my sense of self") on a 7-point scale ranging from *strongly disagree (0)* to *strongly agree (6).* The two items were highly correlated ($r = 0.75$, $p < 0.001$) and therefore combined into a single measure.

Professional Orientation

Guided by the Parole Officer Professional Orientation measure (Fulton et al., 1997), which was designed to assess how parole

officers perform their job functions and goals, participants indicated the extent to which they adopted a professional orientation (i.e., a surveillance vs. therapeutic orientation). Parole officers were asked to rate how they perform their job and respond to two items: (a) one regarding their subjective job role—"The most important aspect of your job is …" on a 5-point scale ranging from *exclusively social work (1)* to *exclusively law enforcement (5)*; and (b) a second regarding their occupational strategy: "The most effective way to change behavior is through…" on a 5-point scale ranging from *exclusively positive reinforcement (1)* to *exclusively punitive sanctions (5)*. Higher scores on both questions indicate a stronger focus on strategies related to law enforcement. While both questions sought to measure the construct of professional orientation, their low correlation, $r(84) = 0.39$, $p < 0.001$, suggests that they are tapping into different elements of professional orientation. Therefore, the items were not combined and were explored independently.

Demographics

Participants completed a demographics and background questionnaire that included variables such as gender, age, race/ethnicity, length of time as a parole officer, type of caseload, and personal criminal history.

Procedure

Participants were informed that the study's purpose was to examine the relation between parole officers' professional experiences and attitudes. First, participants were randomly assigned to and completed the positive or negative experience writing task and then completed all remaining measures in the order listed above. Finally, participants were fully debriefed about the purpose of the study.

RESULTS

See **Table 2** (entire sample) and **Table 3** (by experience condition) for descriptives and zero-order correlations.[2]

---

[2]Zero-order correlations between the demographic variables (age, length of time as a parole officer), subjective parole officer identification, the professional orientation measures (subjective role, occupational strategy), and implicit criminal-self association scores were conducted for the entire sample and by each condition. The relation between the categorical demographic variables (gender, race/ethnicity, type of caseload, criminal history, district office location) and implicit criminal-self association scores across the entire sample and by condition were analyzed using a series of ANOVAs. For gender, participants who identified as "other" were nominal ($n = 1$), so gender was coded was coded ($1 = male$; $0 = female$, *other*) to facilitate the interpretation of the results; race/ethnicity was coded ($1 = African$ $American/Black$ and $Hispanic$; $0 = all$ $others$), type of caseload was coded ($1 = sex$ $offender$ $unit$; $0 = all$ $other$ $units$), criminal history was coded ($1 = arrested$, $convicted$, and/or $incarcerated$; $0 = was$ not $arrested$, $convicted$, and/or $incarcerated$), and district office location was coded ($1 = online$ $participant/unknown$ $district$; $2 = Northern$ New $Jersey$, $3 = Central/Southern$ New $Jersey$). The ANOVAs and zero-order correlation analyses revealed that no demographic and measured variables were statistically significantly related to implicit criminal-self association scores.

To test our main hypothesis of the effect of making parole officer experiences salient on implicit associations, SC-IAT scores (criminal-self association strength) were subjected to a one-way analysis of variance (ANOVA). As predicted, participants who were reminded of a positive experience with their parolees ($M = -0.05$, $SD = 0.14$) exhibited stronger associations between criminal and self when compared to those who were reminded of a negative experience with their parolees ($M = -0.13$, $SD = 0.15$), $F(1, 82) = 5.20$, $p = 0.025$, $d = 0.50$ (medium effect size).[3] These findings support our hypothesis that parole officers who are reminded of positive experiences exhibit a stronger implicit association (i.e., a reduced cognitive distance) between criminal and self in comparison to parole officers who are reminded of negative experiences.

We next explored if individual differences in subjective group identification or the two professional orientations (i.e., subjective job role and occupational strategy) moderated the effect of past experiences on criminal-self associations. To this end, we conducted three hierarchical multiple regression analyses in which scores on the three individual differences measures (mean-centered) and parole officer-parolee experience condition (coded $0 = negative$ $experience$, $1 = positive$ $experience$) were entered in the first step and their interaction term in the second step. Consistent with our main hypothesis results above, the main effect of the Experience Condition was significant across the three models, $\Delta Fs(2, 81) > 2.81$, $ps < 0.036$, $R^2s > 0.07$, $\beta s > 0.23$, $ps < 0.025$. However, there was no significant main effect of Subjective Parole Officer Identification, $\Delta F(2, 81) = 3.29$, $p = 0.247$, $R^2 = 0.08$, $\beta = -0.13$, $p = 0.247$, and no significant Subjective Parole Officer Identification X Experience Condition interaction, $\Delta F(3, 80) = 0.66$, $p = 0.417$, $R^2 = 0.08$, $\beta = 0.12$, $p = 0.417$; no significant main effect of Subjective Job Role, $\Delta F(2, 81) = 2.82$, $p = 0.493$, $R^2 = 0.07$, $\beta = -0.08$, $p = 0.693$, and no significant Subjective Job Role X Experience Condition interaction, $\Delta F(3, 80) = 0.01$, $p = 0.929$, $R^2 = 0.07$, $\beta = -0.01$, $p = 0.929$; and, finally, no significant main effect of Occupational Strategy, $\Delta F(2, 81) = 3.72$, $p = 0.146$, $R^2 = 0.08$, $\beta = -0.16$, $p = 0.063$, and no significant Occupational Strategy X Experience Condition interaction, $\Delta F(3, 80) = 1.40$, $p = 0.241$, $R^2 = 0.10$, $\beta = 0.17$, $p = 0.241$.

GENERAL DISCUSSION

This research extends earlier findings regarding implicit self-expansion in general, contributes to the sparse literature on employment-based self-expansion, and provides evidence of a self-expansion mechanism to a highly stigmatized group. The study examined the extent to which parole officers implicitly associate self with the group criminal as a function of a memory-based manipulation. The primary finding that officers who are reminded of a positive experience with a parolee demonstrate stronger implicit self-criminal association on the

---

[3]We conducted the same analyses using the subsample of participants who did not have a criminal history ($n = 69$). Results were similar to those with the full sample (positive experience: $M = -0.05$, $SD = 0.15$, negative experience: $M = -0.12$, $SD = 0.16$; $F(1, 67) = 4.11$, $p = 0.047$, $d = 0.45$, medium effect size).

TABLE 2 | Zero-order correlations and descriptives for all participants (N = 84).

	1	2	3	4	5	6
1. Implicit criminal-self association	–					
2. Subjective parole officer identification	−0.13	–				
3. Subjective job role	−0.04	0.12	–			
4. Occupational strategy	−0.18	−0.12	0.39*	–		
5. Age	−0.02	−0.39*	−0.01	−0.14	–	
6. Years as PO	−0.06	−0.48*	0.06	0.05	0.67*	–
M	−0.09	4.40	3.15	2.87	37.10	8.64
SD	0.15	1.41	0.63	0.58	7.39	6.63

*$p < 0.01$.

TABLE 3 | Zero-order correlations and descriptives by experience condition.

	1	2	3	4	5	6	M	SD
1. Implicit criminal-self association	–	−0.03	−0.09	−0.02	−0.22	−0.24	−0.05	0.14
2. Subjective parole officer identification	−0.22	–	−0.09	−0.19	−0.43**	−0.57**	4.40	1.34
3. Subjective job role	−0.06	0.31*	–	0.46**	0.04	0.07	3.23	0.64
4. Occupational strategy	−0.29	−0.04	0.35*	–	0.85	0.87	2.82	0.54
5. Age	0.13	−0.36*	−0.11	−0.23	–	0.71*	37.95	7.74
6. Years as PO	0.14	−0.39*	0.07	0.07	0.66**	–	8.25	6.55
M	−0.13	4.41	3.08	2.93	36.15	9.08	–	–
SD	0.15	1.51	0.62	0.62	6.95	6.79	–	–

Numbers above the diagonal are data from participants in positive experience condition (n = 44). Numbers below the diagonal are data from participants in negative experience condition (n = 40).
*$p < 0.05$; **$p < 0.01$.

IAT than those who were reminded of a negative experience with parolees supports other group-based implicit self-expansion studies. These former studies, however, target one or more cross-group friends and self-expansion to their groups (e.g., having one or more Latinx friends and the group Latinxs; Page-Gould et al., 2010). The present study, like McIntyre et al.' (2014), is based within a professional work environment. While McIntyre et al. (2014) used "job" as a proxy group identity, the present research tested self-expansion to the client group of parolees after manipulating a memory in which officers imagined an interaction with an individual parolee on their caseload.

For self-expansion at the group level to occur, interactions with one or more members of a group must be: (1) close, (2) frequent, and (3) positive. We suggest that these self-expansion requirements occur in parole officers' relationships with a parolee in the workplace. Closeness is promoted through their relationship interactions with individual parolees who represent the group and its characteristics. The role of the parole officer requires that they interact with parolees for numerous hours a week. During this time parole officers engage with parolees and support them to lead pro-social lives (e.g., assistance with treatment, job placement, and housing). High frequency is assumed due to the time spent and average number of parole officer-parolee interactions on any given day. And, finally, positive (vs. negative) interactions were experimentally manipulated in the present research. Under these conditions, the results demonstrated that a single recalled positive versus

negative memory had a significant effect on implicit associations between parole officers' self and the group criminals.

While our main hypothesis was supported, the nature of the effect needs further understanding. Self-expansion is about taking on aspects of the close other or of the close other's group. In our research, we infer evidence for self-expansion from parole officers' stronger implicit self-criminal associations on the IAT following the reminder of a past positive relative to a negative experience with a parolee. The mean IAT score in the positive experience condition, however, was near the midpoint of the scale, whereas the mean IAT score in the negative experience condition significantly more. Since the IAT in general is a measure of relative associations, presently self-criminal associations relative to other-criminal associations, its scoring is a function of the difference between reaction times to categories simultaneously paired on the computer screen. Thus, another yet complementary suggestion from our data is that parole officers in the negative experience condition increased their cognitive distance between their self-concept and the group parolee, whereas those in positive experience condition demonstrated a decrease in this cognitive distance.

The potential positive impact of these findings is related to the malleability of this effect. If a single positive reminder affects implicit associations between the officer and the parolee, the quality of the parole officer-parolee relationship may be similarly affected. In general, self-expansion results in the acquisition of perspectives, causing the self to be concerned with the needs of the other (Wegner, 1980; Deutsch and Mackesy, 1985;

Aron and Aron, 1986). Thus, even in relationships with people belonging to stigmatized groups, self-expansion results in the reduction of cognitive distance. Within this definition, self-expansion may be a cognitive driver of empathy, which requires one to take the perspective of the other (Selman, 1980; Wegner, 1980; Aron et al., 2004; Batson, 2009).

Empathy is an essential component of the therapeutic relationship, and it can facilitate collaboration, trust, and understanding between the practitioner and client (Murphy and Baxter, 1997) and is related to positive client outcomes (Taxman, 2002; Ross et al., 2008; Taxman and Ainsworth, 2009; McCambridge et al., 2011). Empathy is also related to bonding, feelings of closeness, attitudes towards stigmatized groups (Finlay and Stephan, 2000), and changes in behavior, such as helping the other (Batson et al., 2002). Therefore, it is plausible that empathy follows from self-expansion and, as such, positive reminders and messaging may be used to improve the quality of the interaction and, thus, improve parolee success (Gunnison and Helfgott, 2011).

Although the present study does not directly measure empathy, our data suggest that one way to potentially bolster empathy between parole officers and parolees is to provide environments that foster positive interactions. Practices such as Motivational Interviewing have been increasingly adopted by corrections agencies due to their positive effects on parolees (Dowden and Andrews, 2004; Landenberger and Lipsey, 2005), and have an effect on parole officers as well (Iarussi and Powers, 2018). Because Motivational Interviewing promotes rapport and requires the parole officer to take the perspective of the parolee, it is therefore possible that Motivational Interviewing also provides parole officers with opportunities to acquire the perspective of their parolees and promote helping behaviors that support parolees' successful reentry and long-term desistance.

In the absence of these experiences, reminding parole officers of positive or successful experiences with parolees may serve to strengthen implicit self-criminal associations, which can have occupational behavioral effects. Positive experiences can be promoted through the use of bulletin boards that showcase events in which parole officers and parolees work together or celebrate the accomplishments of parolees (e.g., educational or employment successes). This practice may serve as a reminder that successful and positive experiences between parole officers and parolees do happen and can result in positive outcomes. Future research should examine the relation between implicit self-criminal associations and empathy and its downstream consequences for parole officer behaviors and other outcomes.

Additionally, we explored the moderating role of parole officers' subjective professional identity (i.e., professional orientation and parole officer identification) on the relation between past experiences and implicit self-criminal associations. None of the measures of subjective professional identity were related to implicit associations, nor did they moderate the effect of past experiences on implicit associations. This suggests that making a positive memory salient is powerful enough to override (i.e., regardless of) individual-level professional characteristics.

While self-expansion may impact parole officer performance and parolee outcomes, it is important to note that the present

study cannot answer the fundamental question about whether implicit cognitions in this setting affect parole officer behavior. To be clear, this study is not longitudinal nor does it include measures of actual parole officer behavior or parolee success. However, research suggests that self-expansion can influence not only cognition but also behavior (Aron et al., 1991; Aron and Aron, 1996; Davis et al., 1996; Cialdini et al., 1997). Moreover, implicit social cognitions predict behavioral actions, often having greater explanatory power in behavioral outcomes than explicit social cognitions (Greenwald et al., 2009; Charlesworth and Banaji, in press). Further, research demonstrates that occupations which promote self-expansion are related to increased job satisfaction and commitment (McIntyre et al., 2014). Based on prior work, we speculate that self-expansion may have implications for the well-being of parole officers as well.

This study utilizes data from parole officers, a unique and often inaccessible sample. However, obtaining data from this sample is not without its limitations. New Jersey employs nearly 600 parole officers across 16 parole offices throughout the state (New Jersey State Parole Board, 2020). First, the sample size was relatively small due to the time consuming and costly nature of in-person data collection from on-duty officers during department-wide meetings. The researcher responsible for data collection visited several parole offices numerous times over the course of two years. The intention was to achieve a high response rate. However, some officers were not willing to participate in this research for various reasons ranging from lack of interest to job demands. Ideally, future research with a larger sample could expand on the current research. Second, because the study was limited to ten minutes per participant, the amount of data that could be obtained was limited. For example, measures of explicit association were not collected. While implicit social cognition research demonstrates that implicit and explicit associations with stigmatized groups are often poorly correlated or uncorrelated (Greenwald et al., 2009), self-expansion research demonstrates that implicit and explicit associations are often correlated (Coats et al., 2000; Page-Gould et al., 2010). Future research should examine the effect of experiences on both implicit *and* explicit associations to provide a comprehensive understanding of self-expansion with this particular group.

Finally, while this study examines the effect of reminders of past experiences with a single parolee, this study does not answer the question of repeat experiences. Do parole officers who have a positive experience with a parolee tend to repeatedly engage in positive experiences with their parolees? If so, parole officers who frequently engage in positive experiences may exhibit self-expansion even in the absence of contexts which facilitate positive experiences, and, therefore, translate to continuous positive experiences with parolees. This is squarely in line with research that demonstrates that individuals who had high quality relationships with cross-group friends did not exhibit stress or anxiety following conflict (Page-Gould, 2012). For parole officers, negative experiences with parolees may occur, but parole officers who self-expand may be better equipped to overcome such experiences, thereby potentially improving their relationships with parolees. Despite these limitations, the present data

suggest a promising step in understanding the conditions which strengthen implicit associations with parolees.

CONCLUSION

The present study is the first experiment at the intersection of psychology and criminology to apply the self-expansion model and implicit social cognition measurements within a criminal justice setting to understand the cognitive effects of a parole officer's relationship with parolees. Results demonstrate that parole officers who are reminded of positive experiences with a parolee exhibit stronger associations between themselves and the group parolee in comparison to those reminded of a negative experience. Positive experiences can increase parole officers' positive perceptions of parolees, as well as bolster their overall working relationship with parolees. This may lead to benefits such as increased job satisfaction and may have downstream consequences for parolees; namely, desistance and successful reentry – a primary goal of parole and of the criminal justice system.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Rutgers University IRB. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

Abadinsky, H. (2009). *Probation and Parole: Theory and Practice.* Upper Saddle River, NJ: Pearson Prentice Hall.

Andrews, D. A. (2011). "The principles of effective correctional programs," in *Correctional Contexts: Contemporary and Classical Readings.* eds. E. J. Latessa and A. M. Holsinger (New York, NY: Oxford University Press), 228–237.

Aron, A., and Aron, E. N. (1986). *Love and the Expansion of Self: Understanding Attraction and Satisfaction.* Washington: Hemisphere Publishing Corp/Harper & Row Publishers.

Aron, A., and Aron, E. N. (1996). "Self and self expansion in relationships," in *Knowledge Structures in Close Relationships: A Social Psychological Approach.* eds. G. J. O. Fletcher and J. Fitness (Mahwah, NJ: Lawrence Erlbaum Associates), 325–344.

Aron, A., and Aron, E. N. (1997). "Self-expansion motivation and including other in the self." in *Handbook of Personal Relationships. 2nd Edn., Vol.* 1. W. Ickes (Section Ed.) and S. Duck (Ed.), (Chichester, New York: Wiley), 251–270.

Aron, A., Aron, E. N., and Norman, C. (2004). "Self-expansion model of motivation and cognition in close relationships and beyond," in *Self and Social Identity.* eds. M. B. Brewer and M. Hewstone (Blackwell Publishing), 99–123.

Aron, A., Aron, E. N., and Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *J. Pers. Soc. Psychol.* 63, 596–612. doi: 10.1037/0022-3514.63.4.596

Aron, A., Aron, E. N., Tudor, M., and Nelson, G. (1991). Close relationships as including other in the self. *J. Pers. Soc. Psychol.* 60, 241–253. doi: 10.1037/0022-3514.60.2.241

Aron, A., and McLaughlin-Volpe, T. (2001). "Including others in the self: extensions to own and partner's group memberships," in *Individual Self, Relational Self, and Collective Self: Partners, Opponents, or Strangers?* eds. C. Sedikides and M. B. Brewer (Philadelphia, Hove England: Psychology Press), 89–109.

Aron, A., Paris, M., and Aron, E. N. (1995). Falling in love: prospective studies of self-concept change. *J. Pers. Soc. Psychol.* 69, 1102–1112. doi: 10.1037/0022-3514.69.6.1102

Asencio, E. K. (2011). Familiarity, legitimation, and frequency: The influence of others on the criminal self-view. *Sociol. Inq.* 81, 34–52. doi: 10.1111/j.1475-682X.2010.00352.x

Batson, C. D. (2009). "These things called empathy: eight related but distinct phenomena," in *The Social Neuro-Science of Empathy.* eds. J. Decety and W. Ickes (Cambridge: MIT Press), 3–15.

Batson, C. D., Chang, J., Orr, R., and Rowland, J. (2002). Empathy and attitudes: can feeling for a member of a stigmatized group improve feelings toward the group? *J. Pers. Soc. Psychol.* 28, 1656–1666. doi: 10.1177/014616702237647

Becker, T. E., and Kernan, M. C. (2003). Matching commitment to supervisors and organizations to in-role and extra-role performance. *Hum. Perform.* 16, 327–348. doi: 10.1207/S15327043HUP1604_1

Blasko, B. L., Friedmann, P. D., Rhodes, A. G., and Taxman, F. S. (2015). The parolee–parole officer relationship as a mediator of criminal justice outcomes. *Crim. Justice Behav.* 42, 722–740. doi: 10.1177/0093854814562642

Boduszek, D., Adamson, G., Shevlin, M., Mallett, J., and Hyland, P. (2013). Criminal social identity of recidivistic prisoners: the role of self-esteem, family and criminal friends. *J. Police Crim. Psychol.* 28, 15–25. doi: 10.1007/s11896-012-9105-7

Brenner, M. (1973). The next-in-line effect. *J. Verbal Learn. Verbal Behav.* 12, 320–323. doi: 10.1016/S0022-5371(73)80076-3

Bui, H. N., and Morash, M. (2010). The impact of network relationships, prison experiences, and internal transformation on women's success after prison release. *J. Offender Rehabil.* 49, 1–22. doi: 10.1080/10509670903435381

Chamberlain, A. W., Gricius, M., Wallace, D. M., Borjas, D., and Ware, V. M. (2017). Parolee–parole officer rapport: does it impact recidivism? *Int. J. Offender Ther. Comp. Criminol.* 62, 3581–3602. doi: 10.1177/0306624X17741593

Charlesworth, T. E. S., and Banaji, M. R. (in press). "The relationship of implicit social cognition and discriminatory behavior," in *Handbook on Economics of Discrimination and Affirmative Action.* ed. A. Deshpande (New York: Springer).

Cialdini, R. B., Brown, S. L., Lewis, B. P., Luce, C., and Neuberg, S. L. (1997). Reinterpreting the empathy–altruism relationship: when one into one equals oneness. *J. Pers. Soc. Psychol.* 73, 481–494. doi: 10.1037/0022-3514.73.3.481

Clark, M. S., and Mills, J. (1979). Interpersonal attraction in exchange and communal relationships. *J. Pers. Soc. Psychol.* 37, 12–24. doi: 10.1037/0022-3514.37.1.12

Coats, S., Smith, E. R., Claypool, H. M., and Banner, M. J. (2000). Overlapping mental representations of self and in-group: reaction time evidence and its relationship with explicit measures of group identification. *J. Exp. Soc. Psychol.* 36, 304–315. doi: 10.1006/jesp.1999.1416

Cohen, A., and Yang, C. S. (2019). Judicial politics and sentencing decisions. *Am. Econ. J. Econ. Pol.* 11, 160–191. doi: 10.1257/pol.20170329

Davis, M. H., Conklin, L., Smith, A., and Luce, C. (1996). Effect of perspective taking on the cognitive representation of persons: a merging of self and other. *J. Pers. Soc. Psychol.* 70, 713–726. doi: 10.1037/0022-3514.70.4.713

DeMichele, M. T. (2007). *Probation and Parole's Growing Caseloads and Workload Allocation: Strategies for Managerial Decision Making.* Lombardi, IL: The American Probation & Parole Association.

Deutsch, F. M., and Mackesy, M. E. (1985). Friendship and the development of self-schemas: The effects of talking about others. *Personal. Soc. Psychol. Bull.* 11, 399–408. doi: 10.1177/0146167285114006

Dowden, C., and Andrews, D. A. (2004). The importance of staff practice in delivering effective correctional treatment: a meta-analytic review of core correctional practice. *Int. J. Offender Ther. Comp. Criminol.* 48, 203–214. doi: 10.1177/0306624X03257765

Ellemers, N., and Rink, F. (2005). Identity in work groups: the beneficial and detrimental consequences of multiple identities and group norms for collaboration and group performance. *Soc. Identif. Groups* 22, 1–41. doi: 10.1016/S0882-6145(05)22001-5

Fielding-Miller, R., Cooper, H. L. F., Caslin, S., and Raj, A. (2020). The interaction of race and gender as a significant driver of racial arrest disparities for African American men. *J. Urban Health* 97, 112–122. doi: 10.1007/s11524-018-00325-1

Finlay, K. A., and Stephan, W. G. (2000). Reducing prejudice: the effects of empathy on intergroup attitudes. *J. Appl. Soc. Psychol.* 30, 1720–1737. doi: 10.1111/j.1559-1816.2000.tb02464.x

Fulton, B., Stichman, A., Travis, L., and Latessa, E. (1997). Moderating probation and parole officer attitudes to achieve desired outcomes. *Prison J.* 77, 295–312. doi: 10.1177/0032855597077003005

Gawronski, B., and Payne, B. K. (Eds.). (2011). *Handbook of Implicit Social Cognition: Measurement, Theory, and Applications.* New York: Guilford Press.

Gibbons, S. G., and Rosecrance, J. D. (2005). *Probation, Parole, and Community Corrections in the United States.* Boston: Pearson Allyn & Bacon.

Giordano, P. C., Schroeder, R. D., and Cernkovich, S. A. (2007). Emotions and crime over the life course: a neo-Meadian perspective on criminal continuity and change. *Am. J. Sociol.* 112, 1603–1661. doi: 10.1086/512710

Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* 74, 1464–1480. doi: 10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., and Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *J. Pers. Soc. Psychol.* 85, 197–216. doi: 10.1037/0022-3514.85.2.197

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., and Banaji, M. R. (2009). Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *J. Pers. Soc. Psychol.* 97, 17–41. doi: 10.1037/a0015575

Gunnison, E., and Helfgott, J. B. (2011). Factors that hinder offender reentry success: a view from community corrections officers. *Int. J. Offender Ther. Comp. Criminol.* 55, 287–304. doi: 10.1177/0306624X09360661

Hogg, M. A., and Turner, J. C. (1987). Intergroup behaviour, self-stereotyping and the salience of social categories. *Br. J. Soc. Psychol.* 26, 325–340. doi: 10.1111/j.2044-8309.1987.tb00795.x

Huebner, B. M., and Bynum, T. S. (2008). The role of race and ethnicity in parole decisions. *Criminology* 46, 907–938. doi: 10.1111/j.1745-9125.2008.00130.x

Iarussi, M. M., and Powers, D. F. (2018). Outcomes of motivational interviewing training with probation and parole officers: findings and lessons learned. *Fed. Probat.* 82, 28–35.

Ireland, C., and Berg, B. (2007). Women in parole: gendered adaptations of female parole agents in California. *Women Crim. Just.* 18, 131–150. doi: 10.1300/J012v18n01_05

Jolliffe, D., and Farrington, D. P. (2007). A rapid evidence assessment of the impact of mentoring on re-offending: A summary. Available at: https://www.ojp.gov/ncjrs/virtual-library/abstracts/rapid-evidence-assessment-impact-mentoring-re-offending-summary (Accessed June 3, 2022).

Karpinski, A., and Steinman, R. B. (2006). The single category implicit association test (SC-IAT) as a measure of implicit social cognition. *J. Pers. Soc. Psychol.* 91, 16–32. doi: 10.1037/0022-3514.91.1.16

Kennealy, P. J., Skeem, J. L., Manchak, S. M., and Louden, J. E. (2012). Firm, fair, and caring officer-offender relationships protect against supervision failure. *Law Hum. Behav.* 36, 496–505. doi: 10.1037/h0093935

Krueger, J. (2001). "Social categorization, the psychology of," in *International Encyclopedia of the Social and Behavioral Sciences.* eds. N. J. Smelser and P. B. Baltes (Amsterdam: Elsevier Ltd), 14219–14223.

Landenberger, N. A., and Lipsey, M. W. (2005). The positive effects of cognitive–behavioral programs for offenders: a meta-analysis of factors associated with effective treatment. *J. Exp. Criminol.* 1, 451–476. doi: 10.1007/s11292-005-3541-7

McCambridge, J., Day, M., Thomas, B. A., and Strang, J. (2011). Fidelity to motivational interviewing and subsequent cannabis cessation among adolescents. *Addict. Behav.* 36, 749–754. doi: 10.1016/j.addbeh.2011.03.002

McIntyre, K. P., Mattingly, B. A., Lewandowski, G. W. Jr., and Simpson, A. (2014). Workplace self-expansion: implications for job satisfaction, commitment, self-concept clarity, and self-esteem among the employed and unemployed. *Basic Appl. Soc. Psychol.* 36, 59–69. doi: 10.1080/01973533.2013.856788

Meyer, J. P., Stanley, D. J., Herscovitch, L., and Topolnytsky, L. (2002). Affective, continuance, and normative commitment to the organization: a meta-analysis of antecedents, correlates, and consequences. *J. Vocat. Behav.* 61, 20–52. doi: 10.1006/jvbe.2001.1842

Morash, M., Kashy, D. A., Smith, S. W., and Cobbina, J. E. (2014). The effects of probation or parole agent relationship style and women offenders' criminogenic needs on offenders' responses to supervision interactions. *Crim. Justice Behav.* 42, 412–434. doi: 10.1177/0093854814551602

Murphy, C. M., and Baxter, V. A. (1997). Motivating batterers to change in the treatment context. *J. Interpers. Violence* 12, 607–619. doi: 10.1177/088626097012004009

National Research Council (2008). Parole, desistance from crime, and community integration. Available at: https://www.nap.edu/catalog/11988/parole-desistance-from-crime-and-community-integration (Accessed June 3, 2022).

New Jersey State Parole Board (2020). Annual Report. Available at: https://www.state.nj.us/parole/docs/reports/AnnualReport2019.pdf (Accessed June 3, 2022).

Oudekerk, B., and Kaeble, D. (2021). *Probation and parole in the United States, 2019.* Bulletin (NCJ 256092). Bureau of Justice Statistics.

Page-Gould, E. (2012). To whom can I turn? Maintenance of positive intergroup relations in the face of intergroup conflict. *Soc. Psychol. Personal. Sci.* 3, 462–470. doi: 10.1177/1948550611426937

Page-Gould, E., Mendoza-Denton, R., Alegre, J. M., and Siy, J. O. (2010). Understanding the impact of cross-group friendship on interactions with novel outgroup members. *J. Pers. Soc. Psychol.* 98, 775–793. doi: 10.1037/a0017880

Petersilia, J. (2003). *When Prisoners Come Home: Parole and Prisoner Reentry.* New York: Oxford University Press.

Rivera, L. M., and Veysey, B. M. (2015). Criminal justice system involvement and gender stereotypes: consequences and implications for women's implicit and explicit criminal identities. *Albany Law Rev.* 78, 1109–1128.

Rivera, L. M., and Veysey, B. M. (2018). Implicit self-criminal cognition and its relation to criminal behavior. *Law Hum. Behav.* 42, 507–519. doi: 10.1037/lhb0000300

Ross, E. C., Polaschek, D. L., and Ward, T. (2008). The therapeutic alliance: a theoretical revision for offender rehabilitation. *Aggress. Violent Behav.* 13, 462–480. doi: 10.1016/j.avb.2008.07.003

Seiter, R. P. (2002). Prisoner reentry and the role of parole officers. *Fed. Probat.* 66, 50–54.

Seiter, R. P., and West, A. D. (2003). Supervision styles in probation and parole: an analysis of activities. *J. Offender Rehabil.* 38, 57–75. doi: 10.1300/J076v38n02_04

Selman, R. L. (1980). *The Growth of Interpersonal Understanding.* New York: Academic Press.

Sigler, R. T., and Mcgraw, B. (1984). Adult probation and parole officers: influence of their weapons, role perceptions and role conflict. *Crim. Justice Rev.* 9, 28–32. doi: 10.1177/073401688400900105

Skeem, J. L., Encandela, J., and Louden, J. E. (2003). Perspectives on probation and mandated mental health treatment in specialized and traditional probation departments. *Behav. Sci. Law* 21, 429–458. doi: 10.1002/bsl.547

Skeem, J. L., Louden, J. E., Polaschek, D., and Camp, J. (2007). Assessing relationship quality in mandated community treatment: blending care with control. *Psychol. Assess.* 19, 397–410. doi: 10.1037/1040-3590.19.4.397

Smith, E. R., Coats, S., and Walling, D. (1999). Overlapping mental representations of self, in-group, and partner: further response time evidence and a connectionist model. *Personal. Soc. Psychol. Bull.* 25, 873–882. doi: 10.1177/0146167299025007009

Smith, E. R., and Henry, S. (1996). An in-group becomes part of the self: response time evidence. *Personal. Soc. Psychol. Bull.* 22, 635–642. doi: 10.1177/0146167296226008

Stone, R., Morash, M., Goodson, M., Smith, S., and Cobbina, J. (2016). Women on parole, identity processes, and primary desistance. *Fem. Criminol.* 13, 382–403. doi: 10.1177/1557085116670004

Stryker, S., and Burke, P. J. (2000). The past, present and future of an identity theory. *Soc. Psychol. Q.* 63, 284–297. doi: 10.2307/2695840

Tatman, A. W., and Love, K. M. (2010). An offender version of the working alliance inventory-short revised. *J. Offender Rehabil.* 49, 165–179. doi: 10.1080/10509671003666560

Taxman, F. S. (2002). Supervision-exploring the dimensions of effectiveness. *Fed. Probat.* 66, 14–27.

Taxman, F. S., and Ainsworth, S. (2009). Correctional milieu: the key to quality outcomes. *Vict. Offenders* 4, 334–340. doi: 10.1080/15564880903227347

Tesser, A., Millar, M., and Moore, J. (1988). Some affective consequences of social comparison and reflection processes: the pain and pleasure of being close. *J. Pers. Soc. Psychol.* 54, 49–61. doi: 10.1037/0022-3514.54.1.49

Travis, J. (2005). *But They All Come Back: Facing the Challenges of Prisoner Reentry.* Washington, DC: Urban Institute Press.

Veysey, B. M., and Rivera, L. M. (2017). Implicit criminal identity and age: implications for criminal persistence and desistance. *Crim. Justice Behav.* 44, 1249–1261. doi: 10.1177/0093854817722173

Walters, S. T., Clark, M. D., Gingerich, R., and Meltzer, M. L. (2007). *Motivating Offenders to Change: A Guide for Probation and Parole.* Washington, DC: US Department of Justice, National Institute of Corrections.

Wegner, D. M. (1980). "The self in prosocial action," in *The Self in social Psychology.* eds. D. M. Wegner and R. R. Vallacher (New York: Oxford University Press), 131–157.

![frontiers] | Frontiers in Psychology

*CORRESPONDENCE
Mercedes Novo
Mercedes.novo@usc.es

Study of informal reasoning in judicial agents in sexual aggression cases

Xaviera Camplá, Yurena Gancedo, Jéssica Sanmarco, Álvaro Montes and Mercedes Novo*

Unidad de Psicología Forense, Facultad de Psicología, Universidad de Santiago de Compostela, Santiago, Spain

Background/Objective: Judicial decisions must rest on formal reasoning. Nevertheless, informal reasoning sources (cognitive and motivational biases) were observed in judicial judgment making. Literature has identified sexual aggression cases as the most favorable for informal reasoning. Thus, a field study was designed with the aim of assessing the incidence and effects of cognitive and motivational biases in judicial agents in a case to rape to a woman.

Methods: As for this, Chilean judicial agents ($N = 217$) assessed an allegation (weak evidence) of sexual assault in a case where the perpetrator was known or unknown to the victim. The judicial agents answered to a measure of the myths about sexual aggression, the attribution of responsibility to complainant, the attribution of responsibility to accused, the attribution of credibility to the complainant testimony, the attribution of a nature of a rape to the alleged facts and an estimation of the probability of false/unfounded accusations.

Results: The results revealed an estimation of false/unfounded accusations of sexual aggression significantly higher than the mean of the best estimates, but into the upper limit of the best estimates; that the studied population did not share, in general, the myths about sexual aggression; and that the sources of attributional biases were driven in favor and against the complainant. Nevertheless, the case study showed that a large number of judicial agents participated of an overestimation of the probabilities of false or unfounded allegations, and of the myths about sexual aggressions and of attributional biases against the complainant.

Conclusion: In conclusion, informal reasoning sources were observed in judicial agents when only formal reasoning should prevail. Thus, judicial agents should be trained to control these sources of bias substituting them by formal reasoning (evidence).

KEYWORDS

cognitive biases, motivational biases, judgment making, myths about sexual aggression, formal reasoning, informal reasoning

Introduction

The literature has theorized and highlighted the impact of cognitive and motivational biases on judgment making (Kruglanski and Azjen, 1983; Montibeller and von Winterfeldt, 2015). Cognitive biases are due to the limitations of the human being to process all the information, which lead to direct attention toward certain information and discard other that could be equally relevant (Nisbett and Ross, 1980). In sum, cognitive biases arise from limitations in human information-processing substituting an expected formal reasoning (for a review see Kruglanski and Azjen, 1983). Thus, in contexts in which judgment making must rest on formal reasoning, such as legal judgments, these biases should not have a place. However, research on sexual violence has shown that judgment making about it is influenced by myths about sexual assault (Lonsway and Fitzgerald, 1994), which serve as descriptive or prescriptive cognitive tools about the causes, contexts and consequences of sexual assaults, as well as perpetrators, victims and their interaction. These types of cognitive schemes allow access to heuristic representations of information to judgment making about sexual violence (McKimmie et al., 2020). Thus, individuals sharing these myths use them to deny, minimize, overgeneralize or justify the violence of men against women (Gerger et al., 2007), while favoring risky sexual behaviors (Álvarez-Muelas et al., 2020), and a different evaluation of the same behaviors by men and women (sexual double standard; Álvarez-Muelas et al., 2021), a contingency with a high prevalence in the Hispanic context (Martínez-Gómez et al., 2021). Conversely, the perception of the complainant as chaste, respectable or sober i.e., gender victim stereotypes (counter-myths) is related to the opposite trend (Schuller et al., 2010). For this purpose, judicial judgment makers, in line with the judicial reasoning (law of precedent), are impelled to use the assignment or not of credibility (reliability in scientific judgment making models; Kaplan et al., 1978) to the testimony of the complainant (Arce et al., 2000; Du Mont et al., 2003; Page, 2007, 2010; Schuller et al., 2010; Anders and Christopher, 2011; Hine and Murphy, 2017).

On the other hand, motivational biases are characterized by a tendency to form and hold beliefs that fulfill the needs of the individual or overestimate the perceived degree of controllability of the environment (Novo and Seijo, 2010). Among the motivational biases, attributional biases have been related (Burger, 1981), which are used as means for judgment making through the attribution of responsibility or credibility (judicial task). In judgment making about sexual assault, expectations persist about how a real victim of a sexual assault behaves, which are a breeding ground for attributional biases (Smith and Skinner, 2017) that are not present in other types of crimes not including the issue of consent that plays a critical role in rape cases (Bieneck and Krahé, 2011). Thus, these sources of bias are used to attribute responsibility to the complainant and aggressor, to attribute credibility to the testimony of the

complainant (proof of charge, while the testimony of the accused is not evidence of the charge, so from its evaluation cannot be derived criminal responsibility) or to attribute to the alleged facts nature of a sexual assault (most sexual assault complaints have to be resolved on whether the facts are an assault or a consensual relationship).

Biases acquire a functional nature as the evidence is weak and lose it when the evidence is strong (Visher, 1987; Kahneman, 2011; Butterfield and Bitter, 2019; Nitschke et al., 2019). In crimes committed in the private sphere (e.g., domestic violence, sexual assault) there are, therefore, few media of burden of proof beyond the testimony of the complainant and the evaluation of the damage to the complainant (Arce, 2017). Hence, trials in these crimes are conducive to the manifestation of bias.

These types of cognitive and motivational biases affect not only judicial judgment making of laypersons (the vast majority of research has been carried out with jurors, i.e., laypersons; Schuller et al., 2010), but also of law professionals (Fitzmaurice and Pease, 1986; Saks and Kidd, 1986; Fariña et al., 2002; Arce et al., 2003). In the European and Anglo-Saxon context, the presence and impact of myths about sexual violence on police and judicial agent samples has been widely documented (Camplá et al., 2017; Smith and Skinner, 2017; Temkin et al., 2018). Nevertheless, not all have the same meaning in this type of population. Thus, one of the most prevalent myths reported in the population of professionals with procedural or judicial competences (Police, Prosecutors, Judges) is the one referring to false (deliberately fabricated)/unfounded (not deliberately fabricated, baseless, groundless) accusations, based on the belief that women allege for revenge, for profit or regret (Lisak et al., 2010; Lonsway, 2010; Ferguson and Malouff, 2016). As well, the Chilean Law of Precedent establishes that the testimony of the complainant is not sufficient evidence if there is some benefit, revenge or repentance in it. Hence, if this were the only proof against the accused, the judicial criterion would classify the case as evidenceless and it would be closed or archived, expanding in this group unfounded to evidenceless.

In any case, the estimation of the probability of false or unfounded accusations has been considered as a source of bias in trials in sexual assault crimes. In this regard, a meta-analytical review (Ferguson and Malouff, 2016) found a high inter-study variability and that the results were subject to the effect of moderators that could not be identified due to lack of studies (possibly moderators of the effects are the definitions of false and unfounded complaint and the type of population). While waiting to know these moderators, the lower (0.012) and upper (0.174) limits reported demarcate the estimates within normality (between the interval of the best estimates), with the lower and upper estimates being outside the normal range.

In Latin-American judicial setting, it has been argued that investigative actions and judicial judgment making in sexual assault cases may be contaminated by prejudices about sexual

assault and the complainant (Salinas et al., 2015). Hence a quasi-experimental field study with Chilean judicial agents (i.e., law enforcement officers, correction officers, prosecutors, and judges), as professionals with procedural or judicial competences, to estimate the probability of false or unfounded complaints, the incidence of myths about sexual assaults, and the incidence and effects of motivational biases in judgment making was designed.

Materials and methods

Participants

Chilean judicial agents participated in the study, of which 60 were gendarmes (correctional officers), 76 police (law enforcement officers), 67 prosecutors and 14 judges. The distribution of the participants by age, sex, seniority in the position, and specialized training in sexual crimes can be seen in Table 1.

Procedure and design

A quasi-experimental field study was designed. Firstly, the Gendarmerie, Investigative Police, National Prosecutor's Office and Judicial Power headquarters were required to authorize the data collection among their members, presenting the investigation design and measures. Once approval was obtained, participants were contacted by their headquarters asking for voluntary participation. Those who agreed to participate were contacted personally by researchers, signed an informed consent, and endorsed the measures. In compliance with Chilean regulations, the ethical principles of beneficence, autonomy, and

justice were respected. Data collection was individual and, once the sociodemographic information was obtained, to counterbalance the interaction between measures, the order in which the measures were obtained was rotated (standard rotation procedure), that is, A, B, C,... F; B, C,.... A;... Data were collected individually from July 2018 to February 2019.

Measure instruments

A sociodemographic questionnaire was created in which the participants reported age, sex, length of service (< 10 and > 10 years, which is the criterion with which judicial agents are considered to be highly experienced public officials) and having completed specialized training in sexual crimes in his/her position (yes vs. no). Additionally, they were asked to estimate the percentage (from 0 to 100), i.e., probability of false/unfounded accusations regarding sexual crimes according to their own experience.

For the measurement of myths about sexual aggression, it was applied the Chilean adaptation (Camplá et al., 2019a,b) of the AMMSA Scale (Acceptance of Modern Myths about Sexual Aggression; Gerger et al., 2007). This adaptation, with a unidimensional structure, consists of 14 items to which people respond on a 7-point Likert-type scale from *completely disagree* (1) to *completely agree* (7). With study participants, the scale showed excellent internal consistency, $\alpha = 0.907$.

For the evaluation of attributional biases in the study population, 2 vignettes ($\pm$ 300 words) with weak evidence (only the accusatory testimony of the complainant) about a rape allegation were developed. The difference between the two scenarios was that the accused was or was not known to the complainant. Previously, a group of 10 researchers with experience in research design and with knowledge of psychology and law evaluated the incriminating evidence in the scenarios on an 11-point scale (Thurstone' procedure) if the charging evidence was weak (1, *extremely weak*) or strong (11, *extremely strong*). The results showed a $Mdn = 1$, Mode = 1, max. score = 3, Q1 = 1, Q2 = 2, and IQR = 1. Thus, the scenarios were assessed for the evaluators as weak evidence. Participants answered to a validated measure of the attributional biases (Arce et al., 2003), consisting in 5 questions:

(1) To what extent do you attribute responsibility to the complainant in the reported facts? Where: 0 = *Not at all responsible*; 1 = *Slightly responsible;* 2 = *Somewhat responsible*; 3 = *Mostly responsible;* 4 = *Completely responsible.*

(2) With what probability do you attribute the complainant's ability to have prevented the reported incident? Where: 0 = *Not probable*; 1 = *Slightly probable;* 2 = *Somewhat probable*; 3 = *Moderately probable*; 4 = *Extremely probable.*

(3) To what extent do you attribute responsibility to the defendant in the reported facts? Where: 0 = *Not at all responsible*;

TABLE 1 Socio-demographic characteristics of the participants.

Variable		Gendarmes	Polices	Prosecutors	Judges
Age		*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
		35.7 (6.2)	41.7 (3.7)	39.2 (6.1)	51.9 (8.5)
		n (%)	*n* (%)	*n* (%)	*n* (%)
Sex[a]	Men	46(78)	70 (92.1)	28 (41.8)	4 (28.6)
	Women	13 (22)	6 (7.9)	39 (58.2)	10 (71.4)
Training in sexual crimes		2 (3.4)	12 (15.8)	35 (52.2)	11 (78.6)
Seniority in the position	<10 years	8 (13.3)	0 (0)	35 (52.2)	0 (0)
	>10 years	52 (86.7)	76 (100)	32 (47.8)	14 (100)
Total		60	76	67	14

M(SD), Mean(Standard Deviation); n(%), Number of participants(observed percentage).
[a]1missing value.

1 = *Slightly responsible*; 2 = *Somewhat responsible*; 3 = *Mostly responsible*; 4 = *Completely responsible*.

(4) To what extent do you attribute credibility to the incriminating testimony given by the complainant? Where: 0 = *Not true*; 1 = *Slightly true*; 2 = *Somewhat true*; 3 = *Moderately true*; 4 = *Completely true*.

(5) With what probability do you attribute nature of a rape to the alleged facts? Where: 0 = *Not probable*; 1 = *Slightly probable*; 2 = *Somewhat probable*; 3 = *Moderately probable*; 4 = *Extremely probable*.

These variables are measuring the same construct ($\alpha = 0.70$; $\bar{r} = 0.244$; $r_s > 0.142$, $p < 0.05$), the attributional bias in judgment making.

Data analysis

Mean comparisons with a test value was computed with one sample t-test, being effect size estimated with Cohens's d and quantifying the magnitude in terms of r (Redondo et al., 2019). Mean comparisons of repeated measures were processed performing a MANOVA estimating multivariate effect size in percentage of explained variance (η^2) and bivariate effect size in Cohen's d (within formula).

Observed contingencies were contrasted with a constant computing Z score for the difference between proportions. The constants were taken as follows (Fandiño et al., 2021): (a) a trivial probability (≤ 0.05, insignificant probability); (b) a common probability (= 0.5, probable, observed in 50% of the population); and (c) a normal probability (≥ 0.90; normal, observed in 90% or more of the population). The magnitude of the increase or decrease of the observed contingency (effect size) was valued in terms of the Effect Incremental Index (EII; Arias et al., 2020).

As to compare the observed probabilities of false or unfounded allegations into the three classification categories (lower than best estimates, within best estimates, higher than best estimates) and between subsamples, the 95% confidence interval for each observed probability was computed. If the confidence intervals overlap, then the observed probabilities are equal, meanwhile if the confidence interval do not overlap, the observed probabilities are significantly different. Equally, the 95% confidence interval of the observed mean in the attributional biases were calculated to compare means between the sources of biases. Likewise, if the confidence intervals overlap, then the observed means are equal, meanwhile if the confidence interval do not overlap, the observed means are significantly different. The lower limit of the 90% confidence interval [i.e., $M – (1.645 * SD)$] for the population distribution in the attributional biases was valued to know if a trivial (insignificant) effect (1 = slightly) was or not within the normal distribution.

The population distribution and mean comparison study is of great scientific relevance, but results are insufficient to whole generalization to population as the effect is not general. Thus, the estimation of the margin of error of the resulting statistical model complements the significant model and should be reported. Hence, the Probability of an Inferiority (PIS) or a Superiority (PSS) Score (Gancedo et al., 2021a), i.e., the probability of subjects of the higher mean score group obtains a lesser score than the mean of the higher group (PIS), or the probability of subjects of the lower mean score group obtains a great score than the mean of the higher group (PSS), was estimated.

Results

Estimation of false or unsubstantiated reports in reports of sexual assault

The mean probability ($M = 0.200$) reported by the study population (22 participants did not answer this question, $N = 196$) of false/unfounded complaints is significantly higher, $t(195) = 10.12$, $p < 0.001$, than the mean ($M = 0.052$) of the existing studies (test value from a meta-analytic review; Ferguson and Malouff, 2016) in the literature, with a large effect size, $d = 0.98$, implying a 44.0% ($r = 0.440$) of increase in the estimate over the average of the best estimate. However, given the high variability in the estimates (heterogeneity in the studies) and that the moderating variables of the effect are unknown, although it is believed that the measure of only false complaints gives rise to lower rates than when it also includes unfounded complaints, the observed mean was contrasted with the upper limit of the estimates (0.174), finding that the mean of the study population was equal to the upper limit, $t(195) = 1.76$, *ns*. Hence, the estimates of false/unsubstantiated allegations are at the upper limit of the best estimates.

Regarding the case study (see **Table 1**), the reported probability was recoded in three categories according to the results of the meta-analytic review by Ferguson and Malouff (2016): Lower than best estimates (estimated probability ≤ 0.012); within the best estimates ($0.012 <$ estimated probability ≤ 0.174) and higher than best estimates (estimated probability > 0.174). The results showed, for the population of judicial agents, a non-trivial underestimation (> 0.05) of the probability of false/unfounded complaints, $Z(N = 196) = 7.25$, $p < 0.001$, resulting the increase over a trivial probability of 69.3% (EII = 0.693), being common (= 0.5) the overestimation (0.454), $Z(N = 196) = 1.29$, *ns*. On the other hand, the observed probability within the best estimates (0.383) is significantly lower than that expected for this contingency (0.90, normal probability), $Z(N = 196) = –24.13$, $p < 0.001$, with the decrease in 57.4% (EII = 0.574) in relation to the normal probability. Comparatively, the

probability of overestimation is significantly higher than that of underestimation, $\chi^2(N = 121) = 26.85$, $p < 0.001$. The sample was divided in agents with incriminating procedural functions (e.g., investigation, apprehension, detention of individuals suspected of criminal offenses, prosecution of the defendant or dismiss the case), law enforcement officers (polices) and prosecutors, and agents without procedural functions (they do not make decisions about the process), gendarmes and judges, the results showed (see Table 2) the same contingencies (confidence intervals for the observed proportions overlap) in both subsamples in lower than best estimates and higher than best estimates.

Study of the incidence of myths about sexual assault

The results (see Table 3) exhibited in the studied population (judicial agents) a systematic trend of disagreement with the myths about sexual assaults, except for the accusation of sexual violence to obtain custody, labeling harmless conduct as sexual harassment in the battle between the sexes; and the interpretation of harmless behaviors as sexual harassment at work, in which the degree of agreement is not positioned. As for the myths overall score (a composite score was computed), the results exhibited that judicial agents ($M = 3.33$) do not share the myths about sexual assault, $t(217) = -8.24$, $p < 0.001$, $d = 0.56$.

In contrast, the population reported participating in the attributional bias of greater provision of psychological support from the community to rape victims than to victims of very violent crimes (robbery with the use of weapons).

However, the case study (see PSS in Table 3), necessary in this type of population because the decisions are individual and the biases derived from the myths are manifested individually, warns of bias rates (in agreement with the myths) in this population they range from approximately 1/4 ($\pm$ 25%) for myths (see the content of the myths in Table 3) 4, 7, 8, 9, and 11; around 1/3 ($\pm$ 33%) for myths 5, 13, and 14; around 1/2.5 ($\pm$ 40%) for myths 3 and 6; approximately 1/2 ($\pm$ 50%) for myths 1, 2, and 10; and more than 1/2 (+50%) in myth 12. In general (composite score), about 1/4 of judicial agents agree (PIS = 0.291) with myths about sexual assault.

Evaluation of the effect of attributional biases in judgment making

The results of the study of the extra-legal evidence (see Table 4) revealed a significant effect (>1, 1 = trivial effect) of the biases of attribution of responsibility to the complainant and the accused, of truthfulness to the complainant, of a nature of rape to the alleged facts, and of a preventive role for the complainant with effect sizes greater than large ($d > 1.20$).

The comparison of the bias attributing responsibility to the complainant and the accused (see Table 5) showed a significantly higher attribution of responsibility or the accused (confidence intervals do not overlap, and the mean is higher for the accused). On the other hand, the bias of attribution to the complainant of the ability to prevent the incident is the one with lower incidence (the upper limit of confidence for the mean is lower than the upper limit of the other biases), while the attribution to the facts of a nature of rape, the one with higher incidence (the lower limit of the confidence interval for the mean is greater than the upper limit of the remaining biases). In an intermediate position is the attribution of veracity to the testimony of the complainant.

The normal interval includes the triviality (1) in the attribution of responsibility to the complainant, but not to the accused; that is, the non-attribution of responsibility to the accused is out of normality (abnormal), meanwhile triviality in the attribution of responsibility to the complainant is normal (falls into normality). It also falls within normality (90% normal interval lower limit < 1) not attributing sufficient veracity to the complainant's testimony, as well as not attributing to the complainant the ability to prevent the incident, while not attributing a nature of rape to the facts of rape is abnormal (90% normal interval lower limit > 1). Succinctly, not attributing responsibility to the accused or not qualifying the facts described as a rape is abnormal in this population.

Performed a repeated measures MANOVA on the extralegal evidence measurement variables i.e., attributional biases, the results showed a significant multivariate effect, [$F(5, 211) = 6.47$, $p < 0.001$, $1-\beta = 0.997$], for the perpetrator factor (between factor: known vs. unknown), which explains 13.3% of the variance. As for univariate effects (see Table 6), the results exhibited a significant higher attribution of responsivity, of

TABLE 2 Contingency table of grouping estimates and population.

Estimation false/Unfounded allegations	Gendarmes/Judges f(p[95% CI])	Polices/Prosecutors f(p[95% CI])	Total f(p[95% CI])
Lower than best estimates (<0.012)	13 (0.169 [0.085, 0.253])	19(0.160 [0.094, 0.226])	32 (0.163 [0.111, 0.215])
Within best estimates (0.012, 0.174)	29 (0.377 [0.269, 0.485)	46 (0.383 [0.296, 0.470])	75 (0.383 [0.315, 0.451])
Higher than best estimates (>0.174)	35 (0.455 [0.344, 0.566])	54 (0.454 [0.365, 0.543])	89 (0.454 [0.384, 0.524])
Total	77	119	196

f(p[95% CI]), frequency observed probability [95% Confidence Interval].

veracity in the testimony and of capacity to prevent the incident to complainant for unknown perpetrators in comparison with known perpetrators. However, 34.0, 30.2, and 28.8% of the judicial agents (see PIS in **Table 6**) would attribute less responsibility, truthfulness in the testimony and prevention capacity to the complainant when the perpetrator is unknown (statistical model error).

Discussion

Regarding the incidence of myths about sexual assault, the results obtained with Chilean judicial agents showed, in general, a systematic tendency of disagreement with the myths about

sexual assault. Bearing in mind the theoretical content categories of the scale (Gerger et al., 2007), they express disagreement with the denial of the scope of sexual violence, as well as with gender stereotypes about male sexuality, the beliefs that exonerate perpetrators of violence, and the naturalization of male coercion. At the same time, they are in favor of the demands of the victims and of the policies designed to address the effects of sexual violence. Nevertheless, there is no systematic trend of agreement or disagreement regarding the false accusation of sexual violence to obtain custody, and the interpretation of harmless gestures as sexual harassment. On the contrary, they participate in the myth of positive discrimination "receive more psychological support" rape victims compared to victims of armed robbery that are not based on the provision of greater support to victims of

TABLE 3 One sample *t*-test for the contrast of the acceptance of the myths about sexual aggression (test value: 4, Neither agree nor disagree).

Myths about sexual assault	*M*	*t*	*d*	PSS
1. Para conseguir la custodia de sus hijos, las mujeres a menudo acusan falsamente a sus exmaridos (o exparejas) de tener inclinaciones hacia la violencia sexual [To get custody for their children, women often falsely accuse their ex-husband of a tendency toward sexual violence]	3.82	−1.57	−0.11	0.456
2. Interpretar gestos inofensivos como "acoso sexual" es un arma muy común en la batalla de los sexos [Interpreting harmless gestures as "sexual harassment" is a popular weapon in the battle of the sexes]	3.82	−1.59	−0.07	0.472
3. Mientras no vayan demasiado lejos, los comentarios e insinuaciones que se hacen a las mujeres simplemente quieren decirle que es atractiva. [As long as they don't go too far, suggestive remarks and allusions simply tell a woman that she is attractive]	3.55	−3.73***	−0.25	0.401
4. La mayoría de las mujeres prefieren ser elogiadas por su físico que por su inteligencia [Most women prefer to be praised for their looks rather than their intelligence]	2.97	−8.28***	−0.55	0.291
5. Aunque a las mujeres les gusta hacerse las tímidas, eso no significa que no quieran sexo. Women like to play coy. This does not mean that they do not want sex.	3.01	−8.08***	−0.47	0.319
6. Muchas mujeres tienden a exagerar el problema de la violencia machista [Many women tend to exaggerate the problem of male violence]	3.59	−3.10**	−0.23	0.409
7. Cuando una mujer soltera invita a un hombre soltero a su casa, está indicando que no es reacia a mantener relaciones sexuales [When a single woman invites a single man to her flat she signals that she is not averse to having sex]	2.65	−12.56***	−0.73	0.233
8. Cuando se habla de "violación en el matrimonio," no hay una distinción clara entre coito conyugal normal y violación [When defining "marital rape," there is no clear-cut distinction between normal conjugal intercourse and rape]	2.74	−10.04***	−0.60	0.274
9. La sexualidad de un hombre funciona como una olla a presión; cuando la presión es muy alta. tiene que "soltar vapor"[A man's sexuality functions like a steam boiler – when the pressure gets too high, he has to "let off steam"]	2.65	−11.11***	−0.66	0.255
10. El debate sobre el acoso sexual en el trabajo ha provocado que muchos comportamientos inofensivos sean malinterpretados como acoso sexual [The discussion about sexual harassment on the job has mainly resulted in many a harmless behavior being misinterpreted as harassment]	3.80	−1.66	−0.10	0.460
11. En las citas lo que suele esperarse es que la mujer "ponga el freno" y el hombre "siga adelante" [In dating situations, the general expectation is that the woman "hits the brakes" and the man "pushes ahead"]	2.85	−10.27***	−0.59	0.278
12. Pese a que las víctimas de robo armado corren un mayor peligro de vida, reciben mucho menos apoyo psicológico que las víctimas de violación [Although the victims of armed robbery have to fear for their lives, they receive far less psychological support than do rape victims]	4.29	2.23*	0.16	0.564
13. El alcohol es a menudo el causante de que un hombre viole a una mujer [Alcohol is often the culprit when a man rapes a woman]	3.31	−5.33***	−0.34	0.367
14. Muchas mujeres tienden a malinterpretar un gesto bienintencionado como "acoso sexual" [Many women tend to misinterpret a well-meant gesture as a "sexual assault"]	3.54	−3.90***	−0.26	0.397

df(216); *p < 0.05; **p < 0.01; ***p < 0.001.

TABLE 4   One sample *t*-test of the attributional bias measures with a trivial attribution as test value (1, slightly).

Source of attributional bias	t	M	d
Attribution of responsibility to the complainant	34.22***	3.08	3.28
Attribution of responsibility to the accused	48.38***	3.43	4.65
Attribution of veracity to the complainant	32.48***	2.89	3.15
Attribution of nature of a rape to the facts?	57.40***	3.57	5.52
Attribution of prevention of the incident to the complainant	15.66***	2.14	1.51

df(433). ***$p < 0.001$.

TABLE 5   Confidence interval for the observed mean and populational lower limit of the normal distribution of the extralegal evidence variables (attributional biases).

Variable	M [95% CI]	90% NI lower limit
Attribution of responsibility to the complainant	3.08 [2.96, 3.20]	0.99
Attribution of responsibility to the accused	3.43 [3.33, 3.53]	1.70
Attribution of veracity to the complainant	2.89 [2.80, 2.98]	0.90
Attribution of nature of a rape to the facts	3.57 [3.49, 3.65]	2.04
Attribution of prevention ability to the complainant	2.14 [2.00, 2.28]	−0.36

N = 434; M [95% CI], Mean [95% Confidence Interval for the mean]; 90% NI lower limit, 90% normal interval lower limit.

sexual assault. De facto, all Chilean victims of violent crimes (e.g., sexual assault, kidnapping, robbery with violence) receive legal assistance, psychological therapy and social support. In any case, the results do not endorse a generalized cognitive bias in this population contrary to those reporting sexual assault victimization, observed in other studies (Sleath and Bull, 2012; Smith and Skinner, 2012; McMillan, 2016; Hine and Murphy, 2017; Temkin et al., 2018).

However, the case study, necessary, as it deals with personal biases (De Neys and Bonnefon, 2013), indicates that a large part of the legal agents participate in the myths about sexual assaults, with agreement rates that oscillate between ± 25%, at +50%. In sum, although a systematic bias trend is not observed in the population studied, a significant and large prevalence of cases has been recorded (ranging from 1/4 to more than 50% of judicial agents, depending on the myths).

Myths are part of sources of informal reasoning (evidence is replaced by the myths), as opposed to formal reasoning (evidence based) that must support judicial judgments (Kruglanski and Azjen, 1983). In practice, myths are within the cognitive bias "preconceived ideas or theories" that predispose the individual to adopt uncertain ideas (e.g., myths) that guide judgment making *via* information processing strategies such as presumed covariation (e.g., correlation between myths and false or unfounded allegations), representativeness (e.g., overestimation of the probabilities of false or unfounded allegations related to myths) or causality (e.g., myths are the causes of false or unfounded allegations) (Fariña et al., 2002). These preconceptions maximize judgments based on myths, avoid discordant information, and lead to cognitive savings for the individual (Ross, 1977).

In relation to the estimation of false or unfounded reports (representativeness cognitive bias), it was found that the mean probability reported by the study population is significantly higher than the mean reported in the literature (0.052; Ferguson and Malouff, 2016), and is in the upper limit of the best estimates. In addition, the study of cases, according to the results of the meta-analytical review by Ferguson and Malouff (2016), has allowed in order to establish for legal agents a common overestimation of false or unfounded complaints, as well as a non-trivial underestimation, although the overestimation is significantly greater than the underestimation (Venema, 2016). In this sense, research has shown that high estimates are related to a lower allocation of credibility and receptivity toward those who report (Lonsway et al., 2009; Mennicke et al., 2014). Likewise, judicial agents with differentiated procedural functions (who activate the search for evidence or file the process), police and prosecutors, and operators without procedural functions (they do not make decisions about the process), gendarmes and judges, did not reveal differences in the estimation of false or unfounded complaints, which may condition the decision about a criminal prosecution or contribute to poor investigations (Mennicke et al., 2014; Hohl and Stanko, 2015; O'Neal et al., 2015; Carboné-López et al., 2016). In any case, it is remarkable that the assessment of the probability of false or unfounded complaints is conditioned in this population by the judicial criterion of subjective incredibility (Law of Precedent) that establishes that the testimony of the complainant is not sufficient proof of incrimination if he/she has any interest in the cause beyond the legitimate conviction of the accused (e.g., economic benefit, resentment, revenge, existence of a previous relationship).

Regarding attributional biases, the results confirmed that, for judgment making and, by extension, judicial decision-making, judicial agents use sources of informal reasoning (attributional biases) in contrast to the formal reasoning expected in this context. These biases facilitate paths of judgment both incriminating (i.e., attribution of responsibility to the accused, attribution of credibility to the complainant, attribution to the facts of a nature of rape) and exculpatory (i.e., attribution of responsibility to the complainant, attribution of the duty of prevention to the complainant). Furthermore, the magnitude of the effects of the biases in the reasoning is more than great. However, the biases in favor of incriminating the

TABLE 6 Univariate effects on the attributional biases for the perpetrator factor.

Variable	M_{Uk}	M_K	F	1-β	d	PIS
Attribution of responsibility to the complainant	3.16	3.00	4.39*	0.550	0.28	0.340
Attribution of responsibility to the accused	3.50	3.36	2.86	0.391	0.23	0.409
Attribution of veracity to the complainant	3.06	2.72	14.33***	0.965	0.52	0.302
Knowledge of the crime as such	3.62	3.52	2.95	0.401	0.24	0.405
Attribution of prevention ability to the complainant	2.35	1.94	16.84***	0.983	0.56	0.288

Within-subjects effects.

df (1, 215); M_{Uk}, mean of the unknown perpetrator condition; M_K, mean of the known perpetrator condition.

*$p < 0.05$; ***$p < 0.001$.

accused (attribution of responsibility to the accused, attribution of credibility to the complainant, attribution to the facts a nature of rape) have a greater weight than those that delegitimize the accusation (attribution of responsibility to the complainant, attribution of the duty of prevention to the complainant). Moreover, the results revealed that these have a smaller effect on known victims than on unknown ones. Succinctly, higher effects for attributional biases are for unknown perpetrators in both directions: to support incrimination (i.e., attribution of higher veracity to complainant testimony) and absolution/dismiss of the judicial proceeding (i.e., higher attribution of responsivity to complainant, higher attribution of capacity to prevent the incident to complainant). Nevertheless, the attribution of credibility to the complainant testimony has a higher incriminating evidence value than the attribution of responsivity and capacity to prevent the incident to the complainant for exonerating criminal responsibility (Arce et al., 2003). In this way, the existence of a close link between the complainant and the accused requires a greater burden of proof, since less verisimilitude is attributed to the complainant testimony (McKimmie et al., 2014), more responsibility is attributed to the complainant in the facts and more ability to have prevented the incident (Hohl and Stanko, 2015; Hine and Murphy, 2017). In the absence of such an increased burden of proof, these attributional biases would predispose judgment making toward the absolution of the prosecuted or the dismiss of the judicial proceeding. Paradoxically, this is related to a lower probability of reporting and abandoning the relationship (Garrido-Macías et al., 2020) and, by extension, less judicial protection for victims of known aggressors. However, the case study warns that around 1/3 of the judicial agents (34.0, 30.2, and 28.8%, respectively), would attribute, respectively, more responsibility, truthfulness, and prevention capacity to the complainant when the perpetrator is known (statistical model error).

Motivational attribution biases refer to a tendency to form and hold beliefs that conform to the needs of the individual, in this case, judgment making and the subsequent decision-making, and manifest when the legal evidence is insufficient or weak (week cases; Butterfield and Bitter, 2019), being irrelevant in strong evidence cases (Visher, 1987; Kahneman, 2011). Under this contingency, judicial agents, in judicial

judgment and decision making, must be guided by strict compliance with the principle of presumption of innocence (Article 11.1 of the Universal Declaration of Human Rights; United Nations, 1948), which implies that none innocent person may be classified as guilty, so the attributional biases supporting the absolution of the accused of the crime would support the motivation of the procedural action taken or the judicial resolution. However, although the judicial resolution or procedural action (i.e., prosecutorial decision making) executed would be correct, it would be based on informal reasoning, which would deviate from normative judicial reasoning (i.e., evidenceless, unfounded). On the contrary, in cases of weak or insufficient legal evidence (burden of proof), resorting to attributional biases to support the case would not only be inadmissible in terms of reasoning (motivation of the actions or judicial resolutions) as it is informal in the face of the expected formal, but also contrary to law (judicial error). In any case, attributional biases, such as irrational beliefs, provide such a high level of support (effect size greater than large) that they give the subject a guarantee of certainty and efficiency in judgment making (Perry, 1988).

In conclusion, although generally and in population terms the effect of biases in judgment making contrary to complainant is not observed, in the case study it was found that a large number of judicial agents participated of biased routes against the complainant. Moreover, this type of bias in judgment making does not occur in other crimes (Bieneck and Krahé, 2011). This research found that in rape cases was attributed more blame to the victim and less blame to the perpetrator compared with robbery cases. Thus, as these sources of bias in judgment making are unconscious for judgment makers and ways of informal reasoning, judicial agents should be trained to control the effects of these sources of bias (Bartels, 2010), promoting debiasing, i.e., substituting informal reasoning (judgment making sustained on biases against the complainant) by formal reasoning sources (evidence, procedural rules, charge of the proof) (Butterfield and Bitter, 2019). In sum, the training and specialization of judicial agents (e.g., courts specialized in sexual assaults, training police forces to obtain the statement from complainants of sexual assault) in sexual violence against women cases is necessary

(Barn and Powers, 2021; Gancedo et al., 2021b); so that, from an orientation of Therapeutic Justice, they can mediate the wellbeing of the victims (Cattaneo and Goodman, 2010; Camplá et al., 2020; Novo et al., 2020).

Limitations

The results of this study are not generalizable to other types of populations, since judicial agents are determined in their judgments by procedural and legal rules and Chilean Law of Precedent. Likewise, caution must be taken in generalizing judicial agents from judicial contexts other than Chile, since the case law may not be equivalent. As participation was voluntary, the results do not represent strictly the population. Finally, the manifestation of these biases can only be generalized to cases of insufficient evidence.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Álvarez-Muelas, A., Gómez-Berrocal, C., and Sierra, J. C. (2020). Relación del doble estándar sexual con el funcionamiento sexual y las conductas sexuales de riesgo: revisión sistemática [Relationship of sexual double standard with sexual functioning and risk sexual behaviors: a systematic review]. *Rev. Iberoamericana Psicol. Salud* 11, 103–116. doi: 10.23923/j.rips.2020.02.038

Álvarez-Muelas, A., Gómez-Berrocal, C., and Sierra, J. C. (2021). Typologies of sexual double standard adherence in Spanish population. *Eur. J. Psychol. Appl. Legal Context* 13, 1–7. doi: 10.5093/ejpalc2021a1

Anders, M. C., and Christopher, F. S. (2011). A socioecological model of rape survivors' decisions to aid in case prosecution. *Psychol. Women Q.* 35, 92–106. doi: 10.1177/0361684310394802

Arce, R. (2017). Análisis de contenido de las declaraciones de testigos: evaluación de la validez científica y judicial de la hipótesis y la prueba forense [Content analysis of the witness statements: evaluation of the scientific and judicial validity of the hypothesis and the forensic proof]. *Acción Psicol.* 14, 171–190. doi: 10.5944/ap.14.2.21347

Arce, R., Fariña, F., and Fraga, A. (2000). Género y formación de juicios en un caso de violación [Gender and juror judgment making in a case of rape]. *Psicothema* 12, 623–628.

Arce, R., Tortosa, F., and Alfaro, E. (2003). Veredictos y análisis del contenido de las deliberaciones de los tribunales de jueces y jurados en el contexto jurídico español [Verdict and deliberation content analysis of judges and juries in the Spanish legal context: a contrastive approach]. *Psicothema* 15, 127–135.

Arias, E., Arce, R., Vázquez, M. J., and Marcos, V. (2020). Treatment efficacy on the cognitive competence of convicted intimate partner violence offenders. *Anal. Psicol. Ann. Psychol.* 36, 427–435. doi: 10.6018/analesps.428771

Barn, R., and Powers, R. A. (2021). Rape myth acceptance in contemporary times: a comparative study of university students in India and the United Kingdom. *J. Interpers. Violence* 36, 3514–3535. doi: 10.1177/0886260518775750

Bartels, B. (2010). "Top-down and bottom-up models of judicial reasoning," in *The Psychology of Judicial Decision Making*, eds D. Klein and G. Mitchell (New York, NY: Oxford University Press), 41–56.

Bieneck, S., and Krahé, B. (2011). Blaming the victim and exonerating the perpetrator in cases of rape and robbery: is there a double standard? *J. Interpers. Violence* 26, 1785–1797. doi: 10.1177/0886260510372945

Burger, J. (1981). Motivational biases in the attribution of responsibility for an accident: a meta-analysis of the defensive-attribution hypothesis. *Psychol. Bull.* 90, 496–512. doi: 10.1037/0033-2909.90.3.496

Butterfield, M. E., and Bitter, A. N. (2019). A social judgment? Extralegal contrast effects in hypothetical legal decision making. *Psychol. Public Policy Law* 25, 30–37. doi: 10.1037/law0000188

Camplá, X., Marcos, V., Fariña, F., and Arce, R. (2019a). "Sexual violence against women: variables involved in judicial decision making," in *2019 Annual Conference of the European Association of Psychology and Law. Abstract book*, eds J. Sanmarco and E. Arias (Santiago de Compostela: European Association of Psychology and Law), 105–106. . doi: 10.1177/0886260520984424

Camplá, X., Novo, M., and Arce, R. (2017). "Fuerzas policiales y mitos sobre las agresiones sexuales: una revisión sistemática de la evidencia," in *Psicología Jurídica: Conocimiento y Práctica. Colección Psicología y Ley N° 14*, eds C. Bringas and M. Novo (Seville: Universidad de Sevilla), 287–303.

Camplá, X., Novo, M., and Vázquez, M. J. (2020). "La víctima adulta de violencia sexual: procesos de victimización y perspectiva de Justicia Terapéutica," in *Justicia Terapéutica en Iberoamérica*, eds F. Fariña, M. S. Oyhamburu, and D. Wexler (Barcelona: Wolters Kluwer), 189–208.

Camplá, X., Novo, M., Sanmarco, J., and Arce, R. (2019b). Escala acceptance of modern myths about sexual aggression: validación para la población chilena [the acceptance of modern myths about sexual aggression scale: validation for chilean population]. *Rev. Mexicana Psicol.* 36, 132–146.

Carboné-López, K., Slocum, L., and Kruttschnitt, C. (2016). "Police wouldn't give you no help": female offenders on reporting sexual assault to police. *Violence Against Women* 22, 366–396. doi: 10.1177/1077801215602345

Cattaneo, L. B., and Goodman, L. A. (2010). Through the lens of therapeutic jurisprudence: the relationship between empowerment in the court system and well-being for intimate partner violence victims. *J. Interpers. Violence* 25, 481–502. doi: 10.1177/0886260509334282

De Neys, W., and Bonnefon, J. (2013). The "whys" and "whens" of individual differences in thinking biases. *Trends Cogn. Sci.* 17, 172–178. doi: 10.1016/j.tics.2013.02.001

Du Mont, J., Miller, K., and Myhr, T. (2003). The role of "real rape" and "real victim" stereotypes in the police reporting practices of sexually assaulted women. *Violence Against Women* 9, 466–486. doi: 10.1177/1077801202250960

Fandiño, R., Basanta, J., Sanmarco, J., Arce, R., and Fariña, F. (2021). Evaluation of the executive functioning and psychological adjustment of child to parent offenders: epidemiology and quantification of harm. *Front. Psychol.* 12:616855. doi: 10.3389/fpsyg.2021.616855

Fariña, F., Arce, R., and Novo, M. (2002). Heurístico de anclaje en las decisiones judiciales [Anchorage in judicial decision making]. *Psicothema* 14, 39–46.

Ferguson, C. E., and Malouff, J. M. (2016). Assessing police classifications of sexual assault reports: a meta-analysis of false reporting rates. *Arch. Sex. Behav.* 45, 1185–1193. doi: 10.1007/s10508-015-0666-2

Fitzmaurice, C., and Pease, K. (1986). *The Psychology of Judicial Sentencing*. Manchester: Manchester University Press.

Gancedo, Y., Fariña, F., Seijo, D., Vilariño, M., and Arce, R. (2021a). Reality monitoring: a meta-analytical review for forensic practice. *Eur. J. Psychol. Appl.Legal Context* 13, 99–110. doi: 10.5093/ejpalc2021a10

Gancedo, Y., Sanmarco, J., Selaya, A., González-Dapía, A., and Novo, M. (2021b). Estudio de los efectos de los mitos sobre las agresiones sexuales en la verosimilitud de una víctima-denunciante de una agresión sexual [Assessing the effects of myths about sexual ggression on the verisimilitude of a victim-complainant of sexual assault]. *Acción Psicol.* 18, 221–248. doi: 10.5944/ap.18.1.30805

Garrido-Macías, M., Valor-Segura, I., and Expósito, F. (2020). Which tactics of sexual violence predict leaving the relationship? The role of dependence towards partner. *Eur. J. Psychol. Appl. Legal Context* 12, 53–60. doi: 10.5093/ejpalc2020a6

Gerger, H., Kley, H., Bohner, G., and Siebler, F. (2007). The acceptance of modern myths about sexual aggression scale: development and validation in german and english. *Aggress. Behav.* 33, 422–440. doi: 10.1002/ab.20195

Hine, B., and Murphy, A. (2017). The impact of victim-perpetrator relationship, reputation and initial point of resistance on officers' responsibility and authenticity ratings towards hypothetical rape cases. *J. Crimin. Just.* 49, 1–13. doi: 10.1016/j.jcrimjus.2017.01.001

Hohl, K., and Stanko, E. (2015). Complaints of rape and the criminal justice system: fresh evidence on the attrition problem in England and Wales. *Eur. J. Criminol.* 12, 324–341. doi: 10.1177/1477370815571949

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York, NY: Macmillan.

Kaplan, M. F., Steindorf, J., and Iervolino, A. (1978). Courtrooms, politics, and morality: toward a theoretical integration. *Pers. Soc. Psychol. Bull.* 4, 155–160. doi: 10.1177/014616727800400134

Kruglanski, A. W., and Azjen, I. (1983). Bias and error in human judgment. *Eur. J. Soc. Psychol.* 13, 1–44. doi: 10.1002/ejsp.2420130102

Lisak, D., Gardinier, L., Nicksa, S. C., and Cote, A. M. (2010). False allegations of sexual assault: an analysis of ten years of reported cases. *Violence Against Women* 16, 1318–1334. doi: 10.1177/1077801210387747

Lonsway, K. (2010). Trying to move the elephant in the living room: responding to the challenge of false rape reports. *Violence Against Women* 16, 1356–1371. doi: 10.1177/1077801210387750

Lonsway, K. A., and Fitzgerald, L. F. (1994). Rape myths: in review. *Psychol. Women Q.* 18, 133–164. doi: 10.1111/j.1471-6402.1994.tb00448.x

Lonsway, K. A., Archambault, J., and Lisak, D. (2009). False reports: moving beyond the issue to successfully investigate and prosecute non-stranger sexual assault. *Voice* 3:1.

Martínez-Gómez, J. A., Bolívar-Suárez, Y., Rey-Anacona, C. A., Ramírez-Ortiz, L. C., Lizarazo-Ojeda, A. M., and Yanez-Perúñuñi, L. Y. (2021). Esquemas tradicionales de roles sexuales de género, poder en las relaciones y violencia en el noviazgo [Traditional schemes of sexual gender roles, power in relationships and dating violence]. *Rev. Iberoamericana Psicol. Salud* 12, 1–16. doi: 10.23923/j.rips.2021.01.041

McKimmie, B. M., Masser, B. M., and Bongiorno, R. (2014). What counts as rape? The effect of offense prototypes, victim stereotypes, and participant gender on how the complainant and defendant are perceived. *J. Interpers. Violence* 29, 2273–2303. doi: 10.1177/0886260513518843

McKimmie, B. M., Masser, B. M., Nitschke, F., Lee, H., and Schuller, R. (2020). The impact of schemas on decision-making in cases involving allegations of sexual violence. *Curr. Issues Crim. Justice* 32, 420–439. doi: 10.1080/10345329.2020.1829780

McMillan, L. (2016). Police officers' perceptions of false allegations of rape. *J. Gender Stud.* 27, 1–13. doi: 10.1080/09589236.2016.1194260

Mennicke, A., Anderson, D., Oehme, K., and Kennedy, S. (2014). Law enforcement officers' perception of rape and rape victims: A multimethod study. *Violence Victims* 29, 814–827. doi: 10.1891/0886-6708.VV-D-13-00017

Montibeller, G., and von Winterfeldt, D. (2015). Cognitive and motivational biases in decision and risk analysis. *Risk Anal.* 35, 1230–1251. doi: 10.1111/risa.12360

Nisbett, R. E., and Ross, L. (1980). *Human Inference: Strategies and Shortcomings of Social Judgment*. New York, NY: Prentice-Hall.

Nitschke, F. T., McKimmie, B. M., and Vanman, E. J. (2019). A meta-analysis of the emotional victim effect for female adult rape complainants: does complainant distress influence credibility? *Psychol. Bull.* 145, 953–979. doi: 10.1037/bul0000206

Novo, M., and Seijo, D. (2010). Judicial judgement-making and legal criteria of testimonial credibility. *Eur. J. Psychol. Appl. Legal Context* 2, 91–115.

Novo, M., Fariña, F., and Vázquez, M. J. (2020). "Justicia Terapéutica en procesos penales con personas menores de edad como víctimas y testigos," in *Justicia Terapéutica en Iberoamérica*, eds F. Fariña, M. S. Oyhamburu, and D. Wexler (Madrid: Wolters Kluwer), 219–240.

O'Neal, E., Tellis, K., and Spohn, C. (2015). Prosecuting intimate partner sexual assault: legal and extra-legal factors that influence charging decisions. *Violence Against Women* 21, 1237–1258. doi: 10.1177/1077801215591630

Page, A. D. (2007). Behind the blue line: investigating police officers' attitudes toward rape. *J. Police Crimin. Psychol.* 22, 22–32. doi: 10.1007/s11896-007-9002-7

Page, A. D. (2010). True colors: police officers and rape myth acceptance. *Feminist Criminol.* 5, 315–334. doi: 10.1177/1557085110384108

Perry, S. R. (1988). Judicial obligation, precedent and the common law. *Oxford J. Legal Stud.* 7, 215–257. doi: 10.1093/ojls/7.2.215

Redondo, L., Fariña, F., Seijo, D., Novo, M., and Arce, R. (2019). A meta-analytical review of the responses in the MMPI-2/MMPI-2-RF clinical and restructured scales of parents in child custody dispute. *Anal. Psicol. Ann. Psychol.* 35, 156–165. doi: 10.6018/analesps.35.1.338381

Ross, L. (1977). "The intuitive psychologist and his shortcomings: distortions in the attribution process," in *Advances in Experimental Social Psychology*, Vol. 10, ed. L. Berkowitz (New York, NY: Academic Press), 174–221.

Saks, M. J., and Kidd, R. F. (1986). "Human processing information: trial by heuristics," in *Judgment and Decision Making. A Interdisciplinary Reader*, eds H. R. Arkes and R. H. Hammond (Cambridge: Cambridge University Press), 213–242.

Salinas, A. M., Meneses, A., Ceric, F., and Aliaga, A. (2015). Decisiones judiciales en fallos de delitos sexuales. *Inter. Cienc. Forenses* 4, 35–52.

Schuller, R. A., McKimmie, B. M., Masser, B. M., and Klippenstine, M. A. (2010). Judgments of sexual assault: the impact of complainant emotional demeanor, gender, and victim stereotypes. *N. Crimin. Law Rev. Int. Interdiscip. J.* 13, 759–780. doi: 10.1525/nclr.2010.13.4.759

Sleath, E., and Bull, R. (2012). Comparing rape victim and perpetrator blaming in a police officer sample: differences between police officers with and without special training. *Crim. Justice Behav.* 39, 646–665. doi: 10.1177/0093854811434696

Smith, O., and Skinner, T. (2012). Observing court responses to victims of rape and sexual assault. *Feminist Criminol.* 7, 298–326. doi: 10.1177/1557085112437875

Smith, O., and Skinner, T. (2017). How rape myths are used and challenged in rape and sexual assault trials. *Soc. Legal Stud.* 26, 441–466. doi: 10.1177/0964663916680130

Temkin, J., Gray, J. M., and Barrett, J. (2018). Different functions of rape myth use in court: findings from a trial observation study. *Feminist Criminol.* 13, 205–226. doi: 10.1177/1557085116661627

United Nations (1948). *Universal declaration of human rights*. Available online at: https://www.undocs.org/en/A/RES/217(III) (accessed September 13, 2021).

Venema, R. M. (2016). Police officer schema of sexual assault reports: real rape, ambiguous cases, and false reports. *J. Interpers. Violence* 31, 872–899. doi: 10.1177/0886260514556765

Visher, C. A. (1987). Juror decision making: the importance of evidence. *Law Hum. Behav.* 11, 1–17. doi: 10.1007/BF01044835

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership