Do They Matter? A Meta-Analytic Investigation of Individual Characteristics and Guilt Judgments

Dennis J. Devine and David E. Caughlin
Indiana University-Purdue University Indianapolis

Many people believe the personal attributes of trial participants substantially impact the decisions of juries, and considerable research has examined the extent to which characteristics of jurors and defendants are associated with juror judgments of guilt. To assess this broad issue, we meta-analyzed empirical studies examining the relationship between 11 juror and defendant characteristics and individual-level judgments of guilt in criminal trial contexts. Three potential moderator variables were also investigated: participant type, outcome type, and case type. In total, 464 effects were obtained from 272 published and unpublished studies. The 11 focal characteristics yielded sample-weighted mean correlations ranging from zero to .22 in magnitude, with the strongest overall relationships emerging for defendant socioeconomic status (−.11), defendant criminal record (.12), juror authoritarianism (.17), and juror trust in the legal system (.22). There was, however, substantial evidence of moderation for 10 of the 11 characteristics, suggesting their overall relationships vary according to one or more other variables. Moderator analyses revealed little support for participant type, some support for outcome type, and good support for case type with regard to their ability to explain variation in the observed effects. Overall, several juror and defendant characteristics were associated strongly enough with guilt judgments to warrant the attention of scholars and legal practitioners, and the results of this work add to our understanding of extralegal bias and juror decision making.

Keywords: juror decision making, extralegal bias, meta-analysis, juries, criminal trials

A fundamental question underlying the use of juries is: To what extent do jurors base their decisions on the evidence presented at trial as opposed to “extralegal” influences that should ideally play no role? This question has long vexed legal scholars as well as the general public, and many potential sources of extralegal influence have been identified and studied by jury researchers since the 1950s. These include what jurors are exposed to before trial (e.g., pretrial publicity), courtroom practices and procedures used during jury trials (e.g., joinder, bifurcation), and what jurors encounter at trial (e.g., inadmissible evidence).

Another source of potential extralegal influence is the personal characteristics of trial participants. These characteristics include race, gender, socioeconomic status (SES), age, religious affiliation, marital status, sexual orientation, political ideology, personality traits, and various attitudes relevant to legal decision making—among others. Over time, a consensus has emerged that jury verdicts should be based on the evidence presented at trial and not the personal characteristics of the trial participants. Whereas jurors were once pointedly selected for their personal characteristics (e.g., community stature) or case-related knowledge, the legal ideal now is that these characteristics play no role at trial.

As a result of their salience in society and relative ease of measurement, there is a large and growing empirical literature on the relationship between the personal characteristics of key trial participants and juror decisions. In light (or sometimes in lieu) of these data, many scholars, reviewers and commentators have weighed in on this issue and generated a disparate set of conclusions (e.g., Abramson, 1994; Adler, 1994; Baldus, Woodworth, &
Pulaski, 1990; Devine, 2012; Kalven & Zeisel, 1966; Vidmar & Hans, 2007). Some conclude legally irrelevant participant characteristics play a decisive role in many trials; others believe these characteristics have little if any consistent influence. The question thus remains: To what extent are the decisions of jurors influenced by their personal characteristics and those of other trial participants? This study was conducted to advance our understanding of this question via a meta-analytic integration of the voluminous empirical literature on participant characteristics.

Theory and Research on Trial Participant Characteristics

A large set of participant characteristics could potentially affect juror decision making. Some of them are observable (i.e., perceived readily by others with a high degree of consistency) and others psychological (i.e., manifested over time through behavior and thus not instantly recognizable). Observable characteristics include physical features such as race, gender, age, and SES. Psychological characteristics include intelligence, personality traits, need for cognition, moral values, general trust in the legal system, and attitudes about particular aspects of it (e.g., various laws, the police). These many characteristics are associated with a variety of individuals who have different functional roles at trial, including: (1) jurors, (2) defendants, (3) victims, (4) witnesses, (5) judges, (6) attorneys, (7) court staff, and (8) spectators. Crossing the many characteristics with the various types of participants yields a large number of participant-characteristic combinations that could be studied, but the bulk of the empirical research has concentrated on the observable characteristics of jurors and defendants.

Early research on juror and defendant characteristics.

Scientific research on juror decision making arguably began with the groundbreaking work of an interdisciplinary group of researchers at the University of Chicago in the latter half of the 1950s. One of many notable contributions of this watershed project was its initial foray into the investigation of participant characteristics, namely juror gender and SES. The first published study of a psychological characteristic deemed relevant to legal decision making—juror authoritarianism—appeared in the late 1960s (Boehm, 1968). This was followed by the first published work on several additional trial participant characteristics in the early 1970s, including defendant SES (Adler, 1973), defendant physical attractiveness (Efran, 1974; Izzett & Leginski, 1974), and defendant criminal record (Hans & Doob, 1976; Sealy & Cornish, 1973). Anticipating later efforts, one of the first studies to examine a broad array of individual-difference variables was conducted in England toward the end of the decade (Baldwin & McConville, 1979, 1980).

Research on participant characteristics then exploded in the 1980s. Based on posttrial questionnaires, field studies of the relationship between a variety of participant characteristics and the reported verdict beliefs of jurors in actual trials began appearing (e.g., Mills & Bohannon, 1980; Moran & Comfort, 1986; Sannito & Arnold, 1982). The experimental study of two juror characteristics—trust in the legal system and need for cognition—was spurred by the development and publication of brief measures such as the Juror Bias Scale (Kassin & Wrightsman, 1983) and Need for Cognition Scale (Petty & Cacioppo, 1986). Another major development during this period was the extensive analysis of participant race in capital trials. Numerous researchers used archival data on death-penalty trials from various jurisdictions to examine the relationship between jury sentence and the race of defendants and/or victims (e.g., Baldus, Pulaski, & Woodworth, 1983; Bowers & Pierce, 1980; Foley & Powell, 1982; Gross & Mauro, 1984; Keil & Vito, 1989; Radelet, 1981). These studies generally revealed Black defendants and defendants accused of killing a White victim to have a greater likelihood of being convicted and sentenced to death—even when controlling for legally relevant aggravating and mitigating factors (e.g., Baldus, Woodworth, Zuckerman, Weiner, & Broffitt, 1998).

Meta-analyses. By the early 1990s, sufficient empirical data were available to allow for meta-analyses of several participant characteristics. The first published meta-analysis of a trial participant characteristic was conducted by Sweeney and Haney (1992), who examined the magnitude of racial bias exhibited by White jurors against Black defendants in sentencing decisions using 19 distinguishable effects from 14 studies. Their overall analysis revealed a modest bias against Black defendants in the sentencing decisions of White mock jurors (d = .17, equivalent to r = .08). However, when the analysis was limited to only those studies involving an all-White sample of participants and cases with White victims, the observed effect was somewhat stronger (d = .26, r = .13). Broadening the scope of inquiry, Mazzella and Feingold (1994) investigated racial bias against defendants regardless of the race of the juror, and included studies involving judgments of sentence as well as guilt. They obtained 63 effects from 29 experimental studies involving the manipulation of defendant race. In contrast to the earlier meta-analysis, Mazzella and Feingold’s overall analysis produced little indication of bias for guilt judgments (d = .01, r = .01) and only a very slight bias for sentencing decisions (d = .06, r = .03).

More than a decade later, Mitchell, Haw, Pfeifer, and Meissner (2005) returned to the domain of participant race to examine the degree of outgroup bias evident in juror decisions. Like Mazzella and Feingold (1994), they analyzed data from 29 experimental studies wherein defendant race was manipulated and judgments of either guilt or sentence were obtained. In contrast to the two earlier investigations, Mitchell and her colleagues examined only those studies that measured juror race and reported separate estimates of outgroup bias for both White and Black mock jurors. Their overall analysis based on all 46 effects yielded relatively small estimates of outgroup bias for both guilt judgments (d = .09, r = .04) and sentencing decisions (d = .19, r = .09), but when the effects were analyzed according to the race of study participants, differences emerged. White mock jurors displayed little evidence of outgroup bias (d = .03, r = .02 for guilt; d = .10, r = .05 for sentencing), whereas Black mock jurors exhibited a moderately strong outgroup bias (d = .43, r = .21 for guilt; d = .73, r = .34 for sentencing). Thus, results have differed somewhat across meta-analyses, with more recent work suggesting racial bias inherently involves an interaction between defendant race and juror race.

In addition to race, several other participant characteristics have now been meta-analyzed. Narby, Cutler, and Moran (1993) examined studies that reported associations between some measure of juror authoritarianism and judgments of defendant culpability in a criminal context. They obtained 20 studies containing 32 separate effects and estimated the mean correlation between authoritarian-
ism and guilt judgments to be modest in magnitude ($\bar{r} = .16$). However, when the effects were reanalyzed based on how broadly the construct was conceptualized, measures of legal authoritarianism were more strongly associated with culpability ($\bar{r} = .19$) than measures of traditional authoritarianism ($\bar{r} = .11$). In their 1994 meta-analytic investigation, Mazzella and Feingold also examined experimental studies of three defendant characteristics aside from race and found small overall effects for defendant physical attractiveness ($d = .19/\bar{r} = .09$ for guilt; $d = .12/\bar{r} = .06$ for sentencing), defendant SES ($d = .15/\bar{r} = .07$ for guilt; $d = .15/\bar{r} = .07$ for sentencing), and defendant gender ($d = -.08/\bar{r} = -.04$ for guilt; $d = .17/\bar{r} = .08$ for sentencing). In essence, criminal defendants tended to be viewed as less guilty and punished less severely when they were more physically attractive, female, and of higher SES. Most recently, Schutte and Hosch (1997) examined the relationship between juror gender and verdict preferences in criminal cases involving rape or child sexual abuse. They analyzed 32 studies containing 36 effects (19 for rape and 17 for child sexual abuse). Both types of cases produced small but reliable effects for mock juror gender, with females more likely to prefer conviction in cases involving rape ($\bar{r} = .13$) or child sexual abuse ($\bar{r} = .21$).

Present Study

A large empirical literature now exists on a wide variety of participant characteristics, and some of those characteristics have been meta-analyzed. Previous meta-analytic examinations have generally yielded small overall effects along with some indication of moderation by other variables. However, a number of participant characteristics have yet to be meta-analyzed, and prior meta-analyses are now both dated and limited by various constraints, including excessively broad inclusion criteria; large proportions of effects treated as zero because of reported statistical nonsignificance; and moderator analyses with very small numbers of effects at one or more levels (i.e., Mazzella & Feingold, 1994), or a limited set of case types (i.e., Schutte & Hosch, 1997). Further, there are now many more studies available for analysis because of vastly improved computerized database search tools, increased electronic accessibility of manuscripts, and the sustained empirical attention of juror researchers over the last two decades. Thus, the time seems right to conduct a broad meta-analytic examination of the most frequently studied participant characteristics.

The theoretical starting point for our work was the well-known story model (Bennett & Feldman, 1981; Pennington & Hastie, 1986, 1993). Rather than viewing jurors as weighting information “cues” based on their probative value and then aggregating them in a linear fashion to arrive at a preferred verdict, the story model proposes that jurors formulate verbal narratives to explain the events leading up to trials. Stories are mental representations of trial-related events that feature actors who have intentions and take action to bring about their goals. A key feature of the story model is that jurors are not seen as undifferentiated automatons who make decisions based on a rational, standardized processing of the evidence. Instead, jurors’ stories are viewed as heavily influenced by their life experiences and perceptions at trial. Empirical research supports the fundamental role of stories (Pennington & Hastie, 1986, 1988, 1992), but relatively little attention has been devoted to the individual differences and cognitive structures that lead jurors in the same trial to adopt different stories.

Devine (2012) recently extended the story model in formulating an integrative model of juror decision making. Based on a systematic review of the empirical literature, the “director’s cut” model identifies a number of juror and defendant characteristics likely to play a role in shaping jurors’ preferred verdicts at trial. Focal juror characteristics include race, gender, SES, trust in the legal system, and need for cognition; focal defendant characteristics consist of race, SES, prior criminal record, physical attractiveness, and courtroom demeanor. In essence, the director’s cut model says that jurors initial mental representations of trial-related events are determined by juror and defendant characteristics along with any information acquired before the trial via the media, the nature of the charges, and the attorneys’ opening statements. Initial trial representations then interact with characteristics of the incoming evidence and serve as the basis for formulating one or more stories, which are then translated into mental models for evaluation. How stories fare when tested via mental simulation then has direct implications for a juror’s preferred verdict.

Figure 1 depicts the conceptual model that guided our meta-analytic investigation. It links a number of frequently studied participant characteristics to jurors’ judgments about criminal guilt via criminal-related stereotypes and case-related scripts. A fundamental premise underlying the model is that existing cognitive structures interact with incoming trial-related information to provide jurors with a set of information used to construct an initial mental representation of the trial. Two types of cognitive structures important to story formation are stereotypes of “criminals” and scripts associated with different criminal cases. Stereotypes are person-related categories consisting of a central label and associated behavioral attributes; scripts represent sets of related events that are understood to occur in a causal sequence. In essence, stories are based on inferences about what is true, but some inferences come from information stored in memory and made accessible via the activation of jurors’ stereotypes and scripts.

A second premise of the model is that jurors’ stereotypes and scripts are a function of life experiences, which are in turn associated with their personal characteristics. In particular, jurors with certain characteristics may be more likely to possess “criminal” stereotypes and case-related scripts that lead them to process defendant-related perceptions and trial-related information in ways conducive to the formation of a proconviction story. Such characteristics would include being attitude-disposed to favor stories offered by the prosecution (particularly those that center on the testimony of authority figures) and being less inclined to formulate or give serious consideration to alternative stories offered by the defense. Jurors might also be expected to be more likely to favor proconviction stories when considering an “outgroup” defendant. Thus:

Proposition 1: Jurors will be more likely to convict when they are: (a) high in legal system trust, (b) high in authoritarianism, (c) low in need for cognition (NC), or (d) deciding a case involving a defendant of a different race.

A third premise of the model is that much of what jurors “know” about crimes and criminal behavior is gleaned from various media sources (e.g., newspapers, TV shows, movies, and Internet websites). Because of consistencies in the way crimes and criminals are depicted in these media, certain defendant characteristics
should be more likely to be incorporated into jurors’ “criminal” stereotypes, making their perception at trial more likely to trigger inferences of culpability. In particular, criminals are often depicted as poor, male, unattractive, young, and violent. They are also typically portrayed as acting from stable internal dispositions that translate into consistent patterns of antisocial behavior. Jurors’ perception of these prototypical “criminal” characteristics may activate “criminal” stereotypes—and associated behavioral inferences conducive to the formation of stories in which the defendant committed the alleged crime. Thus:

**Proposition 2:** Criminal defendants will more likely be convicted when they are: (a) low SES, (b) physically unattractive, (c) male, or (d) known to have a prior criminal history.

At the same time, “criminal” stereotypes may to some extent be crime-specific. Any given defendant characteristic may be viewed as typical for some types of crimes but not others. At trial, jurors likely assess the degree of fit between the defendant’s (perceived) characteristics and the presumed motives and capabilities of individuals seen as (typically) committing that type of crime. For example, perceiving a defendant to be rich may have no story-related implications in a case involving reckless homicide, but it might seem incongruous in a burglary case, leading jurors to be less likely to convict a high-SES defendant in such a case. Furthermore, juror tendencies noted previously with regard to their willingness to formulate and consider different stories may be stronger in some cases than others. For example, the tendency to trust socially prestigious individuals should be more relevant in cases that revolve around the testimony of police officers or “expert” witnesses. This leads to:

**Proposition 3:** Relationships between participant characteristics and judgments of guilt will vary as a function of case type.

In addition to case type, observed empirical relationships may vary across studies because of methodological “noise.” Two methodological variables that have differed across studies of juror decision making are the type of outcome measure obtained and the type of participants used. With regard to outcomes, most experimental studies have involved participants making a dichotomous judgment of guilt (i.e., guilty v. not guilty), but mock jurors are sometimes instructed to indicate their beliefs about a case using continuous rating scales capturing the magnitude and/or certainty of guilt. Having jurors make a dichotomous judgment about a continuous phenomenon could introduce a bias that systematically weakens the magnitude of observed relationships. With regard to the participants, most studies of juror decision making have featured undergraduate students as opposed to samples that are more representative of actual jurors (e.g., those who show up for jury duty, or a broader sampling of community members). College students tend to be less demographically diverse than the jury-eligible population, younger, more affluent, and higher in cognitive ability. These systematic differences may be associated with differences in the magnitude or even direction of relationships observed between participant characteristics and guilt judgments across studies. Thus:

**Proposition 4:** Relationships between focal participant characteristics and juror guilt judgments will vary as a function of both outcome type and participant type.

### Method

**Literature Search**

Four methods were used to identify potential studies for our meta-analyses: (a) computerized search of electronic article databases, (b) manual search of selected empirical journals and recent American Psychology-Law Society (APLS) conference programs, (c) inspection of the references of prior meta-analyses involving juror or defendant characteristics, and (d) emailing established jury researchers. We employed these methods in a converging, complementary fashion in an effort to identify all existing relevant studies conducted as of December 31, 2012. Usable studies pub-
lished in the first half of 2013 were also included if we came across them, but we did not systematically search for them.

With respect to the first three methods, we searched a variety of well-known electronic databases (i.e., Google Scholar, PsycINFO, HeinOnline, LexisNexis, ProQuest) using a broad set of keyword search terms that included: jury, juror, juries, verdict, and decision. Targeted follow-up searches were also conducted for each focal characteristic (e.g., juror gender) using more customized keyword combinations. Multiple searches were conducted using the ProQuest database in order to review unpublished theses and dissertations. We manually searched the following behavioral science journals that regularly publish empirical research on juror decision making: *Law and Human Behavior; Psychology, Public Policy, and the Law; Behavioral Sciences and the Law; Journal of Experimental Psychology: Applied; Journal of Applied Psychology; and Journal of Applied Social Psychology*. Article titles in these journals were scanned from either the earliest published volume or January 1970, whichever was more recent. APLS conference programs were searched for the years 2009–2012 and we requested copies of any paper or poster that appeared to be empirical and involve relevant participant characteristics. Finally, we reviewed the references of earlier published meta-analyses (i.e., Mazzella & Feingold, 1994; Mitchell et al., 2005; Narby et al., 1993; Schutte & Hosch, 1997) and tried to obtain any study they included. Ultimately, we obtained 29 usable theses and dissertations, 12 usable conference presentations, and two other usable manuscripts that were unpublished as of 2012.

We also emailed a large number of individuals known to conduct research on juror decision making for either of two reasons. First, we emailed the lead author of published journal articles and APLS presentations in an effort to obtain unreported effect size and/or sample size information for studies that were otherwise usable. Approximately 45 researchers were contacted regarding specific “high-likelihood” published studies, and roughly two thirds of these inquiries eventually produced sufficient information to include the targeted study. Second, we compiled a list of about 35 persons known to conduct research on juries and attempted to contact them by email to inquire about any unpublished studies they may have conducted that might be usable in our meta-analyses. In contrast to the study-specific inquiries, these general inquiries did not produce many usable studies for several reasons. We were unable to find valid email addresses for some of these individuals; a few emails did not elicit a reply; and those individuals who did respond generally reported having no access to relevant data. Nonetheless, a few general inquiries did yield analyzable effects.

These four search strategies were used iteratively over an extended period of time in order to obtain studies conducted (or published) during the course of the project, as well as to provide checks on earlier searches and obtain studies that were not previously available for some reason (e.g., changes in electronic database access).

### Inclusion and Exclusion Criteria

In general, we included studies in our meta-analyses if they: (a) featured an experimental design that controlled for trial content, (b) manipulated or measured one or more focal participant characteristics, (c) measured an individual-level predeliberation judgment related to guilt in a criminal trial context or sentence in a capital trial, and (d) reported a usable effect-size statistic and corresponding sample size. Studies were excluded if they: (a) involved an academic infraction (e.g., student cheating), (b) used undocumented idiosyncratic measures of a focal juror characteristic, or (c) employed a perfunctory manipulation of a focal defendant characteristic.

A few notes are in order with regard to these criteria. We did not examine studies that involved civil-law claims or reported data solely at the jury level of analysis. With regard to the outcome, we excluded studies wherein participants were asked to indicate only a probability of commission or recommended punishment in a noncapital trial.

With respect to the measurement of focal juror characteristics, we identified acceptable measures based on general acceptance and use in the literature. For juror authoritarianism, these included the F Scale (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950), Legal Authoritarianism Questionnaire (Boehm, 1968), Revised Legal Authoritarianism Questionnaire (RALQ; Kravitz, Culver, & Brock, 1993), Mitchell and Byrne Authoritarianism Scale (MBAS; Mitchell & Byrne, 1973), and Right Wing Authoritarianism Scale (RWAS; Altemeyer, 1981). For legal system trust, we decided to include only those studies that employed the Juror Bias Scale (JBS; Kassin & Wrightsman, 1983) or one of its subscales (i.e., Probability of Commission [PC] or Reasonable Doubt [RD]). Similarly, for juror NC, we included only studies using Petty and Cacioppo’s measure (Petty & Cacioppo, 1982; Cacioppo, Petty, & Kao, 1984) in either its long or short form. A few juror characteristics presented a challenge in having no well-established measure. For prior experience as an actual juror, we deemed acceptable any study that asked respondents to indicate prior service as a juror either dichotomously (yes-no) or continuously (i.e., number of times served). SES has typically been measured in various ad hoc ways using one or more items pertaining to education level, occupational status, and/or annual income. To maximize the number of included studies as well as the interpretability of the corresponding effect, we limited our focus to only the most common indicator—education—as indexed by highest grade level completed or highest academic degree obtained. We excluded studies that only reported measures of occupation and/or annual income.

With respect to defendant characteristics, we included only those studies in which a focal characteristic was experimentally manipulated in the context of a mock trial. We did this because actual trials are susceptible to many confounding variables, the most dangerous being extraneous variation in the strength of the evidence. For defendant physical attractiveness, we included only those studies in which the defendant’s appearance was manipulated visually (i.e., with photographs) as opposed to verbally. For defendant criminal record, studies were required to include a control group that was not exposed to any information about a prior record. For defendant race, we examined only those studies featuring White, Black, and/or Hispanic defendants and reported effects for distinguishable groups of White, Black, and/or Hispanic mock jurors. This excluded studies in which juror race was not measured, or was recorded but the results reported in aggregate form (thereby making it impossible to determine precise effects for specific juror race-defendant race combinations). Finally, as in Mitchell et al. (2005), samples comprising 95% or more White
participants were treated as being all-White in order to include several studies featuring almost entirely white samples.

Studies also had to report a usable statistical effect and an associated sample size (N). Usable statistics included those convertible to a Pearson correlation coefficient (r) via Comprehensive Meta-Analysis (CMA) software (Borenstein, Hedges, Higgins, & Rothstein, 2005), such as $d$, $\chi^2$, $t$, $M/SD$, $2x2$ raw frequencies, or by hand via formulas by Morris and DeShon (1997) and Johnson and Eagly (2000) in the case of $F$ statistics. Regression coefficients associated with multiple regression were not included unless they could be obtained from the first step when no other predictors were included in the model.

Coding

General procedures. For every usable effect, we extracted the necessary information regarding sample size and effect size, and assigned a code for each of the three moderator variables (i.e., participant type, outcome type, and case type). Judgment calls associated with these decisions were made by discussing the issue until a consensus decision was reached. Studies reporting a non-significant relationship but no precise test statistic or effect-size indicator were included using the conservative convention of assigning a zero effect (i.e., $d = .00$ or $r = .00$). In studies where participants made judgments for multiple case scenarios, the mean effect calculated across the various cases was used in the overall analysis unless one scenario could clearly be identified as the most appropriate, in which case its effect was used.

At several points, we double-checked the accuracy of the information entered into the CMA data files and, in the fall of 2012, we conducted a comprehensive audit of all entered data. This involved accessing each study (and any associated emailed information) to verify the accuracy of the entered data. This check revealed an extremely high level of accuracy that was deemed sufficient to warrant conducting the primary analyses.

Moderator variables. For participant type, we sorted available studies into four categories: student, community, venirepersons, or mixed. **Students** were those enrolled in university courses at the undergraduate or graduate level, whereas **community** participants were designated as those recruited from a surrounding area but not in a manner directly associated with a university. **Venirepersons** consisted of individuals who showed up for jury duty, including those who appeared but did not actually serve on a jury. The **mixed** category was used for samples containing both students and participants from one or more nonstudent categories.

For outcome type, we coded the nature of the criterion measure obtained using three categories: dichotomous, continuous, or capital sentence. The vast majority of studies featured outcome measures associated with a juror’s preferred verdict in either dichotomous (e.g., guilty v. not guilty) or continuous form (e.g., guilt ratings or verdict confidence indices). When three or more verdict options were available to participants (e.g., when lesser included charges were present), responses were collapsed into dichotomous categories (i.e., guilty v. not guilty).

Case type was coded using seven categories: homicide, child sexual abuse (child SA), adult sexual assault (adult SA), violent, property-related, mixed, and miscellaneous. **Homicide** trials featured defendants tried for the criminal death of one or more persons, including charges of murder (unspecified), first-degree murder, second-degree murder, voluntary manslaughter, involuntary manslaughter, and reckless homicide. **Child SA** cases involved molestation or sexual abuse of a minor; **adult SA** cases involved rape or other sexual assault against a victim 18 years of age or older. **Violent** trials featured crimes involving physical harm to one or more persons or the threat of it (e.g., robbery or assault). **Property** cases involved the nonviolent illegal taking of money or other material possessions (e.g., theft, burglary, larceny). **Multiple** was used to designate studies where participants read two or more scenarios involving two or more case types, whereas **mixed** was assigned to studies wherein case type was systematically varied in a between-subjects design but the data were reported in an unspecified, aggregated fashion. **Miscellaneous** referred to studies wherein the trial type did not match any of the other categories (e.g., arson, stalking, fraud, obstruction of justice). In addition, for several characteristics with larger empirical literatures (e.g., juror gender), we coded cases involving a death using two categories in addition to **homicide**: homicide-BWS and capital. **Homicide-BWS** involved cases in which a female defendant was charged with killing an abusive partner and the defense invoked the battered woman syndrome (BWS). **Capital** cases involved a charge of first-degree murder and a potential sentence of death.

Meta-Analytic Approach

CMA software (Borenstein et al., 2005) was used for all main and moderator analyses. In keeping with the emerging consensus in the meta-analysis technical literature, random-effects models were used for all analyses (Kepes, McDaniel, Brannick, & Banks, 2013). To determine whether sufficient effect-size heterogeneity existed to warrant moderator analyses, we employed the $Q$ statistic (Rosenthal, 1991) and $I^2$ values (Higgins & Thompson, 2002; Huedo-Medina, Sánchez-Meca, Marin-Martínez, & Botella, 2006). The $Q$ statistic is a chi-square test value associated with the null hypothesis that variation in the observed distribution of effects does not exceed what would be expected by chance; a significant value indicates additional sources of influence on the observed effects beyond sampling error and is consistent with the existence of moderator variables. The $I^2$ statistic estimates the percentage of total variability in a set of effect sizes due to true heterogeneity (Huedo-Medina et al., 2006). $I^2$ has no associated significance test, but higher values (i.e., those closer to 100%) are consistent with the presence of moderators. When our overall analyses produced significant $Q$ statistics and $I^2$ values larger than 50% for a characteristic, we followed up the overall analysis by examining the three potential moderators.

All effect sizes were converted to Pearson correlations ($r$) for meta-analyses and the resulting point estimates represent sample-weighted mean correlations (denoted here as $r$). For juror characteristics of a continuous nature (i.e., authoritariansm, prior experience as a juror, need for cognition, trust in the legal system, and education), a positive $\tilde{r}$ indicates jurors with higher levels of the characteristic were more likely to vote guilty. For juror gender, a positive $\tilde{r}$ indicates female jurors were more likely to vote guilty than male jurors. For the defendant characteristics, a positive $\tilde{r}$ indicates defendants with the following characteristics were more likely to be found guilty: (1) higher SES, (2) greater physical attractiveness, (3) a prior criminal record, and (4) female. For the analysis of defendant race, a positive $\tilde{r}$ indicates an outgroup...
severity bias (e.g., White jurors returning more guilty verdicts for Black defendants than White), whereas a negative \( r \) indicates an ingroup severity bias (e.g., White jurors provided more guilty verdicts for White defendants than Black).

**Results**

Here we present the results of the overall analyses that included every available study for each characteristic, followed by moderator analyses where the available studies were sorted into categories associated with the various levels of the respective moderator variable (i.e., participant type, outcome measure, and case type) and the data reanalyzed at each level. Within each section, results are grouped into two subsections: those pertaining to juror characteristics and those associated with defendant characteristics. We generally refer to the outcome variables as verdicts or guilt judgments despite a small proportion of the effects corresponding to sentencing judgments in the context of a death-penalty trial. All of our estimated meta-analytic effects are of a magnitude less than what Cohen (1992) labeled as “medium” in size (i.e., \(| r | \geq 0.30\), with most falling around his reference value for a “small” effect (i.e., \(| r | = 0.10\)). Table 1 provides a summary of the overall analyses for all 11 participant characteristics.

### Overall Analyses

**Juror characteristics.** The overall analyses for the six juror characteristics produced a range of observed effects. Starting with the weakest, the effects for both education level (\( \bar{r} = 0.00, k = 20 \)) and prior experience as a juror (\( \bar{r} = 0.03, k = 10 \)) were negligible in size and associated with confidence intervals that included zero, suggesting neither characteristic has a robust relationship with guilt judgments. Need for cognition (\( \bar{r} = -0.07, k = 10 \)) and gender (\( \bar{r} = 0.08, k = 215 \)) produced weak effects, with women and those lower in NC being slightly more likely to prefer conviction compared with men and those who had higher levels of NC, respectively. The largest observed relationships were for authoritarianism and trust in the legal system as measured by the JBS. The overall effect for authoritarianism (\( \bar{r} = 0.17, k = 36 \)) reflects a tendency for those with higher levels of the trait to be more likely to prefer conviction than those with lower levels. Higher scores on the JBS correspond to dispositional beliefs that would make a juror more likely to trust that the defendant is the actual perpetrator. The JBS-Total Score based on all 22 items exhibited the strongest overall relationship with judgments of guilt (\( \bar{r} = 0.22, k = 17 \)), although the two JBS subscales each yielded slightly weaker effects that were still notable in magnitude (JBS-PC: \( \bar{r} = 0.16, k = 12 \); JBS-RD: \( \bar{r} = 0.17, k = 11 \)). In essence, jurors with greater trust in the legal system are more likely to convict than jurors with more skeptical beliefs.

**Defendant characteristics.** The overall analyses for the five defendant characteristics yielded a smaller range of effects on guilt judgments compared with the juror characteristics. For three characteristics, the sample-weighted mean effects were very close to zero: gender (\( \bar{r} = 0.02, k = 25 \)), race (\( \bar{r} = 0.03, k = 51 \)), and physical attractiveness (\( \bar{r} = 0.04, k = 12 \)). There is accordingly no evidence of a robust relationship between these defendant characteristics and juror judgments of guilt, but the overall effect for defendant race must be interpreted carefully in light of the way the data were coded. In essence, it reflects a slight observed tendency across all jurors to show favoritism toward defendants of the same race. However, when the studies were sorted by juror race and reanalyzed, differences emerged. There was no indication

---

**Table 1**

*Meta-Analytic Results for Bivariate Relationships Between Juror/Defendant Characteristics and Guilt Judgments*

<table>
<thead>
<tr>
<th>Juror/Defendant characteristics</th>
<th>( N )</th>
<th>( k )</th>
<th>( \bar{r} )</th>
<th>Lower</th>
<th>Upper</th>
<th>( Q )</th>
<th>( F^2 )</th>
<th>IV = 0 (Low)</th>
<th>IV = 1 (High)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defendant Attractiveness</td>
<td>1599</td>
<td>12</td>
<td>-0.04</td>
<td>-0.12</td>
<td>0.04</td>
<td>22.27</td>
<td>51</td>
<td>52</td>
<td>48</td>
</tr>
<tr>
<td>Defendant Gender</td>
<td>4172</td>
<td>25</td>
<td>0.02</td>
<td>-0.04</td>
<td>0.08</td>
<td>78.06</td>
<td>69</td>
<td>49</td>
<td>51</td>
</tr>
<tr>
<td>Defendant SES</td>
<td>4180</td>
<td>20</td>
<td>-0.11**</td>
<td>-0.14</td>
<td>-0.07</td>
<td>20.61</td>
<td>8</td>
<td>56</td>
<td>45</td>
</tr>
<tr>
<td>Defendant Prior Criminal Record</td>
<td>2921</td>
<td>19</td>
<td>0.12**</td>
<td>0.05</td>
<td>0.19</td>
<td>51.02</td>
<td>65</td>
<td>44</td>
<td>56</td>
</tr>
<tr>
<td>Defendant Race</td>
<td>7076</td>
<td>51</td>
<td>0.03</td>
<td>0.03</td>
<td>0.09</td>
<td>267.58</td>
<td>81</td>
<td>49</td>
<td>52</td>
</tr>
<tr>
<td>W. Jurors with W/B or W/H Defendant</td>
<td>5793</td>
<td>39</td>
<td>0.01</td>
<td>-0.05</td>
<td>0.07</td>
<td>224.05**</td>
<td>83</td>
<td>50</td>
<td>51</td>
</tr>
<tr>
<td>W. Jurors with W/B Defender</td>
<td>4476</td>
<td>32</td>
<td>-0.02</td>
<td>-0.09</td>
<td>0.06</td>
<td>198.73</td>
<td>84</td>
<td>51</td>
<td>49</td>
</tr>
<tr>
<td>W. Jurors with W/H Defender</td>
<td>1317</td>
<td>7</td>
<td>0.11*</td>
<td>0.03</td>
<td>0.19</td>
<td>13.65</td>
<td>56</td>
<td>45</td>
<td>56</td>
</tr>
<tr>
<td>B. Jurors with W/B Defender</td>
<td>1029</td>
<td>10</td>
<td>0.13*</td>
<td>0.01</td>
<td>0.25</td>
<td>28.33*</td>
<td>68</td>
<td>44</td>
<td>57</td>
</tr>
<tr>
<td>Juror Need for Cognition</td>
<td>1747</td>
<td>10</td>
<td>-0.07*</td>
<td>-0.13</td>
<td>-0.00</td>
<td>15.91</td>
<td>43</td>
<td>54</td>
<td>47</td>
</tr>
<tr>
<td>Juror Experience</td>
<td>7025</td>
<td>10</td>
<td>0.03</td>
<td>-0.01</td>
<td>0.07</td>
<td>20.03*</td>
<td>55</td>
<td>49</td>
<td>52</td>
</tr>
<tr>
<td>Juror Education</td>
<td>8298</td>
<td>20</td>
<td>0.00</td>
<td>-0.03</td>
<td>0.03</td>
<td>31.04*</td>
<td>39</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Juror Authoritarianism</td>
<td>8205</td>
<td>36</td>
<td>0.17**</td>
<td>0.14</td>
<td>0.20</td>
<td>67.72**</td>
<td>48</td>
<td>42</td>
<td>59</td>
</tr>
<tr>
<td>Juror Trust in Legal System</td>
<td>60480</td>
<td>215</td>
<td>0.08**</td>
<td>0.06</td>
<td>0.10</td>
<td>927.18**</td>
<td>77</td>
<td>46</td>
<td>54</td>
</tr>
</tbody>
</table>

*Note.* \( W = \) White; \( B = \) Black; \( H = \) Hispanic; JBS = Juror Bias Scale (Kassin & Wrightsman, 1983); \( N \) = number of participants pooled across samples; \( k \) = number of unique samples; \( r \) = meta-analytic sample-weighted mean correlation; CI = confidence interval; \( Q \) = test statistic distributed as chi-square when \( \bar{r} \) is a significant value indicates the observed variation in effects is greater than expected by chance; \( F^2 \) = estimated percentage of observed variation in effects because of true heterogeneity; BESD = binomial effect size display (Rosenthal & Rubin, 1982); IV = independent variable.

*\( p < .05 \), two-tailed. **\( p < .01 \), two-tailed.
of outgroup severity bias in the 32 studies where White mock jurors decided a case involving a White versus Black defendant ($\bar{r} = -0.02$), but in the seven studies where White mock jurors decided a case involving a White versus Hispanic defendant, a modest outgroup severity bias emerged ($\bar{r} = .11$). Black mock jurors also demonstrated a modest outgroup bias against White defendants in the 10 studies featuring a White versus Black defendant ($\bar{r} = .13$).

The other two defendant characteristics—prior criminal record and SES—produced overall effects on guilt judgments above the .1 threshold. The modest negative effect for defendant SES ($\bar{r} = -.11, k = 20$) reflects a general tendency for lower-SES defendants to be somewhat more likely to be convicted than high-SES defendants; the modest positive effect associated with prior criminal record ($\bar{r} = .12, k = 19$) indicates defendants known to have a criminal history were somewhat more likely to be convicted.

Assessment of moderation likelihood. A primary question arising in the wake of the overall analyses concerns the likelihood that the relationship between any given focal characteristic and guilt judgments depends on one or more moderator variables. Two statistics commonly used to answer questions of this sort are $Q$ and $F^2$. A review of these values for the 11 overall analyses strongly supports the conclusion that most of the relationships between the focal personal characteristics and guilt judgments are moderated by at least one other variable. In particular, a significant $Q$ statistic resulted for nine of the 11 characteristics (i.e., all but defendant SES and juror NC). Further, with the exception of defendant SES, $F^2$ values were substantially larger than zero for all characteristics (ranging from 39% for juror education to 81% for defendant race). Thus, we proceeded to examine the extent to which the three focal moderators could explain variation in the overall distribution of effects associated with all of our focal characteristics except defendant SES.

A few words are in order about our moderator analyses. The small number of available effects for many of the 11 characteristics precluded some moderator analyses. For any given characteristic, we analyzed a specific moderator when there were five or more effects available for at least two levels (e.g., five based on dichotomous verdicts and five based on continuous guilt ratings). Unfortunately, three focal characteristics yielded an insufficient number of effects to allow any moderator analyses—defendant physical attractiveness, juror need for cognition, and juror experience. Given the results of the overall analyses, all that can be determined from the data at this point is that one or more moderators is likely operating in these domains. Moderator analyses for three other characteristics—defendant gender, defendant criminal record, and juror education—were limited by a skewed distribution of effects across levels of the moderator variable (e.g., 16 effects in one level but only 4 in the other). Our ability to examine moderators for these three characteristics was thus limited. When moderator analyses could be conducted, statistical support for the moderating role of a particular variable exists when: (a) estimated effect-sizes differ across levels of the moderator and (b) within each level of the moderator, the $Q$ statistic is not significant and the $F^2$ value is low (or at least smaller than the $F^2$ for the overall analysis for that characteristic).

**Participant Type Moderator Analyses**

The four participant type categories consisted of student, community, venireperson, and mixed. Sufficient data were available to examine two or more levels of this moderator for five characteristics: juror education level, juror authoritarianism, juror gender, defendant prior criminal record, and defendant gender. For all but juror education level, one of the comparison groups was students. Because of an obvious restriction in range, there were no effects based on students in our distribution of effects for juror education, but there were enough effects to compare community members with venirepersons. Table 2 presents the results of the moderator analyses for participant type.

In general, no particular participant type yielded consistently stronger or weaker effect sizes than the other types. For three of the characteristics—defendant prior criminal record, juror authoritarianism, and juror gender—observed effects were in the same direction (i.e., positive or negative) for all subgroups in the analysis. For both defendant prior criminal record and juror gender, slightly stronger relationships resulted with student samples as opposed to nonstudents (i.e., community residents and/or venirepersons). For defendant prior criminal record, positive relationships were observed for both student participants ($\bar{r} = .14, k = 11$) and community members ($\bar{r} = .09, k = 6$), such that defendants with a known criminal history were somewhat more likely to be found guilty. For juror gender, there was essentially no relationship with guilt judgments for studies involving community residents or venirepersons ($\bar{r} = .01, k = 31$ and $\bar{r} = .00, k = 17$, respectively), but females were somewhat more likely to convict than males in studies based on student participants ($\bar{r} = .10, k = 159$). The pattern of weaker relationships for nonstudent samples did not hold for juror authoritarianism, however, where the observed relationships for both community residents ($\bar{r} = .18, k = 8$) and venirepersons ($\bar{r} = .23, k = 5$) were somewhat stronger than for students ($\bar{r} = .16, k = 22$). For all three groups, those high in authoritarianism were more likely to convict than those who were low.

Of note, the direction of the effect differed across participant types for two of the focal characteristics. For juror education level, a negligible negative relationship was observed for community members ($\bar{r} = -.02, k = 7$) in contrast to a negligible positive relationship for venirepersons ($\bar{r} = .04, k = 10$), although confidence intervals included zero for both groups. The “reversed” direction is more intriguing for defendant gender because of the larger size of the effects involved. Here, a weak but significant positive effect was observed for students ($\bar{r} = .06, k = 17$) in contrast to a modest negative effect for community members ($\bar{r} = -.09, k = 8$). In other words, male defendants were slightly more likely to be convicted than female defendants when judged by students, but slightly less likely to be convicted when judged by community members.

To summarize, differences in the estimated effects across participant types were generally fairly small (i.e., $r$ values within ±.10). In addition, the $Q$ tests remained significant for most participant types within the five focal characteristics, pointing toward the likely existence of other moderators. As such, there is little to suggest the nature of the sample has a substantial impact on the estimated relationship between participant characteristics and guilt judgments.
Table 2
Meta-Analytic Results for Bivariate Relationships Between Juror/Defendant Characteristics and Guilt Judgments With Participant Type As A Moderator

<table>
<thead>
<tr>
<th>Juror/Defendant characteristic</th>
<th>N</th>
<th>k</th>
<th>( \bar{r} )</th>
<th>95% CI</th>
<th>( Q )</th>
<th>( I^2 )</th>
<th>IV = 0 (Low)</th>
<th>IV = 1 (High)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defendant Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>3354</td>
<td>17</td>
<td>.06**</td>
<td>.02</td>
<td>.10</td>
<td>17.66</td>
<td>9</td>
<td>47</td>
</tr>
<tr>
<td>Community</td>
<td>818</td>
<td>8</td>
<td>-.09</td>
<td>-.29</td>
<td>.12</td>
<td>57.41**</td>
<td>88</td>
<td>55</td>
</tr>
<tr>
<td>Defendant Prior Criminal Record</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>1947</td>
<td>11</td>
<td>.14**</td>
<td>.05</td>
<td>.22</td>
<td>30.39**</td>
<td>67</td>
<td>43</td>
</tr>
<tr>
<td>Community</td>
<td>864</td>
<td>6</td>
<td>-.09</td>
<td>-.06</td>
<td>.24</td>
<td>15.90*</td>
<td>69</td>
<td>46</td>
</tr>
<tr>
<td>Juror Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community</td>
<td>3179</td>
<td>7</td>
<td>-.02</td>
<td>-.06</td>
<td>.01</td>
<td>3.33</td>
<td>0</td>
<td>51</td>
</tr>
<tr>
<td>Venireperson</td>
<td>3868</td>
<td>10</td>
<td>.04</td>
<td>-.02</td>
<td>.09</td>
<td>21.70</td>
<td>59</td>
<td>48</td>
</tr>
<tr>
<td>Juror Authoritarianism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>4948</td>
<td>22</td>
<td>.16**</td>
<td>.11</td>
<td>.20</td>
<td>38.94*</td>
<td>46</td>
<td>42</td>
</tr>
<tr>
<td>Community</td>
<td>2005</td>
<td>8</td>
<td>.18**</td>
<td>.12</td>
<td>.24</td>
<td>14.41*</td>
<td>51</td>
<td>41</td>
</tr>
<tr>
<td>Venireperson</td>
<td>900</td>
<td>5</td>
<td>.23**</td>
<td>.14</td>
<td>.32</td>
<td>7.83</td>
<td>49</td>
<td>39</td>
</tr>
<tr>
<td>Juror Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>38185</td>
<td>159</td>
<td>.10**</td>
<td>.08</td>
<td>.13</td>
<td>648.58**</td>
<td>76</td>
<td>45</td>
</tr>
<tr>
<td>Community</td>
<td>12145</td>
<td>31</td>
<td>.01</td>
<td>-.03</td>
<td>.05</td>
<td>104.66**</td>
<td>71</td>
<td>50</td>
</tr>
<tr>
<td>Venireperson</td>
<td>7624</td>
<td>17</td>
<td>.00</td>
<td>-.07</td>
<td>.06</td>
<td>84.77**</td>
<td>81</td>
<td>50</td>
</tr>
</tbody>
</table>

Note. \( N \) = number of participants pooled across samples; \( k \) = number of unique samples; \( \bar{r} \) = meta-analytic sample-weighted mean correlation; CI = confidence interval; \( Q \) = test statistic distributed as chi-square where a significant value indicates the observed variation in effects is greater than expected by chance; \( I^2 \) = estimated percentage of observed variation in effects because of true heterogeneity; BESD = binomial effect size display (Rosenthal & Rubin, 1982); IV = independent variable.
* \( p < .05 \), two-tailed. ** \( p < .01 \), two-tailed.

Outcome Measure Moderator Analyses

There are two categories of primary interest for this moderator: dichotomous verdicts and continuous ratings of guilt. The majority of effects we obtained were based on dichotomous judgments of guilt, but enough were based on continuous ratings to conduct subgroup analyses for four focal characteristics: defendant gender, juror need for cognition, juror authoritarianism, and juror gender. In addition, sufficient effects were available for two of these focal characteristics (i.e., juror gender and juror authoritarianism) to examine a third type of outcome—dichotomous sentencing decision in a capital trial (i.e., life in prison or the death penalty). Table 3 presents a summary of the moderator analyses for outcome type.

Table 3
Meta-Analytic Results for Bivariate Relationships Between Juror/Defendant Characteristics and Guilt Judgments With Outcome Type as A Moderator

<table>
<thead>
<tr>
<th>Juror/Defendant characteristic</th>
<th>N</th>
<th>k</th>
<th>( \bar{r} )</th>
<th>95% CI</th>
<th>( Q )</th>
<th>( I^2 )</th>
<th>IV = 0 (Low)</th>
<th>IV = 1 (High)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defendant Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dichotomous</td>
<td>2480</td>
<td>11</td>
<td>.08**</td>
<td>.04</td>
<td>.13</td>
<td>5.41</td>
<td>0</td>
<td>46</td>
</tr>
<tr>
<td>Continuous</td>
<td>1138</td>
<td>11</td>
<td>-.07</td>
<td>-.22</td>
<td>.09</td>
<td>64.11**</td>
<td>84</td>
<td>54</td>
</tr>
<tr>
<td>Juror Need for Cognition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dichotomous</td>
<td>834</td>
<td>5</td>
<td>-.10*</td>
<td>-.19</td>
<td>-.01</td>
<td>6.46</td>
<td>38</td>
<td>55</td>
</tr>
<tr>
<td>Continuous</td>
<td>913</td>
<td>5</td>
<td>-.04</td>
<td>-.14</td>
<td>.05</td>
<td>8.00</td>
<td>50</td>
<td>52</td>
</tr>
<tr>
<td>Juror Authoritarianism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dichotomous</td>
<td>3859</td>
<td>18</td>
<td>.18**</td>
<td>.14</td>
<td>.23</td>
<td>29.03*</td>
<td>42</td>
<td>41</td>
</tr>
<tr>
<td>Continuous</td>
<td>1942</td>
<td>8</td>
<td>.10**</td>
<td>.03</td>
<td>.16</td>
<td>13.47</td>
<td>48</td>
<td>45</td>
</tr>
<tr>
<td>Sentence</td>
<td>1648</td>
<td>6</td>
<td>.22**</td>
<td>.17</td>
<td>.26</td>
<td>4.06</td>
<td>0</td>
<td>39</td>
</tr>
<tr>
<td>Juror Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dichotomous</td>
<td>49614</td>
<td>182</td>
<td>.10**</td>
<td>.08</td>
<td>.12</td>
<td>644.08**</td>
<td>72</td>
<td>45</td>
</tr>
<tr>
<td>Continuous</td>
<td>4902</td>
<td>19</td>
<td>.04</td>
<td>-.03</td>
<td>.10</td>
<td>105.33**</td>
<td>83</td>
<td>48</td>
</tr>
<tr>
<td>Sentence</td>
<td>3090</td>
<td>7</td>
<td>-.11**</td>
<td>-.15</td>
<td>-.07</td>
<td>7.92</td>
<td>24</td>
<td>56</td>
</tr>
</tbody>
</table>

Note. \( N \) = number of participants pooled across samples; \( k \) = number of unique samples; \( \bar{r} \) = meta-analytic sample-weighted mean correlation; CI = confidence interval; \( Q \) = test statistic distributed as chi-square where a significant value indicates the observed variation in effects is greater than expected by chance; \( I^2 \) = estimated percentage of observed variation in effects because of true heterogeneity; BESD = binomial effect size display (Rosenthal & Rubin, 1982); IV = independent variable.
* \( p < .05 \), two-tailed. ** \( p < .01 \), two-tailed.
In general, observed relationships between the three juror characteristics and culpability measures were somewhat stronger for dichotomous verdicts than continuous ratings. Specifically, this was true for need for cognition (\( \bar{r} = -.04, k = 5 \) for continuous; \( \bar{r} = -.10, k = 5 \) for dichotomous), gender (\( \bar{r} = .04, k = 19 \) for continuous; \( \bar{r} = .10, k = 182 \) for dichotomous) and authoritarianism (\( \bar{r} = .10, k = 8 \) for continuous; \( \bar{r} = .18, k = 18 \) for dichotomous). In other words, the overall tendencies for more antidepressant judgments when jurors were lower in need for cognition, female, or highly authoritarian were somewhat stronger when the outcome was a dichotomous verdict as opposed to a continuous rating of guilt. In addition, the positive relationship between juror authoritarianism and sentencing decisions was slightly stronger (\( \bar{r} = .22, k = 6 \)) than for either type of guilt measure. Jurors higher in authoritarianism were more likely to prefer the death penalty than those who were lower.

In contrast, there were changes in the direction of the effect for the moderator analyses involving gender. For defendant gender, female defendants were convicted more often than males in studies using dichotomous verdicts (\( \bar{r} = .08, k = 11 \)), whereas male defendants were rated as more guilty than female defendants in studies that obtained continuous ratings (\( \bar{r} = -.07, k = 11 \)). With regard to juror gender, the gender tendency for women to be more favorable to the prosecution was reversed for sentences in capital trials, with male jurors being more likely to choose the death penalty than female jurors (\( \bar{r} = -.11, k = 7 \)).

In sum, the general pattern that emerged from the moderator analyses for outcome type was a tendency for effects based on dichotomous verdicts to be somewhat stronger than effects based on continuous ratings. The magnitude of the difference in estimated effect size ranged from a low of .06 for need for cognition to a high of .21 for both juror gender. Further, both defendant gender and juror gender featured changes in the direction of the relationship such that the impact of gender may depend on the type of outcome in question.

**Case Type Moderator Analyses**

Case type was examined as a potential moderator variable based on the notion that jurors’ stereotypes of criminals are to some extent case-specific. For the moderator analyses associated with this variable, we sought to compare seven case types: capital, homicide, homicide-BWS, violent, property, adult SA, and child SA. Unfortunately, the preponderance of homicide and sexual assault trials in the literature limited our analyses of case type to only five characteristics: defendant gender, defendant prior criminal record, defendant race, juror authoritarianism, and juror gender. Table 4 presents a summary of the moderator analyses for case type.

Only two case types could be compared for defendant gender (homicide v. child SA) and defendant prior criminal record (violent v. property), and both comparisons produced effects with the same sign and fairly small differences in magnitude. Of note, although the effect for defendant gender in cases involving child SA was weak (\( \bar{r} = .09, k = 5 \)), its 95% confidence interval did not include zero and it was associated with a nonsignificant \( Q \) test. In addition, the observed effect of defendant criminal record was stronger for property cases (\( \bar{r} = .18, k = 5 \)) than violent crimes (\( \bar{r} = .11, k = 6 \)), hinting that an increased likelihood of conviction for defendants with a known criminal history may be somewhat stronger for cases involving theft and burglary.

There is also some support for the notion that the weak general tendency for jurors to be harsher toward defendants of a different race varies somewhat according to the type of case. Specifically, there was little if any indication of outgroup severity bias for violent cases (\( \bar{r} = -.02, k = 17 \)) or homicide cases (\( \bar{r} = .03, k = 14 \)), but noticeably more bias when trials involved property crimes (\( \bar{r} = .12, k = 5 \)) or adult sexual assault (\( \bar{r} = .13, k = 5 \)).

There is even more support for a moderating role of case type with respect to juror authoritarianism. In keeping with the modest positive relationship observed in the overall analyses, mock jurors higher in authoritarianism were more likely to convict in homicide cases (\( \bar{r} = .20, k = 13 \)) and more likely to decide against the defendant in capital trials (\( \bar{r} = .21, k = 8 \)). Conversely, juror authoritarianism exhibited a much weaker relationship with guilt judgments in adult SA cases (\( \bar{r} = .08, k = 8 \)). In addition to this separation in the magnitude of estimated effects, as further evidence of moderation, the \( Q \) statistic for each of the three case types was not significant and the corresponding \( F \) value was notably smaller than for the overall analysis.

Finally, the relationship between juror gender and guilt judgments appears quite likely to vary by case type. There was little if any relationship observed for three common types of criminal cases: homicide (\( \bar{r} = .01, k = 37 \)), violent (\( \bar{r} = .05, k = 24 \)), and property-related (\( \bar{r} = .06, k = 23 \)). However, consistent with the Schutte and Hosch (1997) meta-analysis, female jurors were notably more conviction-prone than males in cases involving adult sexual assault (\( \bar{r} = .16, k = 52 \)) or the sexual abuse of a child (\( \bar{r} = .18, k = 43 \)). This tendency for female jurors to favor the prosecution did not hold for every type of case, however. In homicide cases involving the killing of an abusive male spouse or domestic partner by a female defendant who invoked BWS as a defense, male jurors tended to convict more frequently than female jurors (\( \bar{r} = -.18, k = 13 \)). Further, male jurors were more likely to decide against the defendant than female jurors in capital trials (\( \bar{r} = -.07, k = 9 \)), with the seven of these effects attached to sentencing decisions in the second phase of a capital trial and the other two pertaining to guilt in the first phase. In general, there was notable separation in the magnitude of the observed effects for the different case types, and their type-specific \( F \) values were lower than the corresponding \( F \) value for the overall analysis. At the same time, the \( Q \) statistics generally remained significant for most case types. Thus, case type appears to moderate the relationship between juror gender and judgments of criminal culpability, but other moderators are probably operating in this domain.

**Follow-Up Analyses**

**Authoritarianism measures.** Given the small positive overall relationship observed between juror authoritarianism and guilt judgments and in light of the diversity of instruments used to measure juror authoritarianism, we conducted an additional moderator analysis for juror authoritarianism to see if the relationship varied as a function of the way in which authoritarianism was measured. Sufficient data were available to estimate the relationship for four instruments: F Scale (\( \bar{r} = .06, k = 9 \)), Right-Wing Authoritarianism Scale (\( \bar{r} = .16, k = 5 \)), Mitchell and Byrne Authoritarianism Scale (\( \bar{r} = .20, k = 5 \)), and Revised Legal Authoritarianism Scale.
### Table 4

*Meta-Analytic Results for Bivariate Relationships Between Defendant Characteristics and Guilt Judgments With Case Type as a Moderator*

<table>
<thead>
<tr>
<th>Defendant Characteristic</th>
<th>N</th>
<th>k</th>
<th>( \bar{r} )</th>
<th>Lower</th>
<th>Upper</th>
<th>( Q )</th>
<th>( F )</th>
<th>IV = 0 (Low)</th>
<th>IV = 1 (High)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Defendant Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homicide</td>
<td>1476</td>
<td>8</td>
<td>.05**</td>
<td>-.01</td>
<td>.11</td>
<td>8.25</td>
<td>15</td>
<td>48</td>
<td>53</td>
</tr>
<tr>
<td>Child SA</td>
<td>1213</td>
<td>5</td>
<td>.09**</td>
<td>.03</td>
<td>.15</td>
<td>2.47</td>
<td>0</td>
<td>46</td>
<td>55</td>
</tr>
<tr>
<td><strong>Defendant Prior Criminal Record</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Violent</td>
<td>870</td>
<td>6</td>
<td>.11</td>
<td>-.01</td>
<td>.24</td>
<td>13.92*</td>
<td>64</td>
<td>45</td>
<td>56</td>
</tr>
<tr>
<td>Property</td>
<td>449</td>
<td>5</td>
<td>.18</td>
<td>-.02</td>
<td>.37</td>
<td>11.64*</td>
<td>66</td>
<td>41</td>
<td>59</td>
</tr>
<tr>
<td><strong>Defendant Race</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All - Homicide</td>
<td>1551</td>
<td>14</td>
<td>.03</td>
<td>-.06</td>
<td>.13</td>
<td>45.91**</td>
<td>72</td>
<td>49</td>
<td>52</td>
</tr>
<tr>
<td>W Jurors with W/B Defendant - Homicide</td>
<td>796</td>
<td>8</td>
<td>-.09</td>
<td>-.21</td>
<td>.02</td>
<td>17.73**</td>
<td>61</td>
<td>55</td>
<td>46</td>
</tr>
<tr>
<td>All - Violent</td>
<td>2042</td>
<td>17</td>
<td>-.02</td>
<td>-.12</td>
<td>.07</td>
<td>69.55**</td>
<td>77</td>
<td>51</td>
<td>49</td>
</tr>
<tr>
<td>W Jurors with W/B Defendant - Violent</td>
<td>1855</td>
<td>14</td>
<td>-.05</td>
<td>-.15</td>
<td>.05</td>
<td>61.03**</td>
<td>79</td>
<td>53</td>
<td>48</td>
</tr>
<tr>
<td>All - Property</td>
<td>971</td>
<td>5</td>
<td>.12</td>
<td>-.07</td>
<td>.30</td>
<td>33.57**</td>
<td>88</td>
<td>44</td>
<td>56</td>
</tr>
<tr>
<td>All - Adult SA</td>
<td>806</td>
<td>5</td>
<td>.13**</td>
<td>.04</td>
<td>.22</td>
<td>6.45</td>
<td>38</td>
<td>44</td>
<td>57</td>
</tr>
<tr>
<td><strong>Juror Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Juror Authoritarianism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capital</td>
<td>1930</td>
<td>8</td>
<td>.21**</td>
<td>.15</td>
<td>.27</td>
<td>12.71</td>
<td>45</td>
<td>40</td>
<td>61</td>
</tr>
<tr>
<td>Homicide</td>
<td>2528</td>
<td>13</td>
<td>.20**</td>
<td>.15</td>
<td>.25</td>
<td>16.62</td>
<td>28</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>Adult SA</td>
<td>1652</td>
<td>8</td>
<td>.08*</td>
<td>.02</td>
<td>.13</td>
<td>9.34</td>
<td>25</td>
<td>46</td>
<td>54</td>
</tr>
<tr>
<td>Juror Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capital</td>
<td>3621</td>
<td>9</td>
<td>-.07*</td>
<td>-.13</td>
<td>.00</td>
<td>147.48**</td>
<td>74</td>
<td>54</td>
<td>47</td>
</tr>
<tr>
<td>Homicide</td>
<td>10283</td>
<td>37</td>
<td>.01</td>
<td>-.02</td>
<td>.04</td>
<td>69.37**</td>
<td>48</td>
<td>50</td>
<td>51</td>
</tr>
<tr>
<td>Homicide-BWS</td>
<td>2756</td>
<td>13</td>
<td>-.18**</td>
<td>-.24</td>
<td>-.12</td>
<td>23.55*</td>
<td>49</td>
<td>59</td>
<td>41</td>
</tr>
<tr>
<td>Violent</td>
<td>6274</td>
<td>24</td>
<td>.05*</td>
<td>.00</td>
<td>.09</td>
<td>46.17**</td>
<td>50</td>
<td>48</td>
<td>53</td>
</tr>
<tr>
<td>Property</td>
<td>7586</td>
<td>23</td>
<td>.06**</td>
<td>.03</td>
<td>.09</td>
<td>29.61</td>
<td>26</td>
<td>47</td>
<td>53</td>
</tr>
<tr>
<td>Adult SA</td>
<td>13733</td>
<td>52</td>
<td>.16**</td>
<td>.12</td>
<td>.19</td>
<td>.00</td>
<td>65</td>
<td>42</td>
<td>58</td>
</tr>
<tr>
<td>Child SA</td>
<td>12387</td>
<td>43</td>
<td>.18**</td>
<td>.15</td>
<td>.21</td>
<td>31.26**</td>
<td>61</td>
<td>41</td>
<td>59</td>
</tr>
</tbody>
</table>

**Note.** W = White; B = Black; Child SA = child sexual abuse; Adult SA = adult sexual assault; Homicide-BWS = homicide involving battered woman syndrome; N = number of participants pooled across samples; k = number of unique samples; \( \bar{r} \) = meta-analytic sample-weighted mean correlation; CI = confidence interval; \( Q \) = test statistic distributed as chi-square where a significant value indicates the observed variation in effects is greater than expected by chance; \( F \) = estimated percentage of observed variation in effects because of true heterogeneity; BESD = binomial effect size display (Rosenthal & Rubin, 1982); IV = independent variable.

* p < .05, two-tailed. ** p < .01, two-tailed.

Questionnaire (\( \bar{r} = .21, k = 11 \)). For all of these instrument-specific effects, the 95% confidence intervals did not include zero, associated \( Q \) statistics were not significant, and corresponding \( F \) values were notably smaller than for the overall effect. The relationship between juror authoritarianism and guilt judgments therefore appears to depend somewhat on the type of authoritarianism measure employed, with the relationship manifestly stronger when dedicated measures of legal authoritarianism are used (e.g., the RLAQ).

**Juror gender moderators.** In light of the effects associated with juror gender displaying some variability as a function of all three moderators, we conducted a multiple regression analysis using weighted least-squares (Lipsey & Wilson, 2001; Wilson, 2005) in which the observed effect in each of the primary studies was simultaneously regressed on dummy-coded versions of the three moderators: participant type (0 = Nonstudent; 1 = Student), outcome type (0 = Continuous, 1 = Dichotomous), and case type (0 = Nonsex crimes; 1 = Sex-related crimes). Collectively, the model explained 35% of the variation in observed effects (\( R = .59; R^2 = .35, p < .01 \)), and all three moderators had significant regression coefficients (\( \beta = .14 \) for participant type, \( \beta = .14 \) for outcome type, \( \beta = .50 \) for case type; all \( p \) values \( \leq .02 \)). In other words, each moderator variable uniquely explained a significant amount of variation in the observed juror gender effects when controlling for the other two, and case type explained the most variation of the three.

### Discussion

**Summary of Major Findings**

The preceding meta-analyses lead to several general conclusions: (1) The focal characteristics of defendants and jurors examined in this study have weak to modest overall relationships with juror predeliberation judgments of criminal culpability, (2) Most of these relationships are likely moderated by other variables and will accordingly vary across trial contexts, (3) The nature of the participants used in research does not appear to have a sizable or consistent effect on estimates of the relationship between participant characteristics and guilt judgments, (4) The nature of the outcome measure has a minor impact on estimated relationships, with slightly stronger effects observed for dichotomous verdicts as opposed to continuous ratings of guilt, and (5) The relationship...
between two juror characteristics (authoritarianism and gender) and judgments of defendant culpability likely varies by case type. See Table 5 for a summary of the effects observed in the preceding meta-analyses.

**Contribution to the Literature on Participant Characteristics**

A primary contribution of this study is its provision of point estimates of the relationship between 11 participant characteristics and judgments of culpability in criminal cases. Some of these estimates constitute initial quantitative assessments of their respective literatures whereas others represent updates of earlier meta-analyses. We obtained sufficient data to conduct first-time meta-analyses for one defendant characteristic (i.e., prior criminal record) and three juror characteristics (i.e., education, prior experience as a juror, need for cognition). We also provided updated estimates for four defendant characteristics (i.e., physical attractiveness, SES, gender, and trust) and two juror characteristics (authoritarianism and trust in the legal system as measured by the JBS) previously meta-analyzed by others. In addition, our analysis for authoritarianism distinguished between dedicated measures of authoritarianism and legal system attitudes as measured by the JBS (Kassin & Wrightsman, 1983). Finally, we conducted the first meta-analysis of juror gender studies that includes all available effects rather than just those from studies featuring sexual crimes.

Turning first to the magnitude of the overall relationships observed for the 11 characteristics, several analyses produced a point estimate indicative of little or no effect. These characteristics were: juror education level \( (r = .00) \), juror prior experience as a juror \( (r = .03) \), defendant physical attractiveness \( (r = .04) \), defendant gender \( (r = .04) \), and defendant race \( (r = .03) \). These overall point estimates did not differ significantly from zero and could be considered weak enough to ignore for practical purposes even if they do represent real effects. Indeed, as shown in Table 1, the binomial effect-size display (BESD; Rosenthal & Rubin, 1982) values associated with effects of this magnitude suggest the difference in the likelihood of a juror preferring conviction who is “high” on the characteristic as opposed to “low” is only a few percentage points. It is worth noting, however, that all of these characteristics were associated with significant \( Q \) tests for heterogeneity and thus featured a distribution of effects more variable than would be expected by chance alone. As a result, one or more moderators is probably operating in each domain, and there may be circumstances when each characteristic has a stronger relationship with guilt judgments.

Three of our focal defendant characteristics were meta-analyzed by Mazzella and Feingold (1994) almost 20 years ago—physical attractiveness, race, and gender. After considerable attention in the 1970s and 1980s, scholarly interest in defendant physical attractiveness apparently waned in the 1990s, and we found relatively few additional studies to analyze despite the passage of two decades. The weak overall effect obtained for defendant physical attractiveness \( (r = .04) \) based on 12 effects suggests a very slight tendency for those perceived as more attractive to be treated more leniently when it comes to guilt judgments. Our effect is also notably smaller than Mazzella and Feingold’s (1994) earlier estimate based on only eight studies \( (d = .19/r = .10) \). The difference is attributable to our exclusion of several questionable studies with relatively large effects included in the earlier analysis (e.g., Efran, 1974; Jacobson & Popovich, 1983), as well as the addition of several newer studies with very weak effects. Although they differ somewhat in the magnitude of their estimates, both meta-analyses indicate the relationship between defendant physical attractiveness and judgments of culpability is neither consistent nor especially strong.

Similarly, for defendant gender, we observed a very slight overall bias against female defendants \( (r = .02) \) in contrast to the weak bias against male defendants found by Mazzella and Feingold \( (d = -.08/r = -.04) \). The discrepancy appears to be primarily attributable to the addition of many new effects from post-1994 research. Given the lack of clear bias, stereotypes about men being more likely to commit crimes may simply not be as strong or as prevalent as suspected, or existing studies of defendant gender may fail to capture a real effect in being unrepresentative of a larger population. At the same time, it should be noted that there was some variability in the relationship between defendant gender and guilt judgments across case type, and many studies on defendant gender involved crimes in which defendant gender was salient (e.g., domestic homicide).

Our meta-analysis of studies that manipulated defendant race is directly comparable to the most recently published meta-analysis of race by Mitchell et al. (2005) in terms of assessing outgroup severity bias against defendants. The overall effect we obtained \( (r = .03) \) indicates a weak bias against defendants of a different race and is close to the corresponding value from the Mitchell et al. (2005) study \( (d = .09/r = .05) \). As did Mitchell and her colleagues, we found evidence of a stronger outgroup bias for Black jurors \( (r = .13) \) than White jurors \( (r = .01) \). Scholars have argued that racism in its modern form tends to be less overt and more likely to manifest itself when race is not a salient factor in the decision context (Sommers, 2007; Sommers & Ellsworth, 2001). It therefore could be that a real outgroup severity bias on the part of White jurors is being masked by the conspicuous nature of the defendant’s race in many experimental studies of juror decision making. However, in this the first published meta-analysis examining racial bias against Hispanic defendants, White jurors were found to exhibit substantially more outgroup severity bias toward Hispanic defendants \( (r = .11) \) than Black defendants \( (r = -.02) \). This analysis provides an initial estimate of outgroup severity bias against Hispanics and suggests social norms that serve to tamp down the display of racial bias by White jurors may be weaker for Hispanic defendants than Black defendants.

Turning to characteristics not previously meta-analyzed, two juror characteristics—education and prior experience as a juror—yielded negligible overall associations with guilt judgments. The null finding for juror education level is not especially surprising given the paucity of studies in which it was hypothesized to affect juror decision making. Conversely, some scholars have argued that serving as a juror tends to harden individuals and make them more cynical about criminal defendants, ultimately causing them to be more likely to convict than “novice” jurors. This view is consistent with several field studies of actual juries that yielded a positive association between the number of experienced jurors on the jury and conviction at the jury level (Dillehay & Nietzel, 1985; Werner, Strube, Cole, & Kagehiro, 1985). Nonetheless, the observed
### Table 5: Summary of Estimated Effects From Overall and Moderator Analyses

<table>
<thead>
<tr>
<th>Juror/Defendant characteristics</th>
<th>Participant type</th>
<th>Outcome type</th>
<th>Case type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall</td>
<td>Student</td>
<td>Community</td>
</tr>
<tr>
<td>Defendant Attractiveness</td>
<td>−.04</td>
<td>.02</td>
<td>.06</td>
</tr>
<tr>
<td>Defendant Gender</td>
<td>.00</td>
<td>.02</td>
<td></td>
</tr>
<tr>
<td>Defendant SES</td>
<td>−.11</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Defendant Prior</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Criminal Record</td>
<td>.12</td>
<td>.14</td>
<td>.09</td>
</tr>
<tr>
<td>Defendant Race</td>
<td>.03</td>
<td>.13</td>
<td>.03</td>
</tr>
<tr>
<td>W Jurors with W/B or</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W/H Defendant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W Jurors with W/B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defendant/italic</td>
<td>−.02</td>
<td>.11</td>
<td>.11</td>
</tr>
<tr>
<td>W Jurors with W/H</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B Jurors with W/B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defendant</td>
<td>.13</td>
<td></td>
<td>.13</td>
</tr>
<tr>
<td>Juror Need for Cognition</td>
<td>−.07</td>
<td>−.10</td>
<td>−.10</td>
</tr>
<tr>
<td>Juror Experience</td>
<td>.03</td>
<td></td>
<td>.03</td>
</tr>
<tr>
<td>Juror Education</td>
<td>.00</td>
<td>.16</td>
<td>.18</td>
</tr>
<tr>
<td>Juror Authoritarianism</td>
<td>.17</td>
<td>.16</td>
<td>.18</td>
</tr>
<tr>
<td>Juror Trust in Legal System</td>
<td>.22</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>JBS-Total</td>
<td>.16</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>JBS-PC</td>
<td>.17</td>
<td>.17</td>
<td></td>
</tr>
<tr>
<td>Juror Gender</td>
<td>.08</td>
<td>.10</td>
<td>.01</td>
</tr>
</tbody>
</table>

Note. W = White; B = Black; H = Hispanic; Adult SA = adult sexual assault; Child SA = child sexual assault; Homicide-BWS = homicide involving a defense of battered woman syndrome; JBS-Total = Juror Bias Scale (Kassin & Wrightsman, 1983) total score; JBS-PC = Juror Bias Scale - Probability of Commission subscale score; JBS-RD = Juror Bias Scale - Reasonable Doubt subscale score. All numeric values represent meta-analytic sample-weighted mean correlations ($r$).
individual-level relationship between prior juror experience and guilt judgments was very weak in our meta-analysis. Although the available data are still insufficient to draw definitive conclusions about these two juror characteristics, existing research certainly does not suggest that either plays a major role in shaping jurors’ judgments about guilt.

We also observed modest overall effects likely to be moderated by other variables for two other juror characteristics: need for cognition (NC) and gender. Scholars have generally not expected juror NC to correlate strongly with verdict preferences, noting potentially offsetting mechanisms wherein high levels of NC might make jurors more critical of the evidence presented by the prosecution, but also more willing to consider complex arguments and/or forms of evidence that support conviction. In keeping with this, researchers have often expected juror NC to interact with other variables related to the nature and complexity of the evidence. We observed a weak overall tendency for jurors with lower NC to convict more often \((r = -0.07)\). This suggests jurors with higher levels of NC may be somewhat more favorable to the defense, but indications of moderation along with a predominance of studies involving homicide make it unwise to draw any general inference about the robustness of the relationship at this point.

The overall effect from the vast empirical literature on juror gender \((r = 0.08; k = 215)\) implies a general tendency for female jurors to be slightly more likely to convict than male jurors. This conclusion does not hold up to closer scrutiny, however, in that the relationship between juror gender and guilt judgments differed considerably as a function of case type. A large proportion of the available effects \((44\%)\) were associated with cases involving adult sexual assault or child sexual abuse and, when we reran the overall analysis with all the effects except those involving child sexual abuse or adult sexual assault, the relationship essentially vanished \((r = -0.01; k = 119)\).

In addition, a follow-up meta-regression of the juror gender effects revealed that case type (coded dichotomously as sex-related v. non-sex-related crimes) explained a significant amount of variability in those effects when controlling for participant and outcome type. Thus, there does not seem to be a general propensity for female jurors to be harsher on criminal defendants than male jurors, but there is substantial support for concluding that female jurors are more likely to convict in cases involving sex crimes (particularly those against children).

Finally, several participant characteristics yielded effects on juror guilt judgments large enough to have some practical significance: defendant SES \((r = -0.11)\), defendant criminal record \((r = 0.12)\), juror authoritarianism \((r = 0.17)\), and juror trust in the legal system \((r = 0.22)\). The modest overall effect of defendant SES on guilt judgments is close to the earlier estimate by Mazzella and Feingold (1994) and reflects a tendency for jurors to convict low-SES defendants somewhat more often than high-SES defendants. Surprisingly, though, the nonsignificant \(Q\) statistic suggests the relationship is fairly robust and not moderated by other variables. This occurred despite considerable diversity with regard to how defendant SES was manipulated in the literature. This operational “noise” might be expected to cause considerable fluctuation in effect sizes across individual studies, but the aggregate data from 20 studies say otherwise, with no more variance in the distribution of effects than expected by chance. Thus, defendant SES appears to have a modest effect on guilt judgments that holds across a variety of trial contexts.

Of all 11 characteristics, defendant criminal record has perhaps the most intuitive relationship with judgments of guilt, but it has not been meta-analyzed before now (although see Tanford & Penrod, 1986, for an early form of quantitative synthesis).

Real jurors are typically prevented from learning of the defendant’s prior criminal history over concern that such information may foster the inference of a criminal disposition that leads them to convict, and several field studies support the utility of this practice in finding a moderate and positive association between jury knowledge of a defendant’s criminal record and conviction (e.g., Givelber & Farrell, 2008; Myers, 1979). Surprisingly, then, the overall effect of defendant criminal record on guilt judgments in our meta-analyses was only modest and likely to depend somewhat on other variables. One potential moderator with a compelling rationale is the degree of similarity between the prior conviction and the present charge. Jurors might be more likely to infer guilt when a defendant has been previously convicted of the same crime as opposed to some other wrongdoing; unfortunately, only a few studies to date have manipulated previous conviction similarity, making it impossible to examine as a moderator. Another potential explanation for the modest effect of prior criminal record is the relatively innocuous manner in which such information has sometimes been conveyed. In more than a few studies, the defendant’s prior criminal record was simply noted in passing with little if any further attention called to it. As such, the full impact of defendant prior criminal record may not be captured by the existing empirical literature.

Only two juror characteristics yielded overall relationships with juror guilt judgments at a magnitude greater than .15—authoritarianism and trust in the legal system as measured by the JBS (Kassin & Wrightsman, 1983). In a previous meta-analysis, Narby et al. (1993) treated the JBS as a measure of authoritarianism and included studies that used it along with those employing traditional measures of authoritarianism. We felt there was sufficient conceptual divergence to warrant distinguishing authoritarianism from trust in the legal system, but our analyses revealed considerable similarity in the profile of these two juror characteristics (i.e., a moderate relationship with guilt judgments likely influenced by one or more other variables). Further, when all JBS items were combined into a total score, the resulting predictive validity was better than either the PC or RD subscales alone, suggesting each dimension does contribute something unique to the explanation of juror verdict preference. The magnitude of the relationship between juror authoritarianism and guilt judgments also varied to some degree as a function of the authoritarianism measure used. Of particular note, the older F Scale displayed a much weaker relationship than the other measures, whereas the newer Revised Legal Attitude Questionnaire yielded the strongest association. The strength of the overall relationships, the evidence of variation by instrument, and the growing number of measures relevant to the legal system (e.g., Pretrial Juror Attitude Questionnaire, Attitudes Toward the Criminal Legal System) highlight the need for careful and informed choices when choosing study measures in this domain.
Viability of Focal Moderators

We anticipated there would be considerable variability in the distribution of effects for our focal characteristics and so coded for three potential moderator variables: (1) outcome measure, (2) participant type, and (3) case type.

Methodological moderators. Most studies of trial participant characteristics have employed undergraduate students, raising concerns about the generalizability of their findings. In light of this, we estimated relationships for several different types of study participants that vary in terms of their representativeness (i.e., students v. community members v. venirepersons) to see how much they differ. In general, the pattern of findings we observed suggests participant type does not explain a large portion of the variability in effects associated with the 11 characteristics. With respect to outcome measure, our primary interest was in comparing effects based on dichotomous judgments versus continuous ratings of guilt. In general, although the typical difference was fairly small, dichotomous verdicts were associated with slightly stronger effects than continuous guilt ratings. This finding clashes with the notion of effects involving the focal participant characteristics, and researchers do not appear to disadvantage themselves (in terms of obtaining significant results) by asking participants to decide whether the defendant is guilty or not as opposed to how certain they are of guilt.

Case type. The rationale for examining case type was that jurors may develop stereotypes of criminals that are to some extent crime-specific. We were able to compare effects across two or more case types for five of the focal characteristics: defendant gender, defendant criminal record, defendant race, juror authoritarianism, and juror gender. Although this situation precludes a definitive assessment with regard to case type acting as a general moderator, two especially noteworthy results emerged from these analyses. First, juror authoritarianism was a better predictor of guilt judgments for cases involving general homicide ($r = .20$) and the death penalty ($r = .21$) than adult sexual assault cases ($r = .08$). This may be because sexual crimes rarely involve third-party eyewitnesses and often present jurors with a difficult choice regarding the issue of consent. In such relatively ambiguous cases where police do not play prominent roles, there may simply be less opportunity for trust in the legal system to have an impact. Second, the effect of juror gender on guilt judgments varied substantially across case type. Consistent with Schutte and Hosch’s (1997) earlier meta-analysis, female jurors were more likely to convict in sex-crime cases ($r = .16$ for adult sexual assault; $r = .18$ for child sexual abuse). Conversely, women convicted less often than men in BWS cases when a female defendant was accused of killing an abusive male partner ($r = -.18$), and women preferred more lenient sentences in capital trials ($r = -.07$). For other case types (e.g., property, violent, and general homicide), juror gender was only weakly related to guilt judgments. The story model would suggest that variation in gender-related effects across case type may be due in part to systematic differences in the life experiences of men and women, with women being more receptive to stories that feature males committing crimes of a sexual nature.

Limitations and Future Research Needs

Data constraints and judgment calls. We are past the point where meta-analyses are treated as providing error-free estimates of “true” relationships simply by virtue of their integrative nature. Rather, it is now recognized that the validity of meta-analytic conclusions rests heavily on both the quantity and quality of the available data as well as the appropriateness of the procedures used by the meta-analysts (Kepes et al., 2013).

With regard to quantity, there is no consensus in the literature regarding the minimum number of effects needed for meta-analysis, but certainly “more is better” and conclusions about characteristics with a smaller number of effects should be drawn with caution. We obtained 10–30 effects for most of our focal characteristics, which we deemed sufficient for preliminary estimates. Although we searched extensively for unpublished research, our observed effects may also have been influenced by the omission of existing studies. We made a concerted effort to obtain unpublished research by contacting jury scholars and thoroughly searching electronic databases for theses and dissertations, but in such a large and multidisciplinary domain as jury decision making, some studies will inevitably be missed.

With regard to the quality of the available literature, it is important to recognize that variation in the operational definitions for focal characteristics represents a potential source of “noise” in some of our analyses. Of special note, studies featuring the manipulation of defendant characteristics employed a wide range of stimulus materials, making each manipulation somewhat unique, and some efforts were clearly more successful than others. Manipulations of defendant prior criminal record and physical attractiveness were particularly diverse and sometimes fairly weak. The consistency of operational definition was better for the juror characteristics, but even those as seemingly straightforward as prior experience and education can be conceptualized and measured differently (e.g., civil v. criminal, continuous v. dichotomized, grade level v. degree obtained). These sources of methodological noise may have truncated the observed relationships to some extent and could be responsible for some variation in the distribution of effects.

In addition to limitations imposed by the data, all meta-analyses involve judgment calls by the researchers. Common judgment calls correspond to when to call off the literature search, how to handle studies that reported no precise effect size but noted a lack of statistical significance, how to quantify effects from repeated-measures designs, and how to code for moderator variables. The most influential judgment calls, however, often involve whether to include a particular study or not. Inclusion criteria and coding schemes can appear nice and neat on paper, but every study is unique and difficult decisions are inevitable when the number of potential studies is large. Nonetheless, a “megastudy” of 196 published meta-analyses revealed little impact of 21 different types of methodological choices and judgment calls, reaffirming the robustness of meta-analytic conclusions (Aguinis, Dalton, Bosco, Pierce, & Dalton, 2011).
Future research. In light of the above, we feel there are several fruitful avenues for further examination of trial participant characteristics. First, researchers should consider manipulating case type in their studies. Case type was varied in only a small percentage of studies and often the choice of a trial stimulus case seemed perfunctory. Systematic examination of case type would benefit from development of a comprehensive taxonomy that aligns with the major types of crime-related scripts that jurors possess. Second, wherever possible, future research on participant characteristics should involve venirepersons or at least community members. This is not to denigrate the value of studies involving undergraduate students, only to call for more work assessing the generalizability of their findings. Third, future studies should involve the measurement of a battery of juror characteristics where practical. This would include one-item demographic variables (e.g., age, education, SES, race, and gender) and brief measures of psychological characteristics such as NC, authoritarianism, and legal system trust. Obtaining, analyzing, and reporting these data would greatly facilitate future meta-analytic examination. Fourth, more attention is needed regarding the conceptualization and measurement of juror attitudes toward the legal system. This broad domain includes several related and potentially overlapping constructs, including authoritarianism, dogmatism, political ideology, general trust in the legal system, and specific legal attitudes (e.g., the death penalty). Multiple measures exist for some of these constructs, and some have known psychometric problems (e.g., LAQ). It would be beneficial to examine the psychometric quality of the available measures and the degree to which they contribute incrementally to the explanation of juror verdict preferences. Finally, research is especially needed on the impact of participant characteristics at the jury level. Studies that manipulate defendant characteristics and/or jury composition on selected characteristics (e.g., race, gender, or authoritarianism) are crucial for determining when juror-level relationships are negated or exacerbated by group interaction.

Practical Implications

The preceding analyses suggest the existence of some non-trivial statistical effects associated with juror and defendant characteristics. In particular, other things being equal, variation on the part of several characteristics could alter the likelihood of an individual juror preferring conviction by up to 20%. To illustrate, assume some defendant characteristic has two levels (A and B) and correlates .20 with preference for conviction, with the A level corresponding to a 60% probability of an individual juror preferring conviction and the B level a 40% probability. In a given trial, the effect of having a defendant with the A level as opposed to the B level would be to increase the likelihood of an individual juror preferring conviction by 20%. Extrapolated to a 12-person jury, 2 to 3 more jurors would be expected to prefer conviction for an “A” defendant as opposed to a “B” defendant. That said, considerable research on social decision schemes (Davis, 1973) has shown that the function relating the number of jurors preferring conviction to the probability of a “guilty” jury verdict is nonlinear. It starts near zero with no initial advocates for conviction and rises slowly at first, increases abruptly in the middle range of adherents, and finally levels off at a very high probability once a strong majority prefers conviction (Devine, Clayton, Dunford, Seying, & Pryce, 2001). An additional 2 to 3 “proconviction” jurors would therefore only have a noteworthy impact on the likelihood of a guilty verdict when there would otherwise be a fairly even split in the initial distribution of verdict preferences. In essence, when either a few—or most—jurors prefer conviction, the impact of adding (or subtracting) a couple should be negligible. Thus, under the most favorable circumstances, juror and defendant characteristics that correlate even moderately well with individual guilt judgments might meaningfully affect the odds of conviction in a relatively small number of “close” cases. In practice, where defendant characteristics will not always be perceived (or perceived uniformly), and attorneys are not free to assemble a jury composed entirely of jurors with the “desired” characteristic, opportunities for participant characteristics to play a decisive role would seem to be even rarer.

Of course, one can acknowledge the likely modest and irregular influence of participant characteristics but still conclude it is too much. Based on the magnitude of their overall or case-specific effects, several participant characteristics arguably warrant the attention of attorneys (or trial consultants) during voir dire in some trials. Specifically, we feel these would generally include two defendant characteristics (SES and prior criminal record), two juror characteristics (authoritarianism and legal system trust), and the race of both jurors and defendants considered jointly. In addition, juror gender appears to be relevant in cases involving domestic homicide and/or child or female victims. These non-negligible effects raise the question of what can be done to limit or further reduce the number of occasions when participant characteristics affect jury verdicts. Several procedural options are possible.

One would be to increase efforts to seat a jury that is maximally diverse with respect to focal venireperson characteristics. The goal would be to ameliorate extralegal bias by diluting the concentration of any given characteristic in the jury pool. As scientists have long known, random selection provides the best overall safeguard against this kind of bias. In practice then, this option would involve taking concerted steps to: (a) improve the comprehensiveness of juror-eligible population lists, (b) ensure that a truly random sample of community members is contacted via the summons process, and (c) increase the “show-up” rate of those called for jury duty.

A second option would be to allow—or even promote—the systematic assessment of selected venireperson characteristics so that information could be used by both sides during voir dire. Most focal psychological characteristics could be measured quickly and easily using brief questionnaires returned by mail or completed online. Alternatively, standardized court-approved sets of questions could be developed and used during voir dire. Attorneys could then use this information to challenge venirepersons for cause (if allowed by informed judges) or via peremptory challenge. Of course, peremptory challenges cannot presently be used to strike individuals on the basis of race or gender, a prohibition that creates a dilemma in light of our findings that juror race and gender are sometimes nontrivially associated with guilt judgments.

A third option would be to limit or delay jurors’ knowledge/awareness of selected personal characteristics of the defendant. This is done routinely for prior criminal record but could be extended to other defendant characteristics such as SES, gender, and physical appearance. For characteristics that are visual in
nature, webcams or closed-circuit TV could be used to allow defendants to monitor or participate in trials, perhaps even testify, without revealing personal characteristics that could influence jurors. As such, at least three options could be employed to limit bias stemming from trial participant characteristics: maximize the diversity of the jury pool, systematically measure and use information on relevant venireperson characteristics, and limit/suppress jurors’ awareness of defendant personal characteristics. There are likely other options worthy of consideration as well.

Conclusion

This study conveys the results of a meta-analytic integration of the large empirical literature on 11 trial participant characteristics, six associated with jurors and five associated with defendants. In general, our results are reminiscent of the proverbial partially filled “glass.” From the “half-full” perspective, most of our focal participant characteristics were associated with weak-modest overall effects likely to be moderated by one or more other variables. Observed effects varied only modestly as a function of the nature of the outcome measure, somewhat more with regard to type of participant, and quite notably by case type. Nonetheless, given the size of typical effects, extra-legal bias associated with these characteristics should rarely be decisive with regard to jury verdicts. From the “half-empty” perspective, several participant characteristics exhibited relationships with guilt judgments large enough to warrant the attention of jury scholars and legal practitioners. These would include defendant prior criminal record, defendant SES, defendant race (in conjunction with juror race), juror authoritarianism, juror legal system trust, and juror gender in some cases (e.g., those involving sex-related crimes). Overall then, the most appropriate answer to the question of whether trial participant characteristics matter would seem to be that some of them do to a modest extent—and more so in some cases than others. Regardless of which perspective one is inclined to take, the results of this research contribute to our understanding of the impact of juror and defendant personal characteristics at trial and inform ongoing efforts to assess and improve the decisions of juries.

References

*References marked with an asterisk indicated studies included in the meta-analysis.


Hosch, H. M., Culhane, S. E., Jolly, K. W., Chavez, R. M., & Shaw, L. H. (2011). Effects of an alibi witness’s relationship to the defendant on the


Nadler, J. (2000). The effects of perceived injustice on deference to the law (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign, Urbana-Champaign, IL.


META-ANALYSIS OF INDIVIDUAL CHARACTERISTICS

133

doctoral dissertation). University of Illinois at Urbana-Champaign, Urbana-Champaign, IL.


Yanchar, N. V. (1982). Subject-juror decision-making in rape cases: Effects of status of the complainant, and gender of the defense attorney, the prosecuting attorney, and the subject-juror (Unpublished doctoral dissertation). Bowling Green State University, Bowling Green, OH.

Received September 9, 2013
Revision received January 5, 2014
Accepted January 7, 2014