

LEGITIMIZING CHARACTER EVIDENCE

*Justin Sevier**

ABSTRACT

Modern consensus among legal commentators is that character evidence—when used to show that an individual behaved in accordance with her predisposition to commit some act—is an illegitimate form of fact-finding proof. This consensus is codified in the Federal Rules of Evidence, which forbids the use of most “propensity evidence” at trial. Defenders of the ban suggest, without empirical proof, that jurors would overvalue the probative worth of propensity evidence and that the public would balk at the inclusion of such evidence as a matter of legal procedure. This Article suggests that this view is misguided, its assumptions are incorrect, and that policymakers should consider lifting the ban on propensity evidence.

This Article reports the results of three original experiments, which examine the conditions under which the public is willing to legitimize legal verdicts that rely on propensity evidence. The psychological literature suggests that two elements must be satisfied for the public to legitimize an evidentiary rule: (1) the public must perceive the rule as promoting decisional accuracy, such that it increases the likelihood that the fact finder reaches the correct verdict, and (2) the rule must promote “procedural justice,” such that people believe that the fact finder has reached its decision according to notions of fair play. Social psychology research on “person perception” suggests that jurors are more competent to evaluate character evidence than legal commentators believe, and research on procedural justice suggests that the inclusion of propensity evidence may increase the popular legitimacy of legal verdicts.

These experiments, which surveyed over 1,200 participants, support the position that propensity evidence is a legitimate form of trial proof. They demonstrate that jurors attend to propensity evidence when it is presented to them, but they afford such evidence significantly less weight than they do most other evidence at trial. Moreover, jurors demonstrate marked competency with propensity evidence: they discriminate between potential accuracy-enhancing

* Charles W. Ehrhardt Professor of Litigation, Florida State University College of Law. I thank Shawn Bayern, Jeffrey Bellin, Avlana Eisenberg, John C. P. Goldberg, Joni Hersch, Mark Spottswood, Tom R. Tyler, Kip Viscusi, Brandi Yoder, Bryce Yoder, the Florida State University College of Law faculty, and the Yale University Department of Psychology for comments regarding this Article. I also thank Elise Berry, Conor Burns, and Jared Dubosar for their excellent research assistance.

and accuracy-diminishing features of such evidence, including the frequency of the defendant's behavior, the duration of the acts, and the similarity of those acts to the act of which the defendant is accused. Finally, these studies suggest that propensity evidence increases—not decreases—the public's perceptions of procedural fairness at trial. These findings have substantial implications for evidential policy and for attorneys who make ground-level decisions regarding the use of character evidence.

INTRODUCTION	443
I. THE LEGAL RISE (& FALL) OF CHARACTER EVIDENCE	447
A. <i>History and Development</i>	447
B. <i>Modern Character Evidence</i>	451
II. THE PSYCHOLOGICAL LEGITIMACY OF CHARACTER EVIDENCE	456
A. <i>Decisional Accuracy: Impression Formation</i>	458
1. <i>Cause for Concern: Overreliance on Personality Traits</i> ...	459
2. <i>Cause for Optimism: The Interactionist Model</i>	461
B. <i>Procedural Justice: Fair Process</i>	462
III. THREE EXPERIMENTS	464
A. <i>Study 1: The Power of Propensity</i>	465
1. <i>Participants in Study 1</i>	466
2. <i>Procedure and Measures in Study 1</i>	466
3. <i>Results of Study 1</i>	470
a. <i>Main Analysis I: Global Attitudes in Study 1</i>	471
b. <i>Main Analysis II: Specific Judgments in Study 1</i>	477
4. <i>Discussion of Findings in Study 1</i>	481
B. <i>Study 2: Testing Decisional Accuracy</i>	482
1. <i>Participants, Procedures, and Measures in Study 2</i>	483
2. <i>Results of Study 2</i>	486
a. <i>Main Analysis of Results in Study 2</i>	486
b. <i>Serial Mediation Analysis of Results in Study 2</i>	488
3. <i>Discussion</i>	491
C. <i>Study 3: Propensity & Procedural Justice</i>	492
1. <i>Participants, Procedures, & Measures in Study 3</i>	493
2. <i>Results of Study 3</i>	495
a. <i>Main Analysis of Results in Study 3</i>	496
b. <i>Path Analysis of Results in Study 3</i>	499
3. <i>Discussion of Results in Study 3</i>	501
IV. IMPLICATIONS AND OBJECTIONS	502
CONCLUSION	507

INTRODUCTION

*“Propensity (n): An often intense natural inclination or preference.”*¹

On the eve of the Megalesia festival in 56 B.C., renowned Roman attorney and orator Cicero faced the magistrate Gnaeus Domitius, along with a gathered crowd, at the *Quaestio de vi*.² Cicero was serving in his capacity as defense attorney for his former student and political adversary, Marcus Caelius Rufus, who had been accused of political violence (known as “*vis*”), the most serious crime in the Roman Republic.³

The charges against Caelius stemmed from the murder of an Alexandrian ambassador who became ensnared in the efforts of King Ptolemy XII of Egypt to recover his throne after he was deposed.⁴ During that time, Caelius had begun a torrid affair with Clodia Pulchra, a recently widowed woman known in Rome for her gambling, drinking, and penchant for sexual scandals.⁵ When Caelius ended the affair, Clodia and her brother reportedly swore revenge against

¹ *Propensity*, MERRIAM-WEBSTER, <https://www.merriam-webster.com/dictionary/propensity> (last visited Nov. 25, 2018). The Merriam-Webster online thesaurus also describes the term as “an established pattern of behavior[.]” “a habitual attraction to some activity or thing[.]” and “aptness.” *See id.*

² *See* T. A. Dorey, *Cicero, Clodia, and Pro Caelio*, in 5 GREECE AND ROME 175, 175 (1958). A *Quaestio de vi* was a specialized commission in the Roman Republic in which a magistrate investigated a criminal matter and reported those findings to the Senate. *See Quaestio*, LECTIC L. LIBR., <https://www.lectlaw.com/def2/q074.htm> (last visited Nov. 25, 2018); *see also* T. COREY BRENNAN, THE PRAETORSHIP IN THE ROMAN REPUBLIC: VOLUME 2: 122 TO 49 B.C., at 439 (2000) (explaining the different Roman courts). The Megalesia festival occurred annually in Ancient Rome from April 4th through April 10th in celebration of Cybele, the mother goddess. *See* Michele Renee Salzman, *The Representation of April in the Calendar of 354*, 88 AM. J. OF ARCHAEOLOGY 43, 47 (1984). The festival included chariot races in the Circus Maximus, religious plays, and displays of wealth by the patrician class. *See, e.g.*, EUGENE N. LANE, CYBELE, ATTIS, AND RELATED CULTS: ESSAYS IN MEMORY OF M.J. VERMASEREN 393–94 (1996); *see also* LYNN E. ROLLER, IN SEARCH OF GOD THE MOTHER: THE CULT OF ANATOLIAN CYBELE 1 (1999).

³ *See* MARCUS TULLIUS CICERO, TEN SPEECHES 187 (James E. G. Zetzel trans., 2009) (discussing the background of the trial and characterizing *vis* as “seditious violence”).

⁴ Tamás Nótári, *Law on Stage—Forensic Tactics in the Trial of Marcus Caelius Rufus*, in 51 ACTA JURIDICA HUNGARICA 199 (2010) (describing the background of the trial). For further background on the deposition of Ptolemy and his restoration (and the life of his daughter, Cleopatra), *see* ERNLE BRADFORD, CLASSIC BIOGRAPHY: CLEOPATRA 28 (Penguin Books 2000) (1971) (discussing the battle wherein King Ptolemy XII defeated the Egyptian frontier forces and regained control of the Alexandrian palace).

⁵ Not much is known about Clodia beyond her characterization in Cicero’s defense of Caelius at trial, but historians suspect that Cicero’s contemporaries had written about her under different names. *See, e.g.*, SUZANNE DIXON, READING ROMAN WOMEN 133–56 (2001) (discussing how Clodia might also be the woman known as Lesbia, the frequently unfaithful woman in the poet Catullus’s love poems). Some historians have disputed these characterizations of Clodia. *Id.*

Caelius.⁶ They later accused him of conspiring to kill the Alexandrian ambassador with funds given to him by Clodia, and attempting to poison her to prevent her from sharing with others his alleged misdeeds.⁷

The two-day trial began on April 3rd with several prosecution witnesses attacking Caelius's moral character, but providing little evidence supporting the allegations against him.⁸ On the trial's second day, Cicero began his defense and gave his famed *Pro Caelio* speech, regarded as one of the best known examples of oratory in Roman history.⁹ In *Pro Caelio*, Cicero took aim at Clodia and her accusations against Caelius. After stating that the prosecution provided few facts supporting its theory of the death of the Alexandrian ambassador, he attacked Clodia's character, specifically her propensity to commit illicit and uncouth acts:

The whole accusation emanates from a house that is malevolent, disreputable, crime-stained and vicious. Whereas the family alleged to have been involved in this shocking deed is notable for its lofty standards, honourable principles, dutifulness and sense of responsibility; and that is the home from which you just heard a sworn affidavit. The question under dispute, therefore, is easy to settle. You are invited to say whether you do not agree that the parties who confront one another are, on the one side, an unstable, evil-tempered nymphomaniac, who has completely fabricated the charge, and on the other, a man of responsibility, wisdom and self-restraint whose evidence has shown the utmost conscientiousness and accuracy.¹⁰

In addition to its oratorical flourishes, the speech is notable for another reason: it is one of the earliest (and perhaps best known) uses of character evidence—specifically propensity evidence—at trial. In describing Clodia's personality as “crime-stained” and “evil-tempered,” Cicero asked the fact finder

⁶ See Nótári, *supra* note 4, at 198–204 (explaining the complex web of events that gave rise to the trial of Caelius and providing a detailed history of the animosity between Cicero himself and Clodia and her brother, stemming from a prior legal proceeding in which they were involved).

⁷ *Id.*

⁸ See CICERO, *supra* note 3, at 205–06 (“There are two charges. One involves gold, the other poison; in both of them one and the same person is concerned. The gold was borrowed from Clodia, the poison was sought to give to Clodia—or so they say. All the rest are not charges but slanders; they belong to a violent quarrel rather than a public court. ‘Adulterer, degenerate, graft-giver.’ That’s brawling, not prosecution. There’s no foundation for these charges, no basis. They’re fighting words thrown out hit or miss by an angry prosecutor with no evidence.”).

⁹ See *id.* at 193–227 (commenting on the speech and including annotations and contextual footnotes).

¹⁰ SELECTED POLITICAL SPEECHES OF CICERO 294 (Michael Grant trans., 1969) [hereinafter GRANT]. Perhaps intending humor, Cicero preceded this quote with the following: “And now I see the origin of a great hatred, with a really vicious breakup. In this case, members of the jury, our whole dispute is with Clodia, a lady not only prosperous but promiscuous—but I won’t say anything about her except to rebut the charges.” CICERO, *supra* note 3, at 206.

to disbelieve Clodia's allegations of murder and poisoning by virtue of her prevailing personality traits.¹¹ Cicero emerged victorious and Caelius was acquitted of the charges against him.¹²

Character evidence embodies the constellation of an individual's acts that indicate her underlying personality traits.¹³ The evidential use of such underlying traits has a storied, complex role in trials throughout history. Until recently, courts welcomed such evidence; a defendant's bad or illicit acts were viewed as circumstantial evidence of a morally bankrupt character, and it was on that basis that the criminal law would punish the defendant for his or her misdeeds.¹⁴ Indeed, courts routinely admitted testimony from character witnesses—including testimony regarding an individual's propensity to commit a relevant act—into evidence at trial.¹⁵ Thus, proceedings such as the Caelius trial were commonplace. It was only in the late nineteenth century that common law courts, responding to cultural changes inspired by egalitarian norms and Enlightenment thought, began to ban certain types of propensity evidence in determining a defendant's criminal or civil liability.¹⁶

This Article argues, with original empirical data, that the ban on propensity evidence leads to a counterintuitive result: it makes the public *less* willing to legitimize legal proceedings. This is so because verdicts are perceived to be *more* accurate if propensity evidence is presented to a fact finder and because the public views the admission of propensity evidence as *more* procedurally fair than when the propensity evidence is excluded.

¹¹ GRANT, *supra* note 10.

¹² See MARCUS TULLIUS CICERO, *CICERO: DEFENCE SPEECHES* 124 (D. H. Berry trans., 2000) (“*Pro Caelio*” chapter).

¹³ We frequently—and often automatically—form impressions of others and make judgments about their character traits in many aspects of our personal and professional lives. Moreover, we extend those character judgments to an individual's behavior, by attributing the former as the cause of the latter. Social science evidence suggests that, in everyday life, these implicit character judgments often serve us well in determining with whom we should associate and whom we should avoid. See *infra* Section II.A.

¹⁴ See, e.g., Claire Finkelstein, *Excuses and Dispositions in Criminal Law*, 6 BUFF. CRIM. L. REV. 317, 317–21 (2002) (discussing the “traditional view” of criminal law that is said to focus exclusively on acts instead of character and noting scholars in recent years have challenged that view that character has had no role to play in the meting out of justice under the criminal law).

¹⁵ For a more thorough discussion of this point, see *infra* Section I.A. explaining the development of character evidence in the courts.

¹⁶ The change in the character evidence rule came during an era in which courts were constraining the expansive power of the modern jury in many ways, including through the regulation of the factual inputs that juries received in reaching a verdict. The courts made lofty, well-intentioned (if tautological) pronouncements that individuals should be judged by their proven behavior, not by the content of their character, and worried that the admission of propensity evidence would make trials less accurate—because juries would overvalue the evidence—and less legitimate as a procedural matter. See *infra* notes 36–50 and accompanying text.

Legal scholars have, from time to time, called for propensity evidence to be admissible under the Federal Rules of Evidence. Columbia Law School Professor Richard Uviller wrote perhaps the best known defense of such evidence many years ago in the *University of Pennsylvania Law Review*.¹⁷ His article, however, contained two critical limitations. First, Professor Uviller did not provide a behavioral theory, backed by social science research, to explain why the use of propensity inferences at trial is a legitimacy-enhancing evidential innovation. Second, he did not provide any empirical data regarding how jurors actually evaluate propensity evidence.

This Article fills that gap. It provides a social psychological theory for how jurors evaluate character evidence and presents data from three original experiments—surveying over 1,200 participants—examining how jurors weigh such evidence and its effects on popular legitimacy. Its findings provide support for lifting the bar on propensity evidence at trial. In barring propensity evidence at trial, common law courts subscribed to an outmoded theory of human behavior, whereby people ascribe others' acts almost entirely to their personality traits. Social psychology research suggests the opposite: the manner in which people form impressions of others is *interactional*, such that they evaluate behavior in light of a person's personality traits and in light of *situational* factors that induced the behavior. This interactionist model suggests that, given the correct tools, jurors evaluate propensity evidence carefully and defensibly.

This Article reports several findings from our original studies.¹⁸ First, we found that jurors are attentive to propensity evidence at trial.¹⁹ They do not, however, weigh the evidence more heavily than other evidence presented at trial, and they carefully evaluate features of character evidence that increase or decrease its diagnosticity with respect to the defendant's legal liability.²⁰ Moreover, the studies show that the public has less confidence in verdicts—and

¹⁷ H. Richard Uviller, *Evidence of Character to Prove Conduct: Illusion, Illogic, and Injustice in the Courtroom*, 130 U. PA. L. REV. 845, 890 (1982) (“Yet today, character evidence most often appears either in burlesque of its function, or as a product of an arcane legalistic wordplay, or as a cruel and senseless shard of forgotten dogma. It is foolish to exclude helpful evidence simply because it tends to prove the fact by proving predisposition to perform it. Relevant is relevant.”). Professor Uviller expressed optimism for a better-constructed character evidence rule while calling the federal rules a “poor example” of good drafting. *Id.* at 891.

¹⁸ The arguments and claims in this Article are the author's own. The word “we” is used throughout to acknowledge the work of the research assistants and others who assisted the author in designing the study and interpreting the results.

¹⁹ See *infra* Section III.A.

²⁰ See *infra* Section III.B.

is less likely to *legitimize* those verdicts—if courts shield propensity evidence from legal fact finders.²¹

This Article proceeds in several parts. Part I provides a brief history of the use of character evidence at trial. It then provides an overview of the current character evidence regime under Articles IV and VI of the Federal Rules of Evidence. Part II examines the applicability of social psychological theory to the evaluation of character evidence. This Part also examines the psychological literature on impression formation from the perspective of the interactionist model, and it examines the circumstances under which the public psychologically legitimizes the decisions of legal actors. Part III presents the results of three original experiments that suggest that the rule barring propensity evidence at trial should be reevaluated. Part IV explores the policy implications of these findings, their limitations, and the future directions of these findings for the law of evidence.

I. THE LEGAL RISE (& FALL) OF CHARACTER EVIDENCE

Part I of this Article briefly defines the term “character evidence” and describes the historical development of the doctrine. It then provides a snapshot of the current state of the doctrine in American courts.

A. *History and Development*

Despite its storied history and the dizzying, “grotesque” array of rules surrounding its application,²² the Federal Rules of Evidence do not formally define character evidence.²³ Legal psychologists Michael Saks and Barbara Spellman have defined it as evidence that is “roughly equivalent to what people think of as ‘the kind of person’ someone is, the set of a person’s traits, one’s

²¹ See *infra* Section III.C.

²² *Michelson v. United States*, 335 U.S. 469, 486 (1948) (“To pull one misshapen stone out of the grotesque structure is more likely simply to upset its present balance between adverse interests than to establish a rational edifice.”). Notably, although the Supreme Court affirmed the prohibition on propensity evidence as circumstantial evidence of a defendant’s illicit act, the Court was profoundly (and candidly) critical of the doctrine: “We end, as we began, with the observation that the law regulating the offering and testing of character testimony may merit many criticisms We concur in the general opinion of courts, textwriters and the profession that much of this law is archaic, paradoxical and full of compromises and compensations by which an irrational advantage to one side is offset by a poorly reasoned counter-privilege to the other.” *Id.* at 485–86.

²³ See MICHAEL J. SAKS & BARBARA A. SPELLMAN, *THE PSYCHOLOGICAL FOUNDATIONS OF EVIDENCE* LAW 143, 302–03 n.2 (2016) (quoting a state supreme court justice stating, in *State v. Williams*, 874 P.2d 12, 25 (N.M. 1994) (Montgomery, C.J., concurring), “I am unable to do what all the text-writers and other legal authorities have failed to do. I am unable to outline the contours of the term ‘character.’”).

personality characteristics or psychological attributes.”²⁴ Even today, what constitutes a character trait for purposes of evidence law sometimes is hotly contested,²⁵ although some elements of character have been litigated so often that they are settled, such as honesty, violence, peacefulness, and temperance.²⁶

As cultural understandings of the meaning of “character” and its social value changed over time, its use as evidence at trial varied as well. The first historical character witnesses can be found in the compurgation procedure advanced in the Middle Ages.²⁷ Under the compurgation process, sometimes called “trial by oath” or “wager of law,” a defendant could establish his innocence or non-liability by bringing to the trial a required number of persons, typically twelve, to swear that they believed the defendant’s oath of innocence.²⁸ This process was considered an advancement from the wager of battle and trial by ordeal that preceded it—insofar as it provided the defendant a form of agency in her defense that did not rely on the luck of battle or the uncertain results of the tasks undertaken at the ordeal.²⁹ And although it also represented a step toward the modern conception of a trial, whereby evidence is gathered in an effort to establish historical truth underlying a legal dispute, it also was a vehicle by which social status and hierarchy could be maintained. It was usually easier for wealthier, powerful members of society to convince twelve compurgators to come to court and swear an oath on behalf of the accused.³⁰

Compurgation also had the effect of cementing, at least through the eighteenth century, the importance and prominence of the use of character

²⁴ *Id.* at 143. They based this definition on the writings of several other evidence scholars. They note that John Henry Wigmore described character as equivalent to disposition, “with a fixed trait or the sum of traits.” *Id.* In his highly regarded treatise, Charles T. McCormick described character as “a generalized description of one’s disposition, or of one’s disposition in respect to a general trait, such as honesty, temperance, or peacefulness.” *Id.*

²⁵ *See, e.g.*, ROGER C. PARK ET AL., EVIDENCE LAW: A STUDENT’S GUIDE TO THE LAW OF EVIDENCE AS APPLIED IN AMERICAN TRIALS 127–28 (3d ed. 2011) (noting the textual ambiguities in the current rule and postulating that “[t]o constitute a character trait, one would think (though this is not settled) that the tendency must arise in some reasonable degree from the person’s *moral* being—from traits over which the person has a substantial element of choice . . .).

²⁶ *Id.* For a rich description of the definition of character evidence, and the social values that inhere in that definition, see Daniel D. Blinka, *Character, Liberalism, and the Protean Culture of Evidence Law*, 37 SEATTLE U. L. REV. 87 (2013) (describing famous cases involving character evidence, providing a history of the doctrine’s evolution, and discussing the doctrinal incongruities within the current doctrine).

²⁷ *See, e.g.*, LAWRENCE M. FRIEDMAN, THE LEGAL SYSTEM: A SOCIAL SCIENCE PERSPECTIVE 272 (1975) (discussing the rise and fall of the “wager of law” in Medieval England).

²⁸ *See* Blinka, *supra* note 26, at 130–32 (discussing the compurgation process).

²⁹ *Id.* The compurgation procedure should not be mistaken for the modern trial process, however. If the tribunal found the defendant guilty, all compurgators could be put to death as well. *Id.*

³⁰ *See* NEIL VIDMAR & VALERIE P. HANS, AMERICAN JURIES: THE VERDICT 21–65 (2007) (discussing the evolution of the jury and the specifics of compurgation).

evidence at trial. Many post-Medieval societies moved away from wagers of battle and trials by ordeal as dispute resolution mechanisms, and they moved instead toward a rudimentary jury system.³¹ Unlike the modern American jury, jurors in these eighteenth century systems often were selected because they had at least some knowledge of the defendant's character and reputation in the community.³² Trials often occurred at breakneck speed, whereby deliberations often took no more than a few minutes with jurors remaining in the courtroom and huddling together to discuss their views of the defendant.³³ Jurors were afforded substantial power and discretion in rendering their trial verdicts, not so that they could methodically ascertain the historical truth underlying a legal dispute, but instead to decide which few defendants would merit a guilty verdict (and a mandatory death sentence) to maintain the social order.³⁴ Evaluating the defendant's character was a central feature of those determinations.³⁵

Major changes occurred, however, in the nineteenth and twentieth centuries. These changes were cultural and structural, and both influenced the decline of character evidence in the modern American trial. Culturally, the Industrial Revolution fundamentally altered the manner in which members of the public interacted with each other, particularly those in different social classes.³⁶ Most critically, the move from an agrarian economy to a more industrial economy promoted the rise of modern cities in which individuals interacted frequently, at the personal and professional level, with people whom they did not know well or at all.³⁷ In societies in which it was commonplace for people to engage in near-anonymous interactions with strangers for goods and services necessary for survival, it was no longer practical for the American legal system to depend so dramatically on a fact finder's knowledge of the defendant's character.³⁸

³¹ See Stephan Landsman & James F. Holderman, *The Evolution of the Jury Trial in America*, 37 LITIG. 32, 32–35 (2010) (detailing the history of the jury system).

³² See VIDMAR & HANS, *supra* note 30 (discussing the requirements of jury service); Landsman & Holderman, *supra* note 31 (same).

³³ VIDMAR & HANS, *supra* note 30, at 50–51. Indeed, the jury box was invented in part to make this process easier. See Blinka, *supra* note 26, at 120.

³⁴ Blinka, *supra* note 26, at 120–21 (expounding on this counterintuitive theory).

³⁵ See David P. Leonard, *In Defense of the Character Evidence Prohibition: Foundations of the Rule Against Trial by Character*, 73 IND. L.J. 1161, 1194, 1196 (1998) (characterizing trials of that era as “a character-based exercise”); see also Blinka, *supra* note 26, at 130 (noting that “the older-style trial . . . placed a premium on a person’s character”).

³⁶ See Leonard, *supra* note 35, at 1196; see also Blinka, *supra* note 26, at 124 (noting that the Industrial Revolution “catalyzed profound social changes”).

³⁷ See Leonard, *supra* note 35, at 1196.

³⁸ *Id.*

At the same time, structural changes in the American legal system threatened the continued viability of “trial by character.” Egalitarian norms and Enlightenment-inspired attitudes toward the purpose of the modern trial began to take root, and each contributed to a shift in the perceived value of character evidence.³⁹ With respect to egalitarian norms, some scholars have suggested that popular attitudes toward a person’s “character” shifted in the nineteenth century.⁴⁰ In an era of social mobility—in part inspired by the Industrial Revolution—people began to view character traits not as immutable fibers of a person, but instead as malleable traits that could be developed to improve one’s social standing.⁴¹ Thus, if character traits are malleable, the nineteenth century American public might have viewed evidence of a person’s character as less useful at trial.⁴²

Enlightenment-inspired thought, and the emergence of empirical social sciences in the twentieth century, also contributed to different conceptions of the nature of legal trials and the value of character evidence.⁴³ Enlightenment thinkers viewed trials not as judgments of a defendant’s moral value, but instead as a scientific search for the historical, objective truth at the heart of the legal dispute.⁴⁴ To better achieve this search for historical truth, several procedural reforms occurred with respect to the modern trial. At the micro level, formal evidential rules were employed and witnesses were required to have firsthand knowledge of the facts to which they testified.⁴⁵ At a macro level, the role and power of the jury began to decrease,⁴⁶ while the role and power of the presiding trial judge increased.⁴⁷ Concomitantly, influential legal scholars began authoring treatises in which they commented on the perceived decline in the importance of character evidence at trial.⁴⁸ Unsurprisingly, stricter common law rules followed, which limited substantially the types of character evidence that juries could consider.

³⁹ *Id.* at 1195–96.

⁴⁰ *See* Blinka, *supra* note 26, at 123–29.

⁴¹ *Id.* at 124.

⁴² *Id.* at 129.

⁴³ *Id.* at 132–33.

⁴⁴ *See* Leonard, *supra* note 35, at 1194–95.

⁴⁵ *Id.*

⁴⁶ *See* Paul Butler, *In Defense of Jury Nullification*, 31 LITIG. 46, 47 (2004) (discussing the history of the jury in the context of its power to refuse to convict guilty defendants).

⁴⁷ *See* VIDMAR & HANS, *supra* note 30, at 41–64 (explaining what they characterize as a tug-of-war between the power of the judge and the jury as trials have evolved).

⁴⁸ *See* Blinka, *supra* note 26, at 129 (discussing Simon Greenleaf’s 1842 evidence treatise in particular as a contributing factor to this phenomenon).

In sum, early trials were often solely judgments of a defendant's character. But cultural and legal developments in the modern era resulted in trials that were perceived as more "truth focused" than trials of the previous era, and evidence of a person's character became, to varying degrees, less helpful in the truth-seeking endeavor.⁴⁹ Against that background, the modern character evidence regime took hold, explicitly endorsed in the Supreme Court's decision in *Michelson v. United States* and later embodied in the Federal Rules of Evidence in 1975.⁵⁰

B. Modern Character Evidence

Against this historical background, the modern doctrine encompasses a morass of rules under Articles IV and VI of the Federal Rules of Evidence that create a labyrinthine structure for admitting or excluding character evidence.⁵¹ These confusing rules address not only the substantive proscriptions involving the use of character evidence, but also the procedural hurdles that parties must overcome when character evidence is admissible.⁵²

Perhaps the most surprising aspect of the rule barring character evidence is how *little* character evidence the rule actually bars. Federal Rule of Evidence 404(a) bars character evidence in civil and criminal trials only when a party proffers the evidence to prove that another party—or someone associated with another party—had a propensity to act in a certain manner and therefore acted in conformity with that propensity.⁵³

⁴⁹ *Id.*

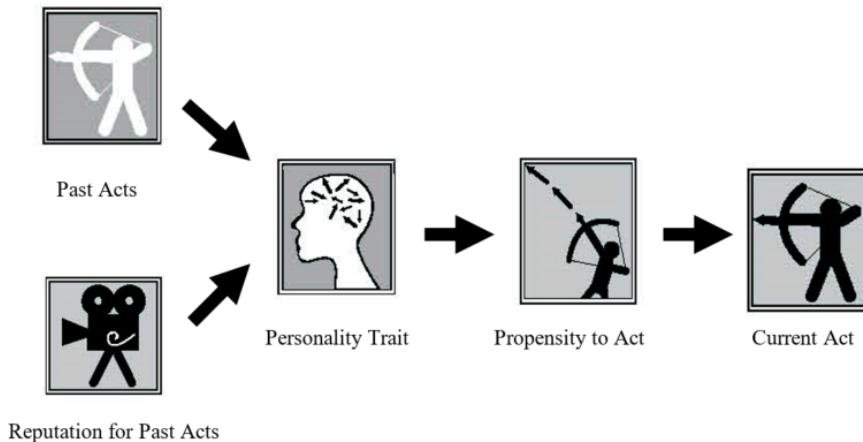
⁵⁰ See 335 U.S. 469, 475–77 (1948); see also FED. R. EVID. 404(a) (explaining the bar against using propensity evidence as proof at trial).

⁵¹ For a discussion of rules bearing on character evidence, see FED. R. EVID. 404 (discussing its substantive import), 405 (discussing its procedural requirements), 406 (distinguishing habit from character), 412 (involving its role in rape cases), 413–15 (discussing its role in civil and criminal sexual assault and molestation cases), and 608 (discussing its role in impeaching a witness). It also appears obliquely in Federal Rule of Evidence (FRE) 803, which establishes a hearsay exception for admissible reputation evidence of a party's character. FED. R. EVID. 803.

⁵² Section I.B. will address the substantive aspects of character evidence only: FRE 405 (and an analogous provision in FRE 608(b)) lays out the procedures governing the form of admissible propensity evidence. Mainly on account of judicial economy, FRE 405 distinguishes between (1) reputation and opinion evidence, which is admissible on direct examination and on cross-examination, and (2) specific acts indicative of an individual's character, which are admissible only on cross-examination. The Rules relax this prohibition on specific act testimony when the evidence is used for a non-propensity purpose or when it involves acts of sexual misconduct pursuant to FRE 413–15. See *infra* Section I.B.

⁵³ See, e.g., *United States v. Lukashov*, 694 F.3d 1107, 1118 (9th Cir. 2012) (affirming the lower court's decision to exclude bad character evidence since that evidence would have "been asking the jury to engage in propensity reasoning"); *Huddleston v. United States*, 485 U.S. 681, 686 (1988) (explaining that before admitting character evidence, FRE 404 demands a court to establish that the evidence is "probative of a material issue

In other words, American courts ban character evidence only if it is proffered to show that a party has a certain undesirable character trait, and *because of that character trait*, the party committed an act relevant to the cause of action. In an important text on the Federal Rules of Evidence, Professors Deborah Merritt and Ric Simmons illustrate this impermissible “propensity inference” visually as follows:⁵⁴



The Federal Rules do not proscribe, however, a party from using character evidence (or evidence that *appears* to be character evidence) for any *other* purpose in court. Indeed, there are five different categories of evidence that appear to be impermissible propensity evidence but are, in fact, admissible in civil and criminal trials: two of which we can conceive of as “exemptions” to the rule barring character evidence, and three of which we can conceive of as “exceptions” to the rule.

Two categories of admissible character evidence are “exemptions” to the propensity bar. Character evidence is forbidden under FRE 404(a) only if the evidence requires the fact finder to reach the end of the impermissible propensity inference chain illustrated by Professors Merritt and Simmons.⁵⁵ If the evidence does not require the fact finder to reach the end of the diagram, then it is deemed

other than character”); *see also* United States v. Canady, 578 F.3d 665, 670–71 (7th Cir. 2009) (describing the analysis a trial court should conduct when determining whether bad character evidence should be admitted).

⁵⁴ Figures substantially similar to the figure above appear in Deborah Merritt & Ric Simmons, *Learning Evidence: From the Federal Rules to the Courtroom* 297, 299, 302 (3d ed. 2017).

⁵⁵ *See generally* FED. R. EVID. 404(a)(1) (“Evidence of a person’s character or character trait is not admissible to prove that on a particular occasion the person acted in accordance with the character or trait.”).

“non-propensity” character evidence and potentially is admissible.⁵⁶ This occurs when a party’s character is directly at issue in the litigation, and when the evidence is proffered as circumstantial evidence of some other fact in the litigation.⁵⁷

With respect to the former, sometimes a party’s character is directly at issue in the litigation because of the substantive law underlying the case. When a party’s character is directly at issue, the party attempts only to prove the character trait itself, not that the other party acted in *conformity* with that trait.⁵⁸ For example, in a defamation case, a famous celebrity may seek to prove that she has a generous character to support a claim that a magazine defamed her by calling her a cheapskate. Other common examples include the personality traits of parents in child custody disputes, an individual’s reputation when a defendant is sued civilly for negligent entrustment or negligent hiring, and claims of entrapment by law enforcement officials.⁵⁹ In these scenarios the character trait itself—and not the behavior that flows from the character trait—is the relevant fact in the litigation.⁶⁰

Regarding the latter, sometimes circumstantial evidence of another relevant fact masquerades as inadmissible propensity evidence. As the Federal Rules of Evidence clarify, in these circumstances, if there is a non-propensity purpose for the use of the evidence (even if it *also* could be proffered for propensity purposes), the evidence is potentially admissible.⁶¹ Examples of this “dual

⁵⁶ See MERRITT & SIMMONS, *supra* note 54 (discussing the flow chart).

⁵⁷ FED. R. EVID. 404 advisory committee’s note (“Character may itself be an element of a crime, claim, or defense. . . . No problem of the general relevancy of character evidence is involved, and the present rule therefore has no provision on the subject.”); see also FED. R. EVID. 404(b) (stating that “evidence may be admissible for another purpose, such as proving motive, opportunity, intent, preparation, plan, knowledge, identity, absence of mistake, or lack of accident”).

⁵⁸ See CHRISTOPHER B. MUELLER & LAIRD C. KIRKPATRICK, 1 FEDERAL EVIDENCE § 4:39 (4th ed. 2013).

⁵⁹ See, e.g., *Cox Broad. Corp. v. Cohn*, 420 U.S. 469, 489–90 (1975) (defamation); *U.S. v. Brown*, 567 F.2d 119, 120 (D.C. Cir. 1977) (entrapment); *Breeding v. Massey*, 378 F.2d 171, 181 (8th Cir. 1967) (negligent entrustment).

⁶⁰ See MUELLER & KIRKPATRICK, *supra* note 58.

⁶¹ For example, consider a recent case in which the government accused the defendant, a former police officer, of robbing prostitutes and their customers in the customers’ vehicles. *United States v. Pindell*, 336 F.3d 1049, 1051 (D.C. Cir. 2003). At trial, the prosecutor proffered evidence that the defendant had himself paid some of the prostitutes for sex, and the defendant claimed on appeal that this was reversible error. *Id.* at 1057. The evidence initially appears to be inadmissible propensity evidence pursuant to the Merritt and Simmons illustration: the prior bad act is circumstantial evidence that the defendant is a lawbreaker, and because he is a lawbreaker, he robbed the prostitutes. But there is, of course, another purpose for which the prosecutor can proffer the evidence: to lay the foundation for the prostitutes’ identification of the defendant as the perpetrator based on their previous interactions with him. Because there was a “non-propensity” purpose for which the prosecutor proffered the evidence, the court deemed the evidence admissible pursuant to FRE 404(b). *Id.* It is

purpose” evidence include evidence of a defendant’s identity, motive, opportunity to commit the crime, overarching scheme or plan, and knowledge of a fact relevant to the alleged crime.⁶² None of these purposes require a jury to make the impermissible propensity inference, and the evidence could be admitted pursuant to FRE 404(b).⁶³

The remaining three categories of character evidence are exceptions to the rule barring propensity evidence. They are exceptions to the rule because in these instances, the fact finder *does* make the full, forbidden propensity inference. Nonetheless, for reasons of public policy, the Federal Rules of Evidence explicitly allows for these exceptions. First, FRE 608(a) allows “a witness’s credibility [to] be attacked or supported by testimony about the witness’s reputation for having a character for truthfulness or untruthfulness, or by testimony in the form of an opinion about that character.”⁶⁴ Critically, such testimony is admissible to prove that the witness is lying on the witness stand now because the witness is, in fact, a liar.⁶⁵ Although this is the forbidden propensity inference—and would otherwise be inadmissible—a witness’s credibility is so important to the fact finder’s ability to render an accurate verdict that the Federal Rules of Evidence explicitly allow for this type of testimony.⁶⁶ Along similar lines, under certain conditions, Rule 609 allows a party to use a witness’s prior convictions as evidence that the witness is lying on the witness stand.⁶⁷

worth noting, however, that the evidence still could have been excluded under FRE 403 as substantially more prejudicial than probative. *See, e.g.,* United States v. Beechum, 582 F.2d 898, 911 (5th Cir. 1978).

⁶² *See, e.g.,* United States v. Cyphers, 553 F.2d 1064, 1069–70 (7th Cir. 1977) (upholding the admission evidence of past bad acts because it established motive); United States v. Johnson, 525 F.2d 999, 1006 (2d Cir. 1975) (same); *see also* United States v. Lemaire, 712 F.2d 944, 948 (5th Cir. 1983) (evidence of prior bad acts was properly admitted since it “indicate[d] the execution of one scheme or plan, rather than separate and distinct offenses”).

⁶³ *See, e.g.,* United States v. Hamilton, 684 F.2d 380, 384 (6th Cir. 1982) (upholding the trial court’s admission of character evidence that was admitted to show intent and identity); United States v. Lambros, 564 F.2d 26, 31 (8th Cir. 1977) (character evidence was properly admitted since that evidence established identity); United States v. Robinson, 560 F.2d 507, 513 (2d Cir. 1977) (holding character evidence was admissible because it established that the defendant had the opportunity to commit the crime he was on trial for).

⁶⁴ FED. R. EVID. 608(a); *see* United States v. Whitmore, 359 F.3d 609, 616–17 (D.C. Cir. 2004) (discussing FRE 608(a) which specifically examines who may offer the applicable character evidence); *see also* United States v. Jewell, 614 F.3d 911, 926 (8th Cir. 2010) (recognizing that the lower court erred when it excluded bad character evidence that attacked the credibility of a witness).

⁶⁵ *Whitmore*, 359 F.3d at 619–20.

⁶⁶ *See* FED. R. EVID. 608 judiciary committee’s note (discussing the rationale of the rule).

⁶⁷ *See, e.g.,* United States v. Charmley, 764 F.2d 675, 677 (9th Cir. 1985) (affirming the trial court’s decision to admit evidence of the defendant’s past convictions under FRE 609).

Second, and the most complex of the exceptions to the bar on propensity evidence, is the “mercy rule” provision of Rule 404(a)(2). The mercy rule grants a criminal defendant the right to proffer otherwise inadmissible propensity evidence—provided that the evidence is pertinent to the charged offense—either to prove her good character or to prove the victim’s bad character.⁶⁸ The rule also allows the prosecutor to respond in kind with propensity evidence under limited circumstances.⁶⁹

The final exception to the rule barring propensity evidence is the most controversial.⁷⁰ In 1995, in response to several high-profile sexual misconduct acquittals, Congress passed Rules 413, 414, and 415. These Rules—which Congress passed over the near-unanimous objection of the Advisory Committee to the Federal Rules of Evidence⁷¹—explicitly allow the government in both civil and criminal cases to proffer propensity evidence regarding a party’s prior acts indicative of sexual assault or child molestation.⁷²

The character evidence provisions of Articles IV and VI create a doctrine that is incoherent, internally inconsistent, and according to some scholars, the legal equivalent of Swiss cheese.⁷³ When examining the complex series of exemptions and exceptions that pervade Articles IV and VI, it becomes clear that the vast majority of character evidence is *admissible*. It is only when the evidence is proffered for just one limited purpose—to prove that a party acted in conformity with her pertinent character trait—that the evidence is inadmissible, and even that proposition is not always true. The Federal Rules of Evidence’s so-called ban on character evidence is therefore quite narrow. It is

⁶⁸ See FED. R. EVID. 404(a)(2)(A); FED. R. EVID. 404(a)(2)(A) advisory committee’s note to 2006 amendments (explaining the framework of the rule).

⁶⁹ *Supra* note 68.

⁷⁰ Significant criticism and debate accompanied the passage and implementation of these controversial rules. See Louis M. Natali, Jr. & R. Stephen Stigall, “*Are You Going to Arraign His Whole Life?*”: *How Sexual Propensity Evidence Violates the Due Process Clause*, 28 LOY. U. CHI. L.J. 1, 2 (1996); see also Dale A. Nance, *Foreword: Do We Really Want to Know the Defendant?*, 70 CHI-KENT L. REV. 3, 10–14 (1994).

⁷¹ Michael S. Ellis, *The Politics Behind Federal Rules of Evidence 413, 414, and 415*, 38 SANTA CLARA L. REV. 961, 961–62, 971 (1998) (citing a report from the Judicial Conference Committee, which noted that “the Advisory Committee on Evidence Rules reported an unanimous decision, but for one dissenting vote by the representative of the Department of Justice[]”; the Committee criticized the adoption of Rules 413, 414, and 415 as superfluous).

⁷² See FED. R. EVID. 413–15; see also *United States v. McCormack*, 700 F. App’x 643, 645 (9th Cir. 2017) (applying FRE 414); *United States v. Willis*, 826 F.3d 1265, 1270–71 (10th Cir. 2016) (applying FRE 413).

⁷³ Cf. Jessica Murphy, *Swiss Cheese That’s All Hole: How Using Reading Material to Prove Criminal Intent Threatens the Propensity Rule*, 83 WASH. L. REV. 317, 320–21, 327–29 (2008) (discussing inconsistencies in the doctrine).

therefore worth examining the rationale for the narrow ban on propensity evidence and determining whether the Advisory Committee is justified in treating propensity evidence differently from other admissible character evidence.

II. THE PSYCHOLOGICAL LEGITIMACY OF CHARACTER EVIDENCE

Against this background regarding the black letter law of character evidence in federal court, this Part explains the primary justifications for the ban on propensity evidence and the reasons for its purported illegitimacy. It also presents psychology research that challenges the assumptions upon which the ban is based.

As this author has written elsewhere, political theorists “describe the concept of legitimacy as the status and acceptance that governed people confer onto their governors’ institutions and conduct based on the belief that those actions constitute an appropriate use of power.”⁷⁴ Indeed, [a]ccording to German sociologist Max Weber, the governed confer legitimacy onto legal actors via an alignment of values between the political actors—that is, through public trust that the government will act in the interests of the governed—and not through the government’s coercion or force.⁷⁵

Social psychologists have studied the concept of legitimacy, as Weber has defined it, with respect to how the public perceives legal regimes. These scholars posit that the public perceives the courts as serving two distinct but related goals: (1) “to get to the truth of a legal matter” (that is, to maximize “decisional accuracy” by correctly finding the facts that underlie the dispute), and (2) “to do so in a manner that the public deems to be fair and just” (termed “procedural justice”).⁷⁶ Recent scholarship has supported this claim with empirical evidence

⁷⁴ See Justin Sevier, *Evidentiary Trapdoors*, 103 IOWA L. REV. 1155, 1169 (2018) (citing JOSEPH RAZ, *THE MORALITY OF FREEDOM* (1986)); see also JOHN R. SCHERMERHORN JR. ET AL., *ORGANIZATIONAL BEHAVIOR* (2011) (discussing “interactional legitimacy” between social actors).

⁷⁵ Sevier, *supra* note 74, at 1169–70 (citing Max Weber, *Politics as a Vocation*, in *FROM MAX WEBER: ESSAYS IN SOCIOLOGY* (H.H. Gerth & C. Wright Mills eds., 1991)). “[T]o the extent that a misalignment develops between the values of the governed and the actions of the government, political legitimacy is endangered.” *Id.* at 1170 (citing JOHN RAWLS, *POLITICAL LIBERALISM* 121 (1993) (“suggesting that political institutions that lack legitimacy exercise their power unjustifiably and will not be obeyed”).

⁷⁶ *Id.* at 1172 (citing John Thibaut & Laurens Walker, *A Theory of Procedure*, 66 CALIF. L. REV. 541, 541 (1978)). As other scholars have noted, the courts—including the Supreme Court—have stated that “a major objective of litigation is to obtain a close correspondence between proven fact and historical truth.” Uviller, *supra* note 17, at 845 n.1. As Professor Uviller notes, Justice White once wrote that the legal system “stresse[s] the importance of arriving at the truth in criminal trials,” and that a “wealth of other recent cases [] have followed this homily [and] that it is fast becoming a major theme of contemporary criminal jurisprudence.” *Id.* Professor

demonstrating that decisional accuracy and procedural justice account for the vast majority of the variance in the public's willingness to legitimize the courts.⁷⁷

Applying these principles to character evidence, legal scholars provide three reasons why propensity inferences are illegitimate. First, the jury may overtly use the evidence for an impermissible purpose: "to penalize the accused for past misdeeds or for being a bad person."⁷⁸ Second, rather than overtly penalizing the defendant because of his past misdeeds, the jury might inadvertently overvalue the probative weight of the evidence.⁷⁹ Finally, commentators have argued that it "seems unfair to require the defendant to be prepared not only to defend against the immediate charges, but to answer for other alleged misdeeds, or more generally to explain his past."⁸⁰

These arguments can be organized along the broad psychological dimensions that compose the public's willingness to legitimize the courts. The first two arguments represent concerns over the *accuracy* of verdicts that are premised in part on propensity evidence: that jurors will intentionally or unintentionally err by relying too much on such evidence. The third argument invokes concerns over *procedural justice* and suggests that admitting propensity testimony violates our shared notions of fair play.

Notably, these concerns are empirical. Researchers can examine just how well jurors evaluate character evidence and how fair the public perceives it to be at trial.⁸¹ A substantial body of research in psychology—in the scholarship on impression formation and person perception—suggests that, if courts provide jurors with the tools to evaluate propensity evidence appropriately, jurors can demonstrate great competency with respect to how they evaluate that evidence. Moreover, the research on procedural justice suggests that jurors are more likely

Uviller penned a follow-up article focusing on the importance of "truth and the adjudicative process." H. Richard Uviller, *Credence, Character, and the Rules of Evidence: Seeing Through the Liar's Tale*, 42 DUKE L.J. 776, 779–93 (1993).

⁷⁷ See, e.g., Tom R. Tyler & Justin Sevier, *How Do the Courts Create Popular Legitimacy?: The Role of Establishing the Truth, Punishing Justly, and/or Acting Through Just Procedures*, 77 ALB. L. REV. 1095, 1097 (2013/2014).

⁷⁸ See MUELLER & KIRKPATRICK, *supra* note 58, at § 4:22.

⁷⁹ *Id.*

⁸⁰ *Id.*

⁸¹ Surprisingly, only a handful of studies have been conducted to date, with inchoate results. See, e.g., Jennifer S. Hunt & Thomas Lee Budesheim, *How Jurors Use and Misuse Character Evidence*, 89 J. APPLIED PSYCHOL. 347, 350, 358 (2004). For a more recent review of the literature, see Jennifer S. Hunt, *The Cost of Character*, 28 U. FLA. J.L. & PUB. POL'Y 241 (2017).

to delegitimize trials when the government shields such evidence from the fact finder. This Part discusses these bodies of research in more detail.

A. *Decisional Accuracy: Impression Formation*

This section examines whether juries are likely to afford propensity evidence too much weight—resulting in inaccurate verdicts—if it is admissible at trial. To do so, we examine the social psychological processes that govern how we form impressions of others in our social world and how those impressions affect the attributions we make regarding their behavior. The analysis now turns to the psychological phenomena of impression formation and person perception.

Forming accurate impressions of others is an intricate process involving several overlapping psychological mechanisms. In social psychology, impression formation refers to the process by which disparate pieces of information about another person are integrated to form a global impression of the individual.⁸² At its core, the process is driven by expectations of coherence (and unity) of attitudes and behaviors in the personalities of others.⁸³ Person perception (sometimes termed “social perception”) is a subset of impression formation that accounts specifically for how we evaluate other human beings.⁸⁴ It is the process by which we observe others, make sense of the information that we extract when we observe them, and use that information to inform our judgments about them.⁸⁵

⁸² See S. E. Asch, *Forming Impressions of Personality*, 41 J. ABNORMAL & SOC. PSYCHOL. 258, 258–62 (1946) (explaining the concept and proposing a theory of its existence).

⁸³ See, e.g., Sanne Nauts et al., *Forming Impressions of Personality: A Replication and Review of Asch’s (1946) Evidence for a Primacy-of-Warmth Effect in Impression Formation*, 45 SOC. PSYCHOL. 153, 154 (2014) (discussing the work of psychologist Solomon Asch and noting his conclusions that “perceivers form coherent, unitary impressions of others”). Two major theories have gained prominence in explaining how we form impressions of others. The first is the Gestalt approach, which views the formation of a general impression as the sum of multiple interrelated impressions. See *id.* (discussing Asch’s example of the meaning of levels of gaiety in an “intelligent man” and a “stupid man”). As a person attempts to derive meaning and coherence from another person’s attitudes or behaviors, previous impressions of that person (stemming from prior behaviors) play a dominating role in contextualizing those current behaviors and interpreting their meaning. See David L. Hamilton, & Steven J. Sherman, *Perceiving Persons and Groups*, 103 PSYCHOL. REV. 336, 337–38 (1996). The cognitive algebraic approach, in contrast, assumes that new information about an individual is integrated and evaluated independent of previous information about that individual, and combines with that previous information to form a dynamic, malleable impression of the attitudes, personality, and behavior of others. See Samuel Himmelfarb, *Integration and Attribution Theories in Personality Impression Formation*, 23 J. PERSONALITY & SOC. PSYCHOL. 309, 310, 312–13 (1972).

⁸⁴ See, e.g., *Person Perception*, PSYCHOL., <https://psychology.iresearchnet.com/social-psychology/social-cognition/person-perception/> (last visited Nov. 25, 2018).

⁸⁵ *Id.*; see also ELLIOT ARONSON ET AL., *SOCIAL PSYCHOLOGY* 83–115 (7th ed. 2010).

There are many ways in which we classify other people in our environment. Social psychologist Gordon Allport's research on trait theory suggests that we organize our impressions of others into general "traits," which are habitual patterns of behavior, thought, and emotion.⁸⁶ We then organize these distinct traits into a hierarchy, prioritizing the "cardinal traits"—the ones most diagnostic of a person's underlying personality—but allowing for central and secondary traits as well.⁸⁷

Through the classic experiments of social psychologist Solomon Asch, four general principles of impression formation and person perception have emerged: (1) individuals have a natural inclination to make global dispositional inferences about the nature of another person's personality; (2) we expect the behaviors we observe in others to reflect those stable personality traits; (3) individuals attempt to fit information about an individual's attitudes and behaviors into a hierarchy of traits that is a meaningful and coherent whole; and (4) we explain away and rationalize inconsistencies between observed behavior and impressions of the individual's cardinal personality traits if they conflict.⁸⁸

1. Cause for Concern: Overreliance on Personality Traits

Our initial impressions of others, however, tell only part of the story with respect to how we interact with those individuals in our social environment. A vast body of research suggests that we make implicit links from our initial

⁸⁶ Gordon W. Allport, *Personality and Character*, 18 PSYCHOL. BULL. 441, 441–45 (1921) (advancing his "trait theory" of psychological impression formation).

⁸⁷ Floyd H. Allport & Gordon W. Allport, *Personality Traits: Their Classification and Measurement*, 16 J. ABNORMAL PSYCHOL. & SOC. PSYCHOL. 6, 8–9 (1921) (discussing the measurement and differences among cardinal traits and secondary traits). Often, these cardinal traits suggest a constellation of other, closely related traits that we believe the individual possesses. For example, if we encode an individual as cardinally friendly, we are more likely to believe that she is happy and generous as well. See, e.g., David J. Schneider, *Implicit Personality Theory: A Review*, 79 PSYCHOL. BULL. 294, 297 (1973) (reviewing the literature).

Recent research suggests that, partly as a result of our social evolution over time, our impressions of an individual's cardinal traits tend to fall along two axes, which account for roughly 80% to 90% of the variance in our impressions. See Susan T. Fiske & Eugene Borgida, *Best Practices: How to Evaluate Psychological Science for Use by Organizations*, 31 RES. ORG. BEHAV. 253, 259 (2011) (citing Bogdan Wojciszke, *Morality and Competence in Person- and Self-Perception*, 16 EUR. REV. SOC. PSYCHOL. 155 (2005)). We tend to evaluate others with respect to (1) how warm and trustworthy they are, and (2) how strong and competent they are, and we tend to do so outside of our conscious awareness. See Susan T. Fiske et al., *A Model of (Often Mixed) Stereotype Content: Competence and Warmth Respectively Follow from Perceived Status and Competition*, 82 J. PERSONALITY & SOC. PSYCHOL. 878, 891 (2002).

⁸⁸ See Hamilton & Sherman, *supra* note 83; see also Edward R. Hirt, *Do I See Only What I Expect? Evidence for an Expectancy-Guided Retrieval Model*, 58 J. PERSONALITY & SOC. PSYCHOL. 937, 937–38 (1990); Curt Hoffman et al., *The Role of Purpose in the Organization of Information About Behavior: Trait-Based Versus Goal-Based Categories in Person Cognition*, 40 J. PERSONALITY & SOC. PSYCHOL. 211, 211–13 (1981).

personality assessments to the behaviors of others. This body of research is referred to as “attribution theory,” named for how we attribute the behavior of others in our environment in the absence of direct access to others’ internal mental states.⁸⁹

Specifically, an attribution is defined as the use of observations about a target individual to (1) gather information regarding the individual’s motivations for her observable behaviors; and (2) predict the individual’s future behaviors.⁹⁰ Attribution theorists focus on identifying the systems that people use to make causal inferences regarding the behaviors of others. Attribution theory plays an important role in our ability not only to predict behavior in the future, but also to “postdict” behavior.⁹¹

Psychologists have enumerated several dangers, however, in attributing all of a person’s behaviors to personality traits. In the 1960s, psychologist Walter Mischel stunned personality theorists when he performed a meta-analysis of the effects of personality on subsequent behavior and found only a moderate correlation ($r = .30$)—meaning that personality provides little predictive ability of subsequent behavior.⁹² A series of classic experiments examining the role of character traits on subsequent behavior supported Professor Mischel’s meta-analysis. Researchers found, almost uniformly, no effects of a person’s personality characteristics on her subsequent behaviors; instead they found effects of innocuous situational variables.⁹³ For example, in a famous study of bystander intervention, when people believed that another room in a building had been set afire, people’s willingness to alert others to the danger was not predicted by their levels of altruism or their locus of control, but instead by the sheer number of other people in the room with them.⁹⁴

⁸⁹ See SAUL KASSIN ET AL., *SOCIAL PSYCHOLOGY* (8th ed. 2010) (giving a brief overview of the field); see also FRITZ HEIDER, *THE PSYCHOLOGY OF INTERPERSONAL RELATIONS* 16–18 (1958) (discussing, from the point of view of the founder of the field, its general tenets).

⁹⁰ See KASSIN ET AL., *supra* note 89 (giving a brief definition of the term attribution).

⁹¹ See SAKS & SPELLMAN, *supra* note 23, at 151 (“We say ‘postdict’ because in a trial the question is whether a defendant did something in the past, though the tools the factfinders are being invited to use are those of intuitive prediction.”).

⁹² WALTER MISCHEL, *PERSONALITY AND ASSESSMENT* 78 (George Mandler ed., 1968); see also SAKS & SPELLMAN, *supra* note 23, at 154.

⁹³ See SAKS & SPELLMAN, *supra* note 23, at 151–54 (putting these findings in context).

⁹⁴ *Id.* at 151–52 (citing BIBB LATAN. . . & JOHN M. DARLEY, *THE UNRESPONSIVE BYSTANDER: WHY DOESN’T HE HELP?* (1970)). Similarly, in an experiment involving the helping behavior of a group of seminary students, researchers found that it was the degree to which they were in a hurry, and not their degree of religiosity or the extent to which they were thinking of helping others, that predicted whether they rendered aid to a perceived-injured bystander. *Id.* at 152 (citing John M. Darley & C. Daniel Batson, *From Jerusalem to Jericho: A Study of Situational and Dispositional Variables in Helping Behavior*, 27 *J. PERSONALITY & SOC. PSYCHOL.* 100 (1973)).

2. *Cause for Optimism: The Interactionist Model*

Insights from social psychology suggest that we do *not* solely attribute a person's behaviors to her personality. Rather, we implicitly base our judgments, with varying degrees of success, on an *interaction* between an individual's personality and situational factors that influence the behavior. This theory of behavioral attribution is referred to as the "interactionist" model and is statistically a better behavioral predictor than personality traits alone.⁹⁵

The best recognized version of the interactionist approach is the two-step model proposed by Daniel Gilbert and Patrick Malone.⁹⁶ Their influential theory posits consecutive stages of attribution. People first make an internal attribution about others and afterward consider the possible external explanations for their behavior.⁹⁷ In other words, we initially attribute the behaviors of others to their internal personality traits, but then we modify this attribution to account for appropriate situational forces.

To be sure, human beings are not perfect calibrators. A body of research from psychologist Lee Ross suggests that this calibration system sometimes breaks down in favor of personality attributions, whereby we automatically make dispositional attributions for a social actor's behavior and then insufficiently adjust for situational influences.⁹⁸ This phenomenon, termed the "fundamental attribution error," suggests that we sometimes overweigh a person's character traits in our behavioral judgments.⁹⁹

Nonetheless, the interactionist model also suggests that, on the whole, people are actually quite *competent* at arriving at appropriate attributions under many

⁹⁵ See Bill D. Bell & Gary G. Stanfield, An Interactionist Appraisal of Impression Formation: The 'Central Trait' Hypothesis Revisited, 9 KAN. J. SOC. 55, 63 (1973).

⁹⁶ Daniel T. Gilbert & Patrick S. Malone, *The Correspondence Bias*, 117 PSYCHOL. BULL. 21, 22 (1995). Psychologists have proposed several models for how people make situational-interactionist attributions about the behaviors of others. See, e.g., Edward E. Jones & Keith E. Davis, *From Acts to Dispositions: The Attribution Process in Person Perception*, in 2 ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY 219, 222–24 (Leonard Berkowitz ed., 1965) (correspondence inference theory); Harold H. Kelley, *Attribution Theory in Social Psychology*, in 15 NEBRASKA SYMPOSIUM ON MOTIVATION 192, 197 (David Levine ed., 1967) (covariation model of attribution).

⁹⁷ Gilbert & Malone, *supra* note 96.

⁹⁸ Lee Ross, *The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process*, in 10 ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY 173, 184 (Leonard Berkowitz ed., 1977).

⁹⁹ *Id.* A common example of the fundamental attribution error would be initially assuming that a person who cuts us off in traffic is rude and impatient (a dispositional attribution) and failing to adjust for a situational reason for his behavior (for example, that he was rushing to the hospital).

circumstances.¹⁰⁰ Indeed, researchers have identified several conditions under which people tend to be more accurate with respect to their social judgments of others: when people have greater history of experiences with others, observe them directly in their presence, are exposed to probability rules (including base rate information), and most importantly, when they are motivated by concerns of open-mindedness and accuracy.¹⁰¹

In sum, the personality and social psychology literature provides support for several propositions about the use of propensity evidence in court. First, jurors are likely to attend to propensity evidence and afford it probative weight in their verdicts, although the degree to which it affects their verdicts is unclear. Second, situational factors are at least as important—and perhaps more important—than personality factors in explaining a social actor’s behavior, and research on the fundamental attribution error suggests that people are not always as attentive to the latter as they are to the former. But third, and most importantly, several factors—many of which are relevant to an individual’s role as a juror in a legal proceeding—focus people on diagnostic, situational cues when making inferences about human behavior.

B. Procedural Justice: Fair Process

A reduction in a tribunal’s decisional accuracy is only one of the concerns raised by the use of propensity evidence.¹⁰² Policymakers have also expressed concern that the use of propensity evidence is anathema to accepted notions of fair evidence-gathering such that the public may refuse to legitimize courts that

¹⁰⁰ This general competency is subject to moderating variables, including aspects of the evaluator, the target, the trait being judged, and the inputs upon which those judgments are made. See David C. Funder, *On the Accuracy of Personality Judgment: A Realistic Approach*, 102 *PSYCHOL. REV.* 652, 656 (1995).

¹⁰¹ See KASSIN ET AL., *supra* note 89. Thus, according to researchers, if the Federal Rules of Evidence bestow the correct tools upon jurors for evaluating propensity evidence, they will make justifiable decisions in weighing it. Specifically, these tools would focus the juror on an individual’s past behaviors instead of on their general reputation or personality traits. A wealth of psychology research suggests that although personality variables do not predict future behavior as much as scientists previously believed, past behavior is, under many circumstances, highly predictive of future behavior. See, e.g., DANIEL L. SCHACTER ET AL., *PSYCHOLOGY* (2d ed. 2010) (discussing Thorndike’s “law of effect”).

¹⁰² See FED. R. EVID. 404 advisory committee’s note (citing with approval the California Law Revision Commission’s conclusion, when evaluating potential changes to the propensity rule in the California Evidence Code, that “[c]haracter evidence is of slight probative value and may be very prejudicial. It tends to distract the trier of fact from the main question of what actually happened on the particular occasion. It subtly permits the trier of fact to reward the good man to punish the bad man because of their respective characters despite what the evidence in the case shows actually happened”).

rely on it.¹⁰³ The psychology literature, however, suggests that this concern may be overstated.

Distributive outcomes matter—and they matter a lot—to our perceptions of whether a governing body’s decisions are just and legitimate.¹⁰⁴ For example, media reports reflect that public outrage over acquittals in recent, high-profile criminal trials stems in part from a belief that the facts adduced in court did not align substantively with popular perceptions of what had truly occurred.¹⁰⁵ But distributive outcomes are not the sole determinant of public perceptions of the justice provided by a governing body, and in fact, outcomes might not be even the strongest predictor of popular justice.

Instead, public perceptions of the justice provided by a governing body—and the legitimacy of that body—stem even more strongly from perceptions of the fairness of the process employed by the body to reach its substantive decisions. Procedural justice theorists therefore argue that “people’s reactions to their experiences with legal authorities are strongly shaped by their subjective evaluations of the justice of the *procedures* used to resolve their case.”¹⁰⁶ Procedural justice researchers have demonstrated, in clever experiments, that the importance of fair process to popular perceptions of a decision maker’s

¹⁰³ See *id.* (noting with concern that “expanding concepts of ‘character’ which seem of necessity to extend into such areas as psychiatric evaluation and psychological testing, coupled with expanded admissibility, would open up such vistas of mental examinations as caused the [United States Supreme] Court concern in *Schlagenhauf v. Holder*, 379 U.S. 104, 85 S. Ct. 234, 13 L.Ed.2d 152 (1964)”).

¹⁰⁴ See, e.g., Robert Folger & Mary Konovsky, *Effects of Procedural and Distributive Justice on Reactions to Pay Raise Decisions*, 32 ACAD. MGMT. J. 115, 122–24 (1989) (reporting the results of an experiment that demonstrated that attitudes regarding the distributive outcome of a pay raise decision strongly predicted participants’ satisfaction with the decision).

¹⁰⁵ See, e.g., Andrew Cohen, *Law and Justice and George Zimmerman*, ATLANTIC (July 13, 2013), <https://www.theatlantic.com/national/archive/2013/07/law-and-justice-and-george-zimmerman/277772/> (noting that the George Zimmerman trial “is above all a blunt reminder of the limitations of our justice system. Criminal trials are not searches for the truth, the whole truth, and nothing but the truth. They never have been. Our rules of evidence and the Bill of Rights preclude it. Our trials are instead tests of only that limited evidence a judge declares fit to be shared with jurors, who in turn are then admonished daily, hourly even, not to look beyond the corners of what they’ve seen or heard in court”); see also Breeanna Hare, ‘What Really Happened?’: *The Casey Anthony Case 10 Years Later*, CNN (June 30, 2018, 12:54 AM), <https://www.cnn.com/2018/06/29/us/casey-anthony-10-years-later/index.html> (interviewing the medical examiner in the Casey Anthony trial, who noted, “what I was most appalled with was the lack of the truth and the lack of substantiated information. You could just say lies and not back it up by any kind of evidence and it was allowed”).

¹⁰⁶ Tom Tyler & David Markell, *The Public Regulation of Land-Use Decisions: Criteria for Evaluating Alternative Procedures*, 7 J. EMPIRICAL LEGAL STUD., 538, 541 (2010) (emphasis added) (citations omitted); see generally E. ALLAN LIND & TOM R. TYLER, *THE SOCIAL PSYCHOLOGY OF PROCEDURAL JUSTICE* (1988) (discussing theories of procedural justice at length).

legitimacy likely stems from the signals that fair processes send to individuals: that they are valued and respected members of society.¹⁰⁷

To that end, subsequent psychology research has clarified that the public conceives of “fair process” in the legal context in specific, concrete ways. As this author has written elsewhere:

Researchers have identified several procedural factors that influence the perceived legitimacy of a decision making body: the decision maker’s neutrality, the degree of respect and dignity that the decision maker confers onto the parties, the level of voice and control that the parties have over the legal dispute, and the degree to which parties can trust the decision maker’s motive to be fair. These factors manifest themselves inside and outside the laboratory in both criminal and civil disputes. In legal adjudication, for example, perceptions of fair process confer legitimacy on actors including judges and juries. People’s views of procedural fairness also inform their perceptions of legitimacy in alternative dispute resolution—including mediation and arbitration—and the decision makers in those paradigms.¹⁰⁸

Turning to the question of propensity evidence, the psychology research suggests that (subject to several nuances), the public may respond negatively to instances in which courts shield fact finders from evidence that would reasonably assist them in arriving at accurate verdicts.¹⁰⁹ It further suggests that the parties’ loss of voice in the proceedings—from the exclusion of the relevant evidence—will lead to verdicts that are *delegitimized* when propensity evidence is disallowed.¹¹⁰

III. THREE EXPERIMENTS

This Article now reports the results from three original experiments, with over 1,200 participants, which examined the two rationales for the bar against propensity evidence in court: (1) jurors will overvalue propensity evidence at

¹⁰⁷ See, e.g., Tom R. Tyler, *The Psychology of Procedural Justice: A Test of the Group-Value Model*, 57 J. PERSONALITY & SOC. PSYCHOL. 830, 837 (1989).

¹⁰⁸ Justin Sevier, *Popularizing Hearsay*, 104 GEO. L.J. 643, 659–60 (2016).

¹⁰⁹ See, e.g., George Loewenstein, *The Psychology of Curiosity: A Review and Reinterpretation*, 116 PSYCHOL. BULL. 75, 93 (1994) (discussing the relationship of “information gap[s]” to the psychology of curiosity, which the author defines as “a discrepancy between what one perceived and what one expected to perceive” in terms of information about one’s environment); see also David R. Shaffer et al., *Effects of Withheld Evidence on Juridic Decisions*, 42 PSYCHOL. REP. 1235, 1236–38 (1978) (finding that mock jurors are attuned to such information gaps and penalize legal actors whom they perceive to be withholding relevant information from them).

¹¹⁰ Loewenstein, *supra* note 109; see also Shaffer et al., *supra* note 109.

the expense of reaching an accurate verdict, and (2) regardless of the effect of such evidence on the accuracy of verdicts, the public is unwilling to legitimize trials in which character evidence is presented because it is procedurally unjust to introduce such evidence.

To test these rationales for the bar on propensity evidence, we designed three experiments. Studies 1 and 2 examine whether propensity evidence threatens the accuracy of legal trials: what weight, if any, do mock jurors afford propensity evidence and can they distinguish between accuracy-enhancing and accuracy-diminishing features of such evidence? Study 3 examines the role of character evidence in perceptions of procedural justice: under what conditions (if any) will the public legitimize verdicts that are procedurally the product of character witness testimony?

A. *Study 1: The Power of Propensity*

Our first study examines the degree to which mock jurors attend to character evidence and the extent to which it affects their trial verdicts. Our participants read a vignette in which they imagined themselves as jurors at a trial at their local courthouse. The study manipulated three variables. First, and most importantly, we manipulated the party that produced the character witness, such that the propensity testimony was either used as part of the defense or as part of the evidence *against* the defendant. Second, we examined whether any effects of the propensity evidence on our mock jurors' verdicts varied with the type of case that was presented: either a fatal shooting, a battery, or an attempted sexual assault. Finally, we manipulated the legal setting, such that the alleged event gave rise to either civil or criminal liability. We measured our participants' attitudes toward the evidence, their verdicts, and their perceptions of whether the defendant committed the acts for which he was accused.

If mock jurors pay attention to character evidence and evaluate it with care, we would expect the identity of the party proffering the evidence to affect our participants' verdicts, such that, compared to a control condition with no propensity evidence, conviction rates should rise when the witness testifies to the defendant's character for violence. Conversely, conviction rates should fall when the witness testifies to the defendant's good character. Moreover, we believe that character evidence does not enjoy special weight in a criminal rather than civil proceeding, and we have no theoretical reason to believe that character evidence has a differential impact based on the subject matter of the trial. The following section reports the methodology and results of Study 1.

1. *Participants in Study 1*

We recruited 812 participants for this online study through the Amazon Mechanical Turk recruitment service. Once recruited, participants received a link to the study, which was hosted on the Qualtrics online survey platform.¹¹¹ Participants were paid \$1.00 for their participation, and they were told that the study was designed to measure their attitudes about a hypothetical legal case. All participants completed the study within fifteen minutes.

The average participant was 36.16 years old (with a standard deviation of 10.73).¹¹² The sample was split evenly by gender, with women composing 52.80% of the sample. The sample reflected the racial diversity of the U.S. population as well, with 26.10% of the sample identifying as non-white.¹¹³ Roughly 59.20% of participants had completed at least a college degree, and the median participant income was approximately \$50,000. The political affiliation of participants varied, although the majority of participants identified as moderate (28.30%) to liberal (30.70%). Table 1 provides descriptive statistics for the participants involved in this study.

2. *Procedure and Measures in Study 1*

After giving their informed consent, participants read materials that asked them to imagine that they had been summoned for jury duty at their local courthouse.¹¹⁴ They were told to imagine themselves in the jury box and to imagine the judge on a raised platform to their right. They were asked to imagine the prosecutor (or plaintiff) seated at a table to their immediate left, and to imagine the defense counsel and the defendant seated at a table further in the distance. The judge then called the trial to order.

¹¹¹ mTurk is an inexpensive platform for collecting high-quality data from a representative sample of the population. See, e.g., Adam J. Berinsky et al., *Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk*, 20 POL. ANALYSIS 351, 366 (2012); Michael Buhrmester et al., *Amazon's Mechanical Turk: A New Source of Inexpensive, yet High-Quality, Data?*, 6 PERSP. ON PSYCHOL. SCI. 3, 5 (2011); Winter Mason & Siddharth Suri, *Conducting Behavioral Research on Amazon's Mechanical Turk*, 44 BEHAV. RES. METHODS 1, 2–3 (2011).

¹¹² All demographic information provided by participants was self-reported.

¹¹³ See, e.g., *QuickFacts*, U.S. CENSUS BUREAU, <https://www.census.gov/quickfacts/fact/table/US/PST045216> (last visited Nov. 25, 2018) (listing current demographic statistics from the U.S. census).

¹¹⁴ We adapted the fact pattern for this study, and for the two studies that follow, from this author's article, Sevier, *supra* note 74, at 1182–83.

Table 1: Participant Demographics (Study 1)

	<i>%</i>	<i>N</i>
Age (Median: 34.00)		
< 30	31.36	254
30-39	36.91	299
40-49	20.25	164
50-59	07.53	61
60-74	03.95	32
Gender		
Male	47.23	383
Female	52.77	428
Race		
Caucasian	73.89	597
African-American	08.66	70
Hispanic	06.31	51
Asian	08.54	69
Other	02.60	21
Education		
High School	09.63	78
Some College	31.23	253
College	44.20	358
Master's	11.85	96
Ph.D. or Professional	03.09	25
Political Affiliation		
Very Conservative	05.43	44
Conservative	18.15	147
Moderate	28.27	229
Liberal	30.74	249
Very Liberal	16.30	132
Other	01.11	09
Income		
Less than \$30,000	25.64	208
\$30,000 - \$49,999	24.16	196
\$50,000 - \$69,999	19.10	155

We subjected our participants to three different experimental manipulations. First, we randomly assigned our participants to one of three cases, all involving an altercation in the early morning hours in a parking lot at an upscale mall. In the first case, the defendant was accused of shooting the victim in a botched robbery stemming from an illicit narcotics transaction. In the second case, the defendant was accused of hitting the victim with a baseball bat during a heated argument. In the third case, the defendant was accused of lying in wait in the parking lot to sexually assault the victim.

Second, we manipulated the judicial setting in which the cases arose. Half of our participants were told that the dispute was a civil matter between the alleged victim and the defendant, whereas half were told that the government had initiated criminal proceedings against the defendant. If our participants were assigned to the civil version of each case, they read about either a wrongful death action filed by the victim's next of kin (the botched narcotics deal described above), a civil battery case (involving the baseball bat), or a hybrid assault and intentional infliction of emotional distress case (stemming from the attempted sexual battery). Participants assigned to the criminal version of each case instead read about a second-degree murder action (the narcotics case), a criminal battery case, or a sexual assault case. These manipulations jointly created six different experimental conditions to which our participants were randomly assigned.

The attorneys next presented their opening statements to the jury. In each version of the experiment, the opening statements suggested that the incident occurred in a mall parking lot and that the identity of the perpetrator was at issue. The defendant denied wrongdoing and focused on the circumstantial nature of the evidence.

The case against the defendant then proceeded, either as a criminal prosecution or as a civil suit. The majority of the evidence against the defendant was the same in each experimental condition: it included the testimony of a police officer, a forensic analyst, and the defendant's brother. The police officer testified to his observations of the scene when he found the victim. The forensic analyst testified to tests he conducted on the weapon alleged to have been used by the defendant. The defendant's brother testified to the defendant's opportunity to commit the crime.

Each case against the defendant involved the discovery of physical evidence at the scene of the incident: a weapon (dropped near the scene) and a ski cap (bearing a local sports team logo) left behind by the perpetrator as he fled. A

complete summary of the evidence against the defendant in one of our scenarios is footnoted below.¹¹⁵

At this point in the trial, we imposed our third (and final) manipulation. For most of our participants, the next person to testify was a character witness in the form of the defendant's co-worker, who would be called either by the prosecution (or the plaintiff in the civil version of the case) or by the defense. For a smaller portion of participants, who served as our experimental controls, no character evidence was presented at the trial.

In our non-control conditions, our third manipulation functioned as follows. For half of these participants, the character witness was the final witness called by the prosecution and testified that the defendant had a bad reputation in the community for being a lawbreaker. For our remaining participants, the defense called the character witness, who testified to the defendant's good and generous character within the community.

The defendant always testified as the final witness at the trial. In each experimental condition, the defendant admitted that he owns a considerable amount of sports memorabilia, but he denied that he owned the cap that was admitted into evidence. He also admitted that he is an avid hunter and owns many weapons. He testified further that he had been on a hunting trip on the day of the murder, and that he was on the trip alone. He also confirmed that he drives a silver Acura sedan.

¹¹⁵ The second-degree murder case proceeded as follows. In her opening statement, the prosecutor suggested that the evidence would show that the victim died during a botched cocaine sale. The prosecutor first called the police officer who responded to the scene. The officer identified the victim and testified that the victim had been shot before 7:00 AM. The officer testified that he observed at the scene an unregistered .45-caliber handgun that appeared to have been recently fired. He also observed a hat bearing the logo of the local sports team, which did not appear to be owned by the victim, as well as a small bag of cocaine in the victim's jacket pocket. The mall's security footage did not provide a clear image of the perpetrator, he testified, but the footage showed the perpetrator speeding away from the scene in a silver or gray sedan. The officer concluded his testimony by stating that he arrested the defendant for the crime later that day, after a swift investigation.

The prosecutor next called a forensic expert to the witness stand. The expert first testified that the bullets in the chamber of the handgun that the officer found at the scene were consistent with the bullet found in the victim's abdomen. The expert next testified to the results of scientific tests that his lab conducted. He testified that the defendant's hands had tested positive for the presence of gunpowder residue when he was arrested. The expert stated that the test has a negligible error rate and that the test is commonly used in criminal investigations.

Finally, the prosecutor called the defendant's co-worker to the witness stand. The co-worker described the defendant as a secretive person who enjoyed hunting and shooting guns, which he owned in abundance. He also testified that the defendant is a die-hard fanatic of the local sports team, and that the defendant owns memorabilia and apparel that bears the local team's logo. On cross-examination, however, he could not be sure that the hat found at the crime scene belonged to the defendant. Finally, he testified that the defendant drives a silver Acura sedan.

Once the defense rested its case, all participants read the parties' closing arguments and jury instructions. The instructions clarified the elements of each alleged cause of action and they specified the standard of proof: either beyond a reasonable doubt (in the criminal case condition) or by a preponderance of the evidence (in the civil case condition).

Participants then rendered a verdict. We also asked them several questions about the trial, including the strength of each witness's testimony, the strength of the case against the defendant, the likelihood that the defendant committed the crime, and the likelihood of finding the defendant civilly or criminally liable.¹¹⁶ Additionally, we asked participants to rate the usefulness of the evidence in addition to collecting demographic and personality trait information from them.¹¹⁷ After they answered these questions, we thanked them for their participation, debriefed them regarding the experimental hypotheses, and concluded the study.

3. *Results of Study 1*

This section proceeds in two parts. First, it reports the effects of our experimental manipulations on our participants' "global" attitudes toward the trial: their verdicts in the case, the likelihood that they would find the defendant liable, the likelihood that the defendant committed the acts of which he is accused, and their confidence in their decisions.

Second, we measured our participants' attitudes toward the testimony produced at the trial, with a focus on their attitudes toward the propensity evidence. From these data, we created a psychological model to account for the effects of propensity evidence on our mock jurors' decisions.

¹¹⁶ Participants rated these phenomena on Likert Scales anchored at 1 (e.g., unwilling to convict, unlikely to have committed the act, not confident) and 7 (e.g., highly willing to convict, highly likely to have committed the act, and highly confident). A Likert Scale is a psychometric scale that is routinely used in questionnaires and is analyzed as an ordinal variable (frequently a range from 1 to 7). See ROBERT M. LAWLESS ET AL., *EMPIRICAL METHODS IN LAW* 145–46 (2d ed. 2016).

¹¹⁷ Participants rated the strength and usefulness of the testimony of the police officer, forensic analyst, defendant's brother, the character witness, and the defendant. We also measured, as control variables in all three studies, participants' levels of authoritarianism, their need for cognition, their need for closure, their attitudes toward social dominance, their belief in a just world, and any negative attitudes they hold toward courts or toward attorneys. We include these variables as controls in the models that we report in Studies 1, 2, and 3. For a list of the personality items that we used in Study 1, see Sevier, *supra* note 74, at 1206 (using similar measures in the context of a study examining the respondeat superior doctrine in agency law).

a. Main Analysis I: Global Attitudes in Study 1

Next, we evaluated the hypotheses underlying Study 1. To determine whether the character witness's testimony, the type of case in which the witness testified, or the civil or criminal setting affected participants' case verdicts, we conducted a stepwise logistic regression on our participants' decisions to find the defendant civilly or criminally liable.¹¹⁸ The results confirmed our hypotheses. As we predicted, our participants found the defendant liable more often in civil cases, where the burden of proof is lower, than in criminal cases, where it is higher.¹¹⁹ Moreover, the type of case to which participants were exposed—a shooting, a beating, or a sexual assault—had no effect on participants' verdicts; they found the defendant liable at roughly the same rate across all three experimental conditions.¹²⁰

Most importantly, we also found an effect of the character evidence on participants' verdicts, such that participants found the defendant liable more often when the witness testified for the prosecution and less often when the witness testified for the defense.¹²¹ To determine the meaningfulness of these differing liability rates against baseline, we compared our control condition (in which no character evidence was adduced at trial) against our experimental character evidence conditions. The tests revealed that the character witness's testimony for the prosecution increased the liability rate from baseline and was

¹¹⁸ A stepwise logistic regression is a series of regression analyses that examines whether several variables independently predict a binary, dichotomous outcome, such as a guilty or not guilty verdict. See LAWLESS ET AL., *supra* note 116, at 299–302 (discussing logistic regressions). Statistical significance in a logistic regression model is determined by a “Wald” statistic and its corresponding p-value. The strength of the variable in the model is designated by its coefficient, “B,” which represents log odds. See ANDY FIELD, *DISCOVERING STATISTICS USING IBM SPSS STATISTICS* 765–66 (4th ed. 2013).

A p-value represents the likelihood that, if the null hypothesis were true (and there is no effect of the predictor variable on the dependent variable), we would see the result that we found in our sample. A statistically significant result is conventionally defined as a p-value below .05; marginally significant results have a p-value below .10, and highly significant results have a p-value below .01. A p-value can be conceived of as reflecting the stability of the experimental finding and (more controversially) a predictor of the likelihood that the effect found in the experiment will replicate outside of the laboratory. See *id.* at 197 (discussing the meaning of p-values).

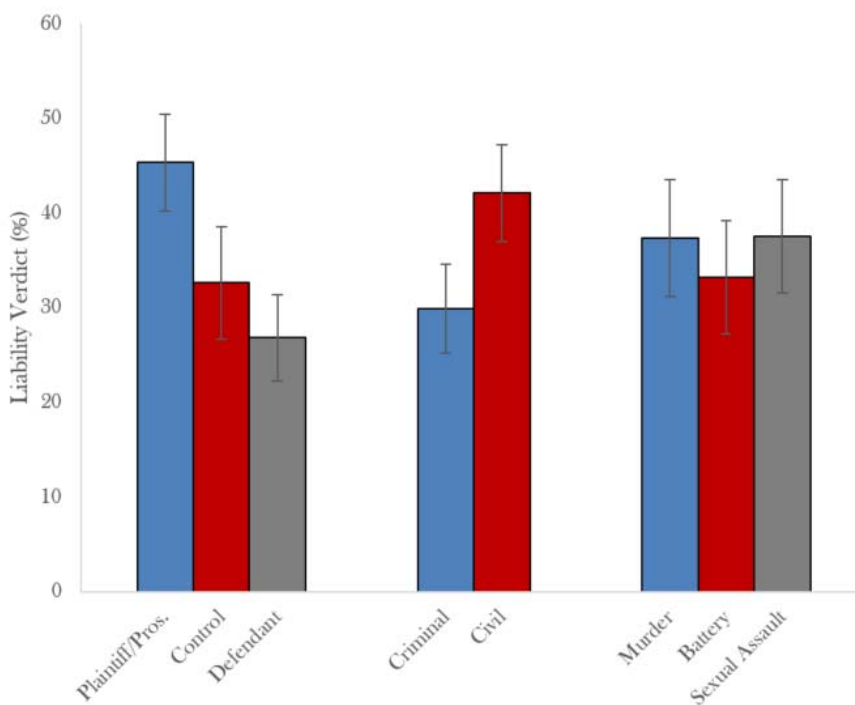
¹¹⁹ In the civil case, 42.10% of our participants found the defendant liable; 29.90% of our participants found the defendant guilty in the criminal case.

¹²⁰ The percentages of participants who found the defendant liable were as follows: 37.30% in the murder case, 33.20% in the battery case, and 37.50% in the sexual assault case (collapsing across civil and criminal legal settings).

¹²¹ When the character witness testified against the defendant, 45.30% of participants found him liable. When she testified for the defense, 26.80% of participants found him liable.

statistically significant.¹²² Conversely, character testimony from the defense decreased the liability rate from baseline; but because the control condition was already skewed toward finding the defendant non-liable, the decrease in this rate when the character witness testified for the defense did not reach statistical significance.¹²³ Graphs of these findings appear below.

Figure 1. Main Effects of Party, Legal Setting, and Case Type on Liability Verdicts



Next, we examined the robustness of the effects of the party proffering the character evidence, and the setting in which the evidence is proffered, on our participants' verdicts. We conducted a stepwise logistic regression that produced three progressive models. First, we examined the effect of our experimental manipulations on participants' liability judgments controlling for several

¹²² The defendant's liability rate increased from 32.60% of participants in the control condition (where no propensity evidence was presented) to 45.30% when a character witness testified against the defendant.

¹²³ The defendant's liability rate decreased from 32.60% of participants in the control condition to 26.80% when the character witness testified on the defendant's behalf.

demographic variables, including gender, age, political affiliation, income, education, and race. Our second model also controlled for several personality variables that correlate with verdicts, including authoritarian tendencies, an individual's need for cognition or closure, social dominance orientation, and belief in a just world.¹²⁴ Our final model also controlled for negative attitudes toward the courts and attorneys that our participants might hold.

As Table 2 below illustrates, our experimental results remain unchanged—and their effect sizes similar—across all three models: there was no effect of the type of case on participants' liability judgments, but they were affected by the party proffering the character witness and the setting in which the trial occurred. All of our models had significant explanatory power, and our most complete model explained 20% of the variance in participants' liability judgments.¹²⁵

Table 2. Effect of Character Evidence on Verdicts (Three Models)

	Model 1	Model 2	Model 3
(Constant)	(-0.35)	(0.80)	(1.59*)
Party	0.87***	0.92***	0.97***
Setting	0.56***	0.52***	0.53***
Case Type			
Murder	-0.03	0.04	0.01
Battery	0.21	0.26	0.21
Demographics			
Gender	0.13	0.18	0.21
Age	0.04***	0.03***	0.03***
Politics	-0.11	0.36*	0.37*
Income	-0.02	0.00	0.00
Education	-0.66*	-0.58	-0.62*
Race	-0.46**	-0.39**	-0.41**
Individual Differences			
Authoritarian		-0.46***	-0.40***
Need Cog.		0.06	0.08

¹²⁴ For more details on these personality controls, see *supra* note 117.

¹²⁵ Some of our control variables, including participants' age, race, authoritarian personality type, and attitudes toward the courts independently predicted their willingness to find the defendant liable. These are interesting findings in their own right, but are not germane to the current experiment.

Need Closure		0.14	0.17
Dominance		-0.16	-0.15
Just World		-0.09	0.10
Legal Attitudes			
Courts			-0.40***
Attorneys			-0.11
Model χ^2	72.43***	104.25***	115.53***
Pseudo-R ²	.13	.19	.21
N	711	711	711

*Note: Asterisks denote statistically significant effects in each model: *** signifies $p < .01$, ** signifies $p < .05$, * signifies $p < .10$ (marginal significance). Verdicts were coded as “0” for liable and “1” for not liable. Sexual Assault served as the comparison category for the Case Type variable, Plaintiff/Prosecution served as the comparison for the Party variable, and Criminal Case served as the comparison for the Setting variable. Coefficients in this logistic regression represent log odds.*

Our final analysis focused on our participants’ willingness to find the defendant liable, their judgments of the likelihood that the defendant committed the act for which he was accused, and their confidence in those judgments.¹²⁶ We included in our analysis the same predictor variables that we included in our final logistic regression model with respect to our participants’ verdicts.¹²⁷

¹²⁶ In technical terms, we conducted a 2 (party proffering the character witness: plaintiff/prosecutor vs. defendant) x 2 (legal setting: criminal vs. civil) x 3 (case type: shooting vs. beating vs. sexual assault) multivariate analysis of covariance (MANCOVA) on participants’ willingness to find the defendant liable, perceived likelihood that the defendant committed the act, and confidence in their judgments.

Our control variables are termed “covariates.” An analysis that includes these covariates would be termed an “analysis of co-variance,” or “ANCOVA,” which is a close cousin of the analysis of variance (ANOVA) linear model. *See, e.g.,* Andrew C. Porter & Stephen W. Raudenbush, *Analysis of Covariance: Its Model and Use in Psychological Research*, 34 J. COUNSELING PSYCHOL. 383, 383 (1987). Both an ANOVA and a MANOVA are statistical tests, which produce Fisher’s F-statistics, that examine whether the means of different groups are statistically different or statistically equal.

A MANCOVA is a special type of analysis of covariance where multiple dependent variables—which are at least moderately correlated with each other—are analyzed in tandem to reduce the likelihood of false positives (“type I error”). *See, e.g.,* Russell T. Warne, *A Primer on Multivariate Analysis of Variance (MANOVA) for Behavioral Scientists*, 19 PRAC. ASSESSMENT RES. & EVALUATION 1, 2 (2014).

¹²⁷ Although our experimental design is factorial, such that each participant was randomly exposed to a trial that contained one legal setting, case type, and party that proffered the character witness, we tested our hypotheses in a main effects model. We did so because we had clear, theoretical predictions with respect to the main effects of these variables on our dependent measures. In contrast, we had no a priori hypotheses regarding whether these variables would interact with one another.

To examine the robustness of our findings, we also conducted the analysis as a series of independent ANOVAs omitting the covariates from the models. Our results were unchanged.

Our hypotheses were confirmed. Participants were more willing to find the defendant liable in a civil setting than in a criminal setting,¹²⁸ and they were more willing to find the defendant liable when the character witness testified for the prosecution (or plaintiff) than when she testified for the defense.¹²⁹ Also as predicted, we found no effect of the type of case in which the character witness testified.¹³⁰

As predicted, we found the same pattern of results regarding our participants' perceptions of the likelihood that the defendant committed the act: the propensity evidence affected our mock jurors' judgments depending on who called the character witness.¹³¹ This time, however, the legal setting had no effect on our participants' judgments.¹³² This suggests that the effect of the trial setting on participants' verdicts (and their willingness to find the defendant liable) is explained by different standards of proof in civil and criminal trials. A graph of the means with respect to our "party" and "legal setting" variables appears below.

¹²⁸ M -civil = 3.25, SE = 0.11; M -criminal = 3.81, SE = 0.10; $F(1, 691) = 17.91, p < .001, \eta^2_p = .03$.

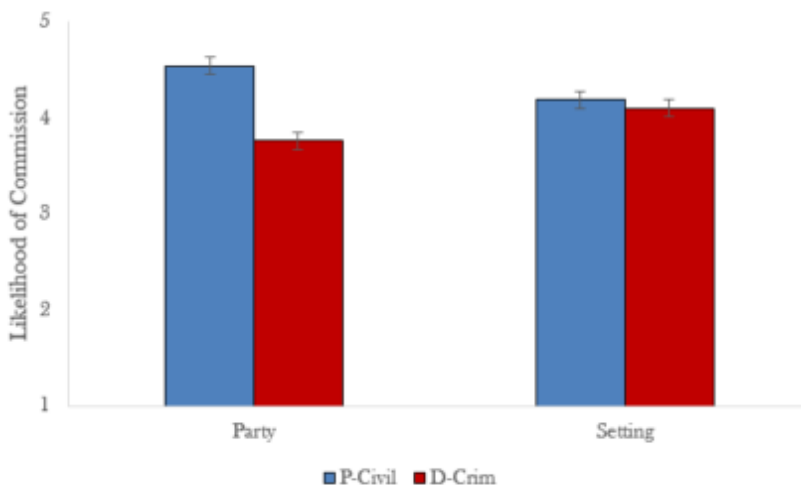
¹²⁹ M -pros/plaintiff = 3.92, SE = 0.11; M -defendant = 3.06, SE = 0.11; $F(1, 691) = 33.69, p < .001, \eta^2_p = .05$.

¹³⁰ $F(2, 691) = 0.59, p = .556, \eta^2_p = .00$.

¹³¹ M -pros/plaintiff = 4.54, SE = 0.09; M -defendant = 3.76, SE = 0.09; $F(1, 691) = 46.35, p < .001, \eta^2_p = .06$.

¹³² $F(1, 691) = 0.62, p = .430, \eta^2_p = .00$.

Figure 2. Effects of Party and Legal Setting on Perceptions of Liability



Finally, we had no strong predictions with respect to our participants' confidence in their judgments of the defendant's liability. We found one statistically significant result: participants were less confident in their judgments when the character witness testified for the defense than when the character witness testified for the plaintiff or prosecution.¹³³ The regression coefficients for our models appear in Table 3 below.

Table 3. Effect of Character Evidence on Legal Outcomes and Confidence

	Likely Convict	Likely Commit	Confidence
Party	-0.20***	-0.25***	-0.22***
Setting	0.15***	0.03	0.04
Case Type			
Murder	0.00	0.03	0.00
Battery	-0.04	-0.04	-0.04
Demographics			
Gender	0.03	0.06	0.02
Age	-0.03***	-0.04	-0.11***

¹³³ M -pros/plaintiff = 3.94, SE = 0.11; M -defendant = 3.06, SE = 0.11; $F(1, 691) = 39.22$, $p < .001$, $\eta^2_p = .05$.

Politics	0.04	0.07	0.07*
Income	0.01	0.01	-0.02
Education	-0.03	-0.03	-0.05
Race	-0.09***	-0.04	-0.04
Individual Differences			
Authoritarian	0.11**	0.10*	-0.16***
Need Cog.	-0.05	0.02	0.02
Need Closure	-0.06	-0.04	-0.02
Dominance	0.05	-0.01	0.02
Just World	-0.03	-0.04	-0.02
Legal Attitudes			
Courts	0.21***	0.20***	-0.19***
Attorneys	0.10***	0.08**	-0.11***
Model Sig. (F-Test)	8.97***	5.35***	7.29***
Model R ²	.16	.09	.13
N	710	711	710

*Note: Asterisks denote statistically significant effects in each model: *** signifies $p < .01$, ** signifies $p < .05$, * signifies $p < .10$ (marginal significance). Party was coded as "0" for plaintiff/prosecutor and "1" for defense, whereas Setting was coded as "0" for criminal and "1" for civil. Sexual Assault served as the comparison category for the Case Type variable.*

b. Main Analysis II: Specific Judgments in Study 1

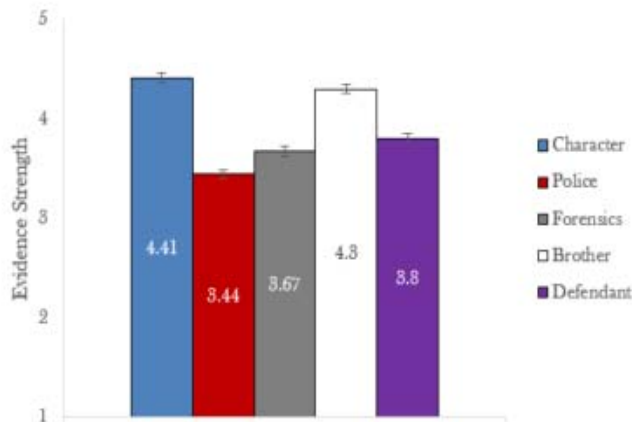
We next examined how our participants viewed the propensity evidence included at the trial. First, our participants evaluated the strength of the following evidence: the police officer's testimony regarding the scene of the incident, the forensic testimony regarding the ski cap, the brother's testimony regarding the defendant's opportunity to commit the crime, the co-worker's testimony regarding the defendant's character, and the defendant's alibi testimony. An ANOVA revealed that jurors did not perceive the evidence presented at the trial to be equally strong.¹³⁴ Post hoc analyses¹³⁵ confirmed that

¹³⁴ M -police = 3.44, SE = 0.07; M -forensics = 3.67, SE = 0.07; M -brother = 4.30, SE = 0.06; M -character = 4.41, SE = 0.06; M -defendant = 3.80, SE = 0.05; $F(2.86, 2060.60) = 58.00$, $p < .001$, $\eta^2_p = .08$. Because the repeated measures data violated the assumption of sphericity (Mauchly's $W = 0.51$, $p < .001$), we applied a Greenhouse-Geisser correction. For the definition and explanation of an ANOVA, see *supra* note 126. A repeated measures ANOVA, also referred to as a within-subjects design, compares multiple responses by the same participant to the experimental stimuli.

¹³⁵ All p -values for the comparisons were less than .001, with the exception of the comparison of the character evidence with the brother's testimony ($p = .694$). An omnibus test, such as an analysis of variance, indicates only whether one of the group's means differs from the others. A statistically-significant omnibus test, however, does not indicate which mean (or means) deviate from the others. Statisticians have created several post hoc tests to make that determination. In this study, we used the "least significant difference" post hoc test

the character evidence was perceived to be stronger than the police officer's testimony, the forensic testimony, and the defendant's alibi testimony. The only piece of evidence that was deemed equally strong was the brother's testimony regarding the defendant's opportunity to commit the crime. A graph of the statistical results appears below.

Figure 3. Participants' Perceptions of Evidence Strength



These results suggest that mock jurors paid attention to the character evidence and viewed it as a strong piece of evidence at the trial.¹³⁶ The results do not tell us, however, how influential the evidence was to our participants' verdicts. We predicted that although our participants may find the propensity evidence to be strong, they will not give propensity evidence disproportionate weight in their legal judgments compared to other evidence. Thus, we hypothesized that our participants would not consider the character witness's testimony as the most—or even the second most—useful and important evidence that they encountered at the trial.

To examine this hypothesis, we presented our participants with the same five pieces of evidence and asked them to rank the evidence, from 1 (“most important

because we employed a (theoretically justified) planned comparisons approach. Even adjusting for family-wise error under a more conservative procedure, our results did not change.

¹³⁶ It is unsurprising that participants viewed the character evidence as strong, because the witness was a friend of the defendant for many years. Also, to ensure that the case that participants read about was a close case legally, we intentionally created forensic evidence and police testimony that was open to interpretation and critique.

to my decision”) to 5 (“least important to my decision”). We evaluated our participants’ rankings via the Friedman test, a non-parametric repeated-measures statistical technique.¹³⁷ The test revealed that our participants differed significantly with respect to the ranking that they gave each piece of evidence.¹³⁸

To determine the nature of that difference, we conducted a Wilcoxon signed rank test (with a Bonferroni correction) to determine the following: (1) which pieces of evidence were ranked differently from the character evidence that our participants encountered; and (2) whether our participants deemed those pieces of evidence more or less important than the character evidence.¹³⁹

A box and whisker plot of our participants’ rankings of each piece of evidence appears below. The whiskers represent the upper and lower bounds of their rankings (each piece of evidence was rated a “1” or a “5” by at least one participant), and the two boxes represent the 25th and 75th percentiles (with the line separating them representing the median rank for each piece of evidence).¹⁴⁰ We also included the average rank for each piece of evidence as a rectangular bullet within each box.

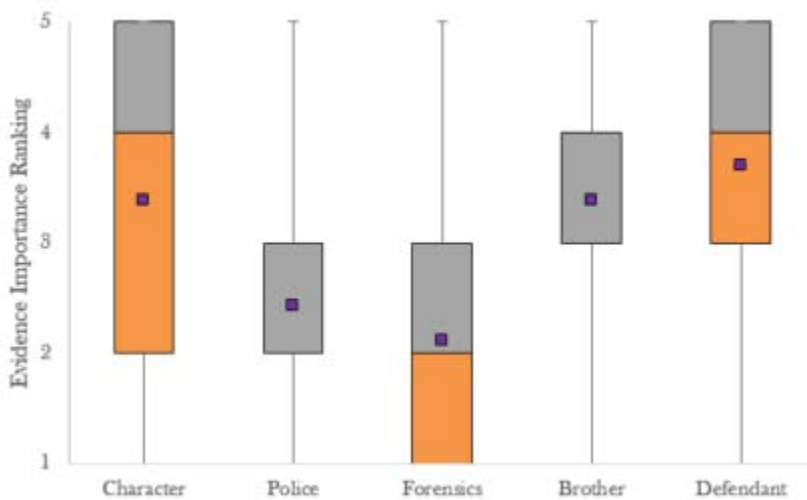
¹³⁷ The Friedman test is a non-parametric statistical test, similar to the repeated measures ANOVA, that is used to detect differences in treatments across multiple responses from the same participant. *Friedman Test in SPSS Statistics*, LAERD STAT., <https://statistics.laerd.com/spss-tutorials/friedman-test-using-spss-statistics.php> (last visited Nov. 25, 2018); see also Milton Friedman, *A Correction*, 34 J. AM. STAT. ASS’N 109, 109 (1939).

¹³⁸ $\chi^2(4) = 542.84, p < .001$.

¹³⁹ The Wilcoxon signed-rank test is a non-parametric statistical test used to compare repeated measurements on a single sample to assess whether their population mean ranks differ. See Frank Wilcoxon, *Individual Comparisons by Ranking Methods*, 1 BIOMETRICS BULL. 80, 80 (1945).

¹⁴⁰ If only one box appears in the graph, the 25th percentile is also the median.

Figure 4. Relative Rankings of Evidence Importance



As the graph above suggests, our participants perceived both the forensic testimony and the police officer's testimony as more important to their verdicts than the character witness's testimony.¹⁴¹ The character witness's testimony tied with the brother's testimony as a distant third in terms of importance.¹⁴² The only piece of evidence that our participants deemed less useful and important than the character witness's testimony was the testimony of the defendant himself.¹⁴³

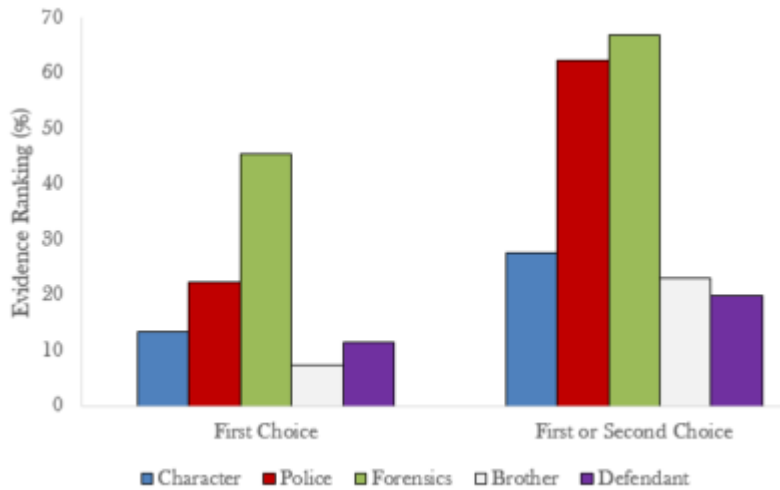
The analysis remains the same if we instead analyze the proportion of our participants who listed the propensity evidence as the most important piece of evidence or even the *second*-most important piece of evidence. As the graph below indicates, the vast majority of participants considered the forensic evidence to be the most important piece of evidence in deciding the case. When the analysis is expanded to participants' first or second choices, the vast majority of participants focused on the forensic evidence and the police officer's testimony. The character evidence, the brother's testimony, and the defendant's testimony remained a distant third.

¹⁴¹ $Z(\text{forensics}) = -13.25, p < .001$; $Z(\text{police}) = -11.00, p < .001$.

¹⁴² $Z(\text{brother}) = -0.25, p = .804$.

¹⁴³ $Z(\text{defendant}) = -3.83, p < .001$.

Figure 5. Evaluations of the Most (and Second-Most) Important Trial Evidence



4. Discussion of Findings in Study 1

Study 1 yielded several important findings bearing on the way in which mock jurors evaluate propensity evidence. Consistent with our experimental hypotheses, the results support the view that jurors consider character evidence in rendering their verdicts. They also support the view that jurors thoughtfully evaluate propensity evidence.

When the plaintiff (or prosecutor) proffered character evidence against the defendant, the percentage of participants who found the defendant liable increased roughly thirteen points on average. Conversely, evidence of the defendant's positive character, proffered by the defense, reduced judgments of the defendant's liability by roughly six points on average.¹⁴⁴ We found these results regardless of the legal setting in which the evidence was proffered (criminal vs. civil), and the results replicated across several different types of cases, including a shooting, a battery, and an attempted sexual assault.

The data suggest that character evidence does not play a disproportionate role in (1) our participants' verdicts, (2) their judgments of their willingness to find the defendant liable, or (3) their perceptions of the likelihood that the

¹⁴⁴ This finding is perhaps even more impressive in light of the fact that all of the cases, at baseline (that is, without the introduction of propensity evidence), favored the defense.

defendant committed the acts for which he was accused. Even when the prosecution (or plaintiff) produced the propensity evidence against the defendant, a majority of our participants still voted not to find the defendant civilly or criminally liable, and they were unwilling to weigh the propensity evidence more than they weighed the forensic evidence or the police officer's testimony. Indeed, they weighed the character evidence more heavily than only one other piece of evidence at trial: the defendant's self-serving, uncorroborated alibi testimony.

Study 1 suggests that mock jurors consider character evidence relevant in rendering their verdicts, but that they are cautious with respect to the weight they place on the evidence. Study 2 builds on these findings by examining more deeply the degree of care with which mock jurors evaluate propensity evidence.

B. Study 2: Testing Decisional Accuracy

Our second study serves two important purposes. First, we use an additional independent sample to replicate the main findings from Study 1: that mock jurors attend to propensity evidence at trial but do not afford it unreasonable probative weight. Second, we extend this finding by examining our mock jurors' sensitivity to accuracy-enhancing and accuracy-diminishing features of propensity evidence.

We evaluated our mock jurors' sensitivity to character evidence by manipulating three different dimensions of the information on which the character witness based her opinion of the defendant: the frequency of the defendant's prior acts, the length of time that had passed between the defendant's prior acts and the act for which he was currently accused, and the similarity of those acts to the acts underlying the current accusation against the defendant. Because we found in Study 1 that the legal setting (a civil or criminal action) and the type of case did not affect our participants' evaluation of propensity evidence, we evaluated our participants' sensitivity to the evidence in the context of the criminal second-degree murder case from Study 1.¹⁴⁵

If participants are not adept at evaluating character evidence, or if they are not sensitive to differences in the acts that form the basis of the character

¹⁴⁵ Additionally, because this Article is examining whether jurors make sensible decisions regarding *inadmissible* character evidence, the propensity witness always testified for the prosecution against the defendant (in a situation in which the defendant had not opened the door to such testimony). Under the mercy rule, *see* discussion *supra* Section I.B., a defendant is already allowed to proffer propensity evidence of a pertinent character trait in a criminal proceeding.

witness's testimony, we should see no differences in their attitudes toward the evidence presented at the trial. But if jurors are attentive and *sensitive* to such information (as past research suggests they might be), we would expect them to evaluate the evidence as stronger when the acts underlying the character witness's testimony were frequent rather than rare, recent rather than old, and similar to the current accusation rather than different from it. Our methods for testing these hypotheses, and the results that we found, appear below.

1. *Participants, Procedures, and Measures in Study 2*

We recruited 246 participants for Study 2, via Amazon Mechanical Turk.¹⁴⁶ The logistics for recruiting our participants mirrored the procedure from Study 1. Our participants again were a representative cross section of the population, and we provide sample statistics in Table 4 below.

Table 4: Participant Demographics (Study 2)

	%	<i>N</i>
Age (Median: 34.00)		
< 30	34.8	83
30-39	41.6	99
40-49	11.2	27
50-59	09.2	22
60-76	03.2	08
Gender		
Male	50.2	120
Female	49.8	119
Race		
Caucasian	75.4	178
African-American	08.5	20
Hispanic	05.5	13
Asian	08.8	21
Other	01.7	04
Education		
High School	08.8	21
Some College	32.2	77

¹⁴⁶ As in Study 1, participants were a representative sample from throughout the United States.

College	46.4	111
Master's	09.6	23
Ph.D. or Professional	02.9	07
Political Affiliation		
Very Conservative	08.4	20
Conservative	18.5	44
Moderate	22.3	53
Liberal	33.6	80
Very Liberal	17.2	41
Other		
Income		
Less than \$30,000	30.3	65
\$30,000 - \$49,999	26.7	68
\$50,000 - \$69,999	17.6	42
\$70,000 or greater	25.4	64

Study 2 followed many of the protocols used in Study 1. Participants read about a fatal shooting at an upscale mall. Because we found no effect of the type of case on our participants' verdicts or impressions of the character evidence in Study 1,¹⁴⁷ all participants in Study 2 read the criminal case in which the government charged the defendant with second-degree murder.¹⁴⁸ With the exception of the character witness, the remaining witnesses and their testimony were identical in all experimental conditions in Study 2. Instead, we varied several facets of the character witness's testimony in Study 2.

As an initial matter, the character witness always testified for the prosecution and provided against the defendant propensity evidence that currently would be inadmissible under the Federal Rules of Evidence. We then varied three facets of the character witness's testimony in Study 2: the frequency of the defendant's acts that underlie the testimony, how recently those acts occurred, and the similarity of the underlying acts to the current dispute.

In each experimental condition, the character witness (who always testified against the defendant) stated that the defendant was known as "a bad guy"

¹⁴⁷ See *supra* note 120 and Figure 1.

¹⁴⁸ Put another way, we made this decision because Study 1 revealed that the effects of character evidence on participants' verdicts were statistically significant regardless of whether the case was a murder, a battery, or a sexual assault—and regardless of whether the case was a civil or criminal matter.

throughout the community and had a reputation for breaking the law. Pursuant to procedures analogous to the procedures outlined in the Federal Rules of Evidence,¹⁴⁹ the character witness revealed the basis for her testimony during cross-examination. In accordance with our first manipulation, half of our participants learned that the character witness's testimony was based on either five violent incidents in the past (the "frequent" condition) or just one incident (the "rare" condition).

We also varied the length of time between the alleged commission of the murder and the act (or acts) that provided the basis for the character witness's testimony against the defendant. Half of the participants learned that the character witness based her testimony on acts performed by the defendant over the past year (the "recent" condition); half of our participants learned that five years had passed between the defendant's prior act (or acts) and the alleged commission of the murder.

Finally, we varied the similarity between the act (or acts) that provided the basis for the character witness's testimony against the defendant and the crime of which the defendant was accused. In the "similar" condition, the character witness based her testimony on the defendant's previous firing of a gun at pedestrians in a park. In the "different" condition, the character witness based her testimony on an incident (or incidents) in which the defendant was drunk and disorderly at a local bar.¹⁵⁰

Study 2, like Study 1, was subject to a "factorial" design, such that each participant was randomly assigned to one frequency condition, one time condition, and one similarity condition. To test our participants' sensitivity to these aspects of the character evidence, we asked participants questions similar to those that we posed in Study 1.¹⁵¹ This time, however, we focused predominantly on their impressions of the character evidence, the strength of the prosecution's case, and the likelihood that they would convict the defendant. We

¹⁴⁹ See FED. R. EVID. 405(a) (requiring, under most circumstances, that character evidence take the form of an opinion or testimony regarding a person's general reputation; specific instances of conduct are generally reserved for cross-examination). The procedure used in this vignette is "analogous to the procedures outlined in the Federal Rules of Evidence" because under the current Rules, the prosecution's character witness would be prohibited from testifying against the defendant unless the defendant invoked the mercy rule provisions of FRE 404(a)(2).

¹⁵⁰ This condition was purposely designed so that, although drunk and disorderly behavior is sufficiently different from the shooting for which the defendant is accused, it still bears on the defendant's capacity for violence. The testimony is therefore pertinent to the current case against the defendant. See, e.g., FED. R. EVID. 404(a)(2)(A) (allowing into evidence a criminal defendant's *pertinent* character trait).

¹⁵¹ See *supra* notes 116–17 and accompanying text.

then posed several demographic and personality questions to our participants before we concluded the experiment.

2. *Results of Study 2*

This section proceeds in two parts. First, it examines mock jurors' sensitivity to the factors that separate stronger propensity evidence from weaker propensity evidence: the frequency of the acts underlying the character witness's testimony, the amount of time that has passed since those acts occurred, and the similarity of the past acts to the current accusation against the defendant. Second, to the extent that jurors are sensitive to these features of propensity evidence, we examine statistically whether this sensitivity affects the likelihood that our mock jurors will find the defendant liable for the crime of which he is accused. We report our results below.

a. *Main Analysis of Results in Study 2*

To test the jurors' sensitivity to factors that affect the strength or weakness of propensity evidence, we examined the effect of the frequency of the act underlying the character witness's testimony, the length of time that had passed, and the similarity of the past act on (1) our participants' assessments of the strength of the evidence; and (2) their assessments of the strength of the evidence on the prosecutor's case.¹⁵²

The results supported our hypotheses. Participants distinguished between the different types of character evidence on all three dimensions. They found the propensity evidence less persuasive when the act underlying the character witness's testimony occurred rarely than when it was a frequent occurrence.¹⁵³ They also credited the character witness's testimony less if the act (or acts) occurred five years ago than if the act (or acts) occurred within the past year.¹⁵⁴ The strongest effect we found was with respect to the similarity of the prior act: participants found the character witness's testimony far more persuasive when the prior act was similar in kind to the act that formed the basis of the current charges against the defendant than when it was a different act (even though it still bore on the defendant's propensity for violence).¹⁵⁵ We found the same

¹⁵² We conducted a 2 (frequency: often vs. rare) x 2 (time: recent vs. old) x 2 (similarity: same vs. different) between-subjects MANCOVA on participants' assessments of the evidence strength and the strength of the prosecutor's case. We report the estimated marginal means in this section in addition to the standard error of the means.

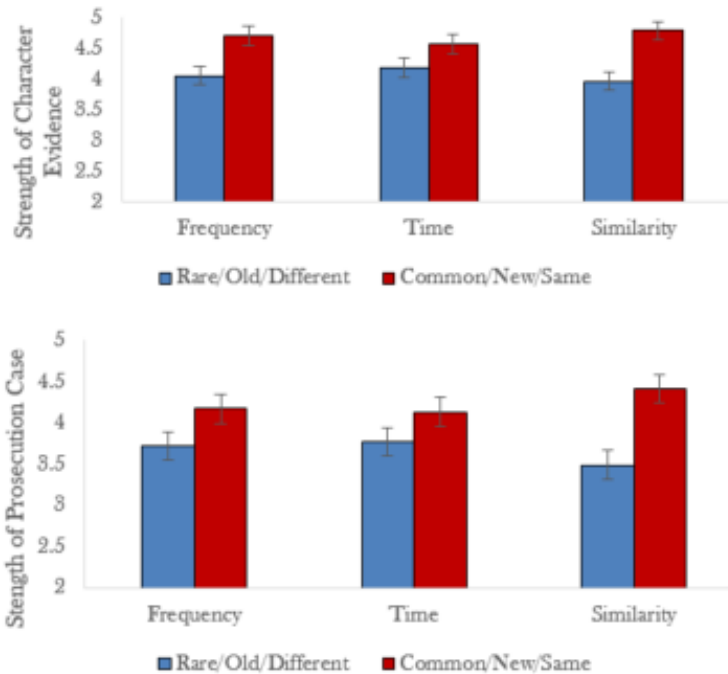
¹⁵³ M -rare = 4.06, SE = 0.15; M -common = 4.71, SE = 0.16; $F(1, 242) = 8.72, p = .003, \eta^2_p = .04$.

¹⁵⁴ M -old = 4.06, SE = 0.15; M -recent = 4.58, SE = 0.16; $F(1, 242) = 3.14, p = .078, \eta^2_p = .01$.

¹⁵⁵ M -similar = 4.80, SE = 0.15; M -common = 3.97, SE = 0.16; $F(1, 242) = 14.29, p < .001, \eta^2_p = .06$.

pattern of effects with respect to the strength of the prosecutor's case, although the effect sizes were smaller. Graphs of the estimated marginal means for each experimental condition appear below.

Figure 6. Perceived Evidence and Case Strength (on the Likert Scale) as a Function of the Frequency, Duration, and Similarity of the Past Act to the Accused Crime



Next, we examined the robustness of these effects on our participants' perceptions of the strength of the character evidence and the prosecution's case. We included in our model three sets of control variables: demographic, personality, and attitudinal. As the table below indicates, our mock jurors were robustly sensitive to accuracy-enhancing and accuracy-diminishing features of propensity evidence. Our models accounted for 17% and 20% of the variance in participants' perceptions of the strength of the prosecutor's case and the strength of the character evidence, respectively.¹⁵⁶ And most importantly, the effects of

¹⁵⁶ In other words, roughly 20% of the change in our participants' ratings of the strength of the evidence and the strength of the prosecution's case could be explained by just the factors that we included in this model.

the frequency of the underlying act, the time frame in which it occurred, and its similarity to the accused crime remained statistically significant predictors of our mock jurors' perceptions of the strength of the propensity evidence taking into account thirteen different control variables. We found similar effects with respect to the strength of the prosecution's case.

Table 5. Sensitivity to Accuracy-Enhancing Features of Character Evidence

	Evidence Strength	Case Strength
Frequency	0.20***	0.11*
Time	0.10*	0.08
Similarity	0.15**	0.17***
Demographics		
Gender	0.10*	0.03
Age	-0.06	-0.18***
Politics	0.13*	0.06
Income	-0.04	-0.01
Education	0.04	0.04
Race	-0.09	-0.07
Individual Differences		
Authoritarian	0.08	0.20*
Need Cog.	0.11*	0.04
Need Closure	0.10	0.02
Dominance	0.11	0.04
Just World	0.02	-0.04
Legal Attitudes		
Courts	0.24***	0.18**
Attorneys	0.05	-0.08
Model Sig. (F-Test)	4.63***	4.10***
Model R ²	.20	.17
N	240	240

*Note: Asterisks denote statistically significant effects in each model: *** signifies $p < .01$, ** signifies $p < .05$, * signifies $p < .10$ (marginal significance).*

b. Serial Mediation Analysis of Results in Study 2

The results from Study 2 support our experimental hypotheses. In sum, jurors appear unlikely to overvalue propensity evidence. Our results from Study 1 suggest that although jurors consider character evidence in rendering their verdicts, it does not move the percentage of liability judgments substantially in

either direction: either toward liability when the witness testifies against the defendant, or toward non-liability when the witness testifies on behalf of the defendant. Study 2 builds on these findings by demonstrating that jurors are careful when evaluating the frequency, timing, and similarity of the acts that underlie propensity evidence used at trial. What we have not yet shown, however, is whether our participants' sensitivity to differences in the frequency, timing, and similarity of propensity evidence directly affected their willingness to convict the defendant. This section examines this question through a statistical technique called a "serial mediation analysis."

A serial mediation consists of a set of regression analyses that are designed to determine the psychological processes that underlie the effect of a predictor variable on an outcome.¹⁵⁷ The psychological process (or processes) that are hypothesized to underlie the effect are termed "mediators" of the effect. A mediation analysis is designed to show that the effect of the predictor on the outcome can be explained—either fully or in part—by the psychological mediators.¹⁵⁸ A "serialized" mediation builds on this concept and involves more than one mediator.¹⁵⁹ A serialized mediation analysis tells us that a predictor variable is associated with one psychological mediator (first mediator), which is associated with *another* psychological mediator (second mediator) which, in turn, is associated with the outcome.¹⁶⁰

Because its effect size was the strongest of our three experimental manipulations, we chose to explore the "similarity" variable in our serialized mediation analysis. We constructed our model as follows: (1) the predictor variable is the similarity of the prior act to the crime for which the defendant is accused; (2) the outcome is our mock jurors' willingness to convict the defendant; (3) the first mediator involves our mock jurors' impressions of the strength of the character witness's testimony; (4) and the second mediator involves our mock jurors' perceptions of the strength of the prosecution's case.

¹⁵⁷ Mediation analysis detects "when a predictor affects a dependent variable indirectly through at least one intervening variable, or mediator." Kristopher J. Preacher & Andrew F. Hayes, *Asymptotic and Resampling Strategies for Assessing and Comparing Indirect Effects in Multiple Mediator Models*, 40 BEHAV. RES. METHODS 879, 879 (2008). The mediation analysis reported in this Article is performed using a linear regression analysis and reports unstandardized coefficients, "B," and standard errors, "SE." It also reports a "t" statistic, which determines whether the coefficients are statistically significant. A linear regression is a statistical test that estimates the independent effects of several predictor variables on a continuous dependent variable. See LAWLESS ET AL., *supra* note 116, at 29, 300–31.

¹⁵⁸ See Preacher & Hayes, *supra* note 157 (discussing the theoretical and statistical import of mediation analyses).

¹⁵⁹ *Id.*

¹⁶⁰ *Id.*

The analysis then proceeds as a series of regression analyses to determine if the effect of the similarity of the prior act on our mock jurors' willingness to convict the defendant is explained as follows: (1) the similarity (or lack of similarity) of the prior act affects our participants' perceptions of the strength of the character witness's testimony; (2) the strength of the character witness's testimony affects our participants' views of the strength of the prosecution's case; and (3) the strength of the prosecutor's case predicts our mock jurors' willingness to convict the defendant. This hypothesis is tested below.

The similarity of the defendant's prior act to the current charge affected our mock jurors' willingness to convict the defendant, such that mock jurors were less likely to convict the defendant if the underlying act was different from the current charge.¹⁶¹ As predicted, the similarity of the prior act was associated with the perceived strength of the character witness's testimony, such that our mock jurors found the witness's testimony less persuasive when the underlying act was different than when it was similar to the current charge.¹⁶² Also as predicted, mock jurors' perceptions of the strength of the character witness's testimony affected their perceptions of the strength of the prosecution's case, such that the weaker their perceptions of the character witness's testimony were, the weaker their perceptions of the prosecution's case were as well.¹⁶³ Finally, perceptions of the prosecution's case significantly predicted the degree to which mock jurors were willing to convict the defendant, such that lower perceptions of the prosecution's case were associated with a lower likelihood of convicting the defendant.¹⁶⁴ The mediation further revealed that this indirect pathway significantly accounts for the effect of the similarity of the defendant's prior act on our mock jurors' willingness to convict the defendant.¹⁶⁵ An illustration of this pathway, which includes the beta coefficients from the regression analyses that we performed, appears below.¹⁶⁶

¹⁶¹ $B = 1.13, SE = 0.27, t = 4.24, p < .001.$

¹⁶² $B = 0.83, SE = 0.22, t = 3.75, p < .001.$

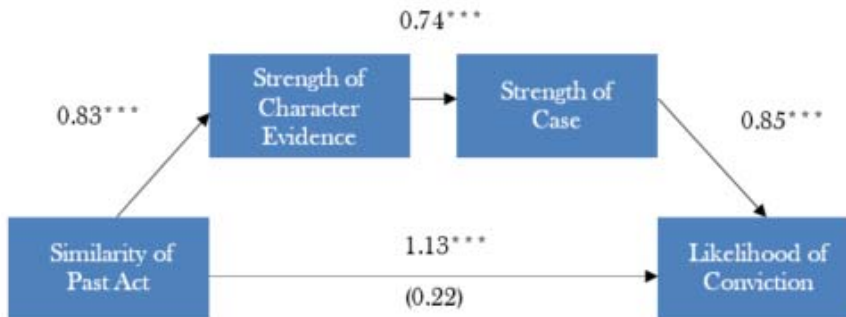
¹⁶³ $B = 0.74, SE = 0.05, t = 13.72, p < .001.$

¹⁶⁴ $B = 0.85, SE = 0.05, t = 18.77, p < .001.$

¹⁶⁵ $B = 0.52, SE = 0.14, 95\% CI [0.25, 0.81].$

¹⁶⁶ Asterisks in the mediation analysis indicate statistically significant associations.

Figure 7. Serial Mediation Investigating the Relationship Between the Similarity of a Past Act to the Accused Act and the Defendant's Likelihood of Conviction



3. Discussion

Study 2 builds on the results from Study 1 in several ways. Study 2 replicated the most important finding from Study 1: that our participants carefully attend to propensity evidence when it is produced at trial. Study 2, however, suggests that our participants do not blindly accept as determinative the inferences that follow from the use of propensity evidence. Instead, and contrary to the views of evidence policymakers and common law courts, our mock jurors were robustly sensitive to both accuracy-enhancing and accuracy-diminishing features of the propensity evidence. In other words, they appeared to distinguish between more and less probative versions of the evidence and weighed the evidence accordingly. This was true for all three variables that we manipulated in our study: the frequency of the prior act, the length of time between its commission and the commission of the charged crime, and the similarity between the prior act and the charged crime. These findings suggest that our mock jurors displayed a sophisticated degree of competency when evaluating this otherwise forbidden evidence.

Moreover, we constructed a psychological pathway to illustrate the nature of our mock jurors' competency with inadmissible character evidence as it related to the similarity of the defendant's prior conduct. When the prior act was similar to the current offense, our participants rated the weight of the character witness's testimony more strongly, which significantly affected their views of the prosecution's case. These considerations explain fully our participants' willingness to convict the defendant as a function of the similarity of his prior acts.

Study 2, paired with Study 1, provides important new information regarding mock jurors' competency with respect to evaluating propensity evidence (and the likelihood that they will reach a more accurate verdict when such evidence is admissible at trial). To legitimize propensity evidence, however, researchers must do more than demonstrate that the inclusion of such evidence has the tendency to potentially enhance a fact finder's decisional accuracy. It also requires researchers to demonstrate that the inclusion of character evidence *heightens* the public's perceptions of the procedural fairness of the fact-finding process. We examine this counterintuitive phenomenon in Study 3.

C. Study 3: Propensity & Procedural Justice

In our final study, we examine the degree to which the public is willing to legitimize trial court verdicts that rely in whole or in part on propensity evidence. In contrast to Studies 1 and 2, participants now read a vignette in which they imagined themselves as *spectators* at a murder trial at their local courthouse. They then read the same trial scenario from Study 2, but with two different experimental manipulations. First, and most importantly, at the conclusion of the trial, one of the parties attempted to admit a surprise witness who would testify to the defendant's character. After listening to each party's arguments, the judge then ruled the proposed character evidence either admissible or inadmissible. Second, because prior research suggests that jurors sometimes have differing attitudes toward a court's admissibility decisions based on the identity of the party that proffers the evidence,¹⁶⁷ we manipulated whether the character witness was proffered by the prosecution (against the defendant) or by the defense (to demonstrate the defendant's good character). The experiment then examined whether the inclusion of the character witness's testimony increased or decreased participants' perceptions of the trial's accuracy, the fairness and legitimacy of the evidence-gathering process, and participants' willingness to legitimize the court's ultimate verdict.

If, as prior research suggests, the public believes that character evidence is relevant and helpful in rendering legal verdicts, several results in Study 3 would follow. First, and counterintuitively, the public's perceptions of the court's decisional accuracy should increase when the judge *admits* the propensity evidence compared to when the propensity evidence is ruled inadmissible. Second, to the extent that the public perceives propensity evidence as helpful to the fact finder, jurors' perceptions of the trial's procedural fairness should also

¹⁶⁷ See Sevier, *supra* note 74, at 1999–2000 (finding that jurors differentially delegitimized trials in which either the prosecutor's or the defense's evidence was admitted).

increase when the evidence is admitted. Finally, the public's perceptions of the court's ability to reach an accurate verdict—as well as their perceptions of the fairness and legitimacy of the fact-gathering process by which that verdict is attained—should predict their willingness to legitimize the trial court's verdict. The following section reports our methodology and results.

1. Participants, Procedures, & Measures in Study 3

We recruited 241 participants for Study 3, via Amazon Mechanical Turk.¹⁶⁸ The logistics for recruiting our participants mirrored the procedure from Studies 1 and 2. Our participants again were a representative cross section of the population, and we provide sample statistics in Table 6 below.

Table 6: Participant Demographics (Study 3)

	%	N
Age (Median: 34.00)		
< 30	29.46	71
30-39	37.34	90
40-49	16.18	39
50-59	12.04	29
60-72	04.98	12
Gender		
Male	44.58	107
Female	55.42	133
Race		
Caucasian	79.08	189
African-American	09.62	23
Hispanic	04.60	11
Asian	05.86	14
Other	00.84	02
Education		
High School	15.00	36
Some College	27.50	66
College	45.42	109
Master's	10.00	24
Ph.D. or Professional	02.08	05
Political Affiliation		
Very Conservative	04.98	12

¹⁶⁸ As in Study 1, participants were a representative sample from throughout the United States.

Conservative	18.26	44
Moderate	31.54	76
Liberal	29.88	72
Very Liberal	14.11	34
Other	01.23	03
Income		
Less than \$30,000	26.25	63
\$30,000 - \$49,999	25.42	61
\$50,000 - \$69,999	20.42	49
\$70,000 or greater	27.91	67

Study 3 followed the protocols of Studies 1 and 2, but with several important differences. This time, participants were asked to imagine themselves as spectators observing a trial, rather than jurors. As in Study 2, we used only the criminal second-degree murder scenario. Participants therefore learned from the prosecutor's statement that the defendant allegedly murdered the victim in the early morning hours at an upscale mall. The prosecutor suggested that the evidence would show that the victim died during a botched cocaine sale. The defense's statement, as in Study 2, focused on the circumstantial nature of the evidence and asserted that the prosecutor would produce no compelling evidence to support the claim that the murder occurred in the context of a cocaine transaction.

The prosecutor then presented the testimony of the same three witnesses who testified in Studies 1 and 2: the police officer who arrived at the scene, the forensic analyst who conducted tests on the murder weapon, and the defendant's brother who provided evidence bearing on the defendant's opportunity to commit the crime. This time, however, the prosecutor then rested her case, and the defense attorney called only the defendant to the stand.

Next, our experimental manipulations unfolded. Participants were told that the trial adjourned for the day and that closing arguments would begin the next morning. Participants were also told, however, that one of the parties made a surprise request to the judge that morning. Half of our participants learned that the prosecutor moved to proffer additional testimony to the jury based on recently discovered information about the defendant. The remaining participants learned that the defense moved to proffer the additional testimony.

In both versions of the experiment, the surprise evidence came in the form of a character witness who would testify about the defendant's propensity for committing the crime. If the character witness testified for the prosecution, the witness stated that the defendant had a reputation for being unsavory and "a bad

guy” who is violent. If the character witness testified for the defense, he stated that the defendant is a “good, non-violent guy” who has often acted to improve the local community.

We then manipulated whether the character evidence was admitted successfully by varying whether the judge granted or denied the motion. After hearing the arguments, the judge either admitted the evidence, at which point the participants read that the witness testified in front of the jury, or the judge excluded the evidence, at which point the participants learned that the witness would not testify.

Once the judge made his ruling, all participants read the parties’ closing arguments and read the instructions that were presented to the jury. Participants then answered several questions regarding everything they had observed. The questions covered three topics: (1) participants’ impressions of the fairness and legitimacy of the judge’s decision to admit or exclude the character evidence;¹⁶⁹ (2) their impressions of the likelihood that the jury would reach an accurate verdict in light of the judge’s decision to admit or exclude the evidence; and (3) their willingness to legitimize the court’s verdict in light of the evidence that was presented at the trial.¹⁷⁰ As in Study 1 and Study 2, we also collected information related to certain personality variables and demographic information. After participants answered these questions, we thanked them for their time, debriefed them with respect to the experimental hypotheses, and concluded the study.

2. *Results of Study 3*

This section proceeds in two parts. First, it reports the main results of the study: how the judge’s decision to admit or exclude propensity evidence affected

¹⁶⁹ We posed four questions to measure perceived accuracy: (1) in light of the judge’s evidentiary decision, how likely is it that the jury will reach an accurate decision in this case? (2) in light of the evidentiary decision in this case, how likely is it that the court will reach the right answer? (3) in light of the judge’s ruling, how likely is it that the court will uncover the true facts that underlie this proceeding? and (4) in light of the judge’s decision, how likely is it that the court will discover the truth of what happened?

We posed three questions to measure perceived fairness of the judicial process: (1) how fair was it to exclude the propensity evidence? (2) was the procedure that the court used to decide what evidence could come in at trial unbiased? and (3) did the court’s procedure for deciding what evidence could be admitted align with your values?

A principal component analysis revealed that these sets of questions measured different psychological constructs and, when each set of questions was averaged together, composed two different, reliable scales (Cronbach’s alpha of 0.95 for accuracy and 0.89 for fairness, and they jointly explained 83.90% of the variance).

¹⁷⁰ We posed five different questions with respect to the legitimacy of the decision to admit or exclude the character evidence and with respect to the legitimacy of the trial overall.

our participants' perceptions of the accuracy, procedural fairness, and legitimacy of the trial. Second, it reports a path analysis that explores how perceptions of the accuracy and fairness of the judge's decision affected our mock jurors' willingness to legitimize the legal tribunal.

a. Main Analysis of Results in Study 3

We hypothesized that admitting propensity testimony into evidence will *increase* laypeople's perceptions of the accuracy of the legal tribunal. Moreover, we expected that our participants would view the judge's exclusion of propensity evidence as less fair than if the judge had admitted the evidence. We therefore tested whether (1) the judge's admissibility ruling and (2) the identity of the party that proffered the propensity evidence affected our participants' views of the tribunal's ability to reach an accurate decision, the fairness of the procedure by which it reached its evidentiary ruling, and our participants' willingness to ultimately legitimize the trial verdict. To do so, we conducted a 2 (ruling: admissible vs. excluded) x 2 (party: prosecutor vs. defendant) MANCOVA on our participants' perceptions of the trial's accuracy and their perceptions of the fairness of the judge's admissibility decision.

The results confirmed our hypotheses. As expected, it made no difference whether it was the prosecutor or the defense that produced the surprise witness;¹⁷¹ the judge's ruling, however, affected the perceived accuracy, fairness, and legitimacy of the judge's admissibility decision. Our participants perceived the judge's decision to be fairer when the judge *admitted* the evidence than when she excluded it.¹⁷² Similarly, our participants perceived the admissibility decision as more legitimate when the character witness was *allowed* to testify than when she was prevented from testifying.¹⁷³ Graphs illustrating the means for our participants' perceived fairness and legitimacy, as a function of the judge's evidentiary ruling, appear below.¹⁷⁴

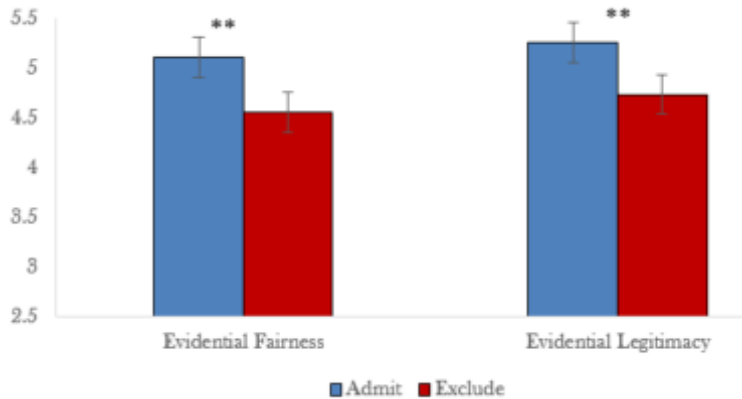
¹⁷¹ All F -values < 2.00, all p -values > .05.

¹⁷² $F(1, 218) = 5.35, p = .022, \eta^2_p = .02$.

¹⁷³ $F(1, 218) = 4.31, p = .039, \eta^2_p = .02$.

¹⁷⁴ Perceptions of the fairness and legitimacy of the judge's admissibility decision were measured as index variables on a seven-point Likert scale.

Figure 8. Perceptions of the Fairness and Legitimacy of the Tribunal’s Decision to Admit or Exclude Character Evidence

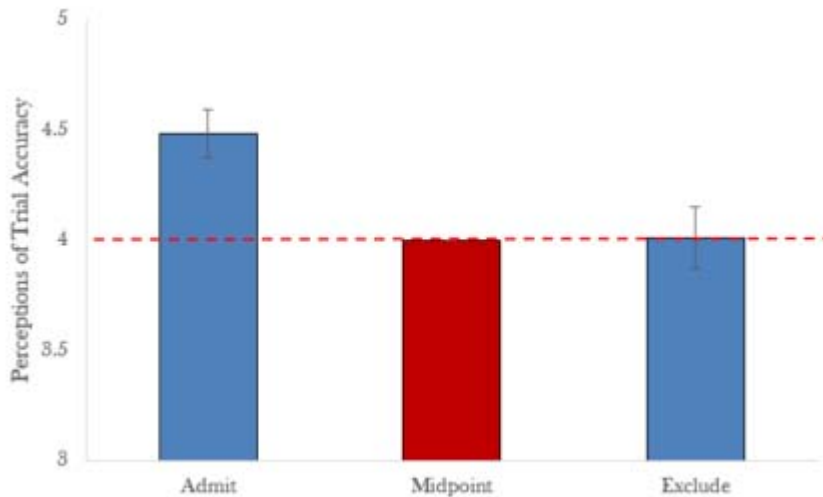


As with our participants’ perceptions of the fairness and legitimacy of the judge’s evidentiary ruling, the ruling also affected our participants’ views of the court’s ability to reach an accurate decision. Our participants believed that the court would reach a *less* accurate decision when it excluded the propensity evidence than when it admitted the evidence.¹⁷⁵ To further evaluate this finding, we compared, against the midpoint of the scale, the mean accuracy rating when the evidence was accepted and when it was excluded. Because the seven-point scale was anchored at “not at all accurate” (1) and “highly accurate” (7), the midpoint (4) would indicate a neutral view of the court’s accuracy, whereas a score statistically above (4) would indicate an increase in accuracy.¹⁷⁶ We conducted a one-sample t-test comparing the midpoint of the scale to the mean perceived accuracy levels in the “propensity evidence admitted” experimental condition and the mean perceived accuracy levels in the “propensity evidence excluded” condition. The results appear in the graph below.

¹⁷⁵ $F(1, 218) = 4.22, p = .041, \eta^2_p = .02$.

¹⁷⁶ A score statistically below a 4 would therefore indicate a decrease in accuracy.

Figure 9. Perceptions of Fact Finder Accuracy as a Function of the Admissibility of Character Evidence



As illustrated in the graph above, when the propensity evidence was excluded, participants were neutral with respect to the effect of the judge's admissibility decision on the ability of the court to reach an accurate verdict.¹⁷⁷ But when the evidence was admitted, participants believed, to a statistically significant degree, that the decision would *increase* the likelihood that the court would reach an accurate verdict having considered the propensity evidence.¹⁷⁸

Finally, as in Studies 1 and 2, we examined the robustness of the effects of our experimental manipulations accounting for thirteen demographic, personality, and attitudinal variables. As illustrated in the table below, the effects were robust.¹⁷⁹ Taking these variables into account, our participants still believed that admitting the propensity evidence (either against the defendant or in his defense) would increase (1) the ability of the court to reach an accurate verdict, and (2) the fairness and legitimacy of the process by which the court rendered that verdict. Notably, our model of the court's ability to accurately

¹⁷⁷ M -exclude = 4.02, SD = 1.50, $t(117)$ = 0.17, p = .867.

¹⁷⁸ M -admit = 4.48, SD = 1.20, $t(122)$ = 4.39, p < .001.

¹⁷⁹ The table reports the standardized regression coefficients for the variables in each model.

render its verdict explained over 30% of the variance in our participants' responses.¹⁸⁰

Table 7. Effect of Character Evidence Ruling on Perceptions of the Trial

	Trial Accuracy	Ev. Fairness	Ev. Legitimacy
Ruling	-0.11**	-0.15**	-0.13**
Party	-0.08	-0.08	-0.06
Demographics			
Gender	0.07	0.09	0.02
Age	-0.14**	-0.02	0.00
Politics	-0.03	0.01	-0.03
Income	-0.11*	-0.03	-0.06
Education	-0.08	-0.09	-0.08
Race	-0.08	-0.03	-0.06
Individual Differences			
Authoritarian	0.03	0.06	0.01
Need Cog.	0.12**	0.01	0.01
Need Closure	0.16***	0.07	0.07
Dominance	0.05	0.04	0.08
Just World	-0.01	-0.18**	-0.16*
Legal Attitudes			
Courts	0.43***	0.27***	0.33***
Attorneys	-0.13**	-0.07	-0.04
Model Sig. (F-Test)	7.92 ***	2.15**	2.40**
Model R ²	.31	.07	.08
N	235	234	235

*Note: Asterisks denote statistically significant effects in each model: *** signifies $p < .01$, ** signifies $p < .05$, * signifies $p < .10$ (marginal significance). 'Ruling' was coded as "0" for admit and "1" for exclude.*

b. Path Analysis of Results in Study 3

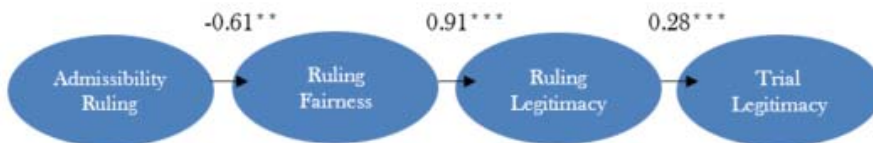
Our final analysis examined the psychological processes that underlie the relationship between the judge's admissibility decision with respect to the propensity evidence at trial and participants' willingness to legitimize the tribunal's verdict. We hypothesized that there would be an indirect relationship between these two variables, which would be mediated by our participants' perceptions of the fairness and legitimacy of the evidential ruling. Specifically,

¹⁸⁰ In other words, over 30% of the change in our participants' ratings of the trial's accuracy could be explained by just the factors that we included in the model.

we hypothesized that our participants would view the court's exclusion of the propensity evidence as unfair to the proffering party, which would (1) cause them to view the process by which the court collected its evidence as less legitimate, and would therefore (2) make them less willing to legitimize the court's ultimate verdict.

We tested this hypothesis by performing a path analysis, which is a more complex version of the analysis that we performed at the conclusion of Study 2.¹⁸¹ This time, instead of predicting a direct relationship between the judge's admissibility decision and the legitimacy of the court's verdict, we predicted a direct, negative association between the judge's decision to exclude the evidence and our participants' perceptions of the fairness of that decision. We also predicted a positive association between our participants' perceptions of the fairness of the decision and the perceived legitimacy of the fact-gathering process. Finally, we expected a direct, positive association between our participants' perceptions of the legitimacy of the fact-finding process, and the legitimacy of the trial. The indirect path analysis that we performed confirmed our hypotheses.¹⁸² An illustration of the pathway, and the coefficients that correspond with each portion of our regression analysis, appears below.

Figure 10. Path Analysis from Admissibility Ruling to Perceptions of Fact Finder Legitimacy



We performed one follow-up analysis as well, this time focusing on our participants' perceptions of the accuracy of the trial court's verdicts as a second indirect pathway between the court's decision to admit or exclude propensity evidence and the perceived legitimacy of the trial. We predicted that our participants would view the court's decision to exclude the propensity evidence as decreasing the likelihood that the court would reach an accurate verdict. We also predicted that their perceptions of the likelihood that the court would reach

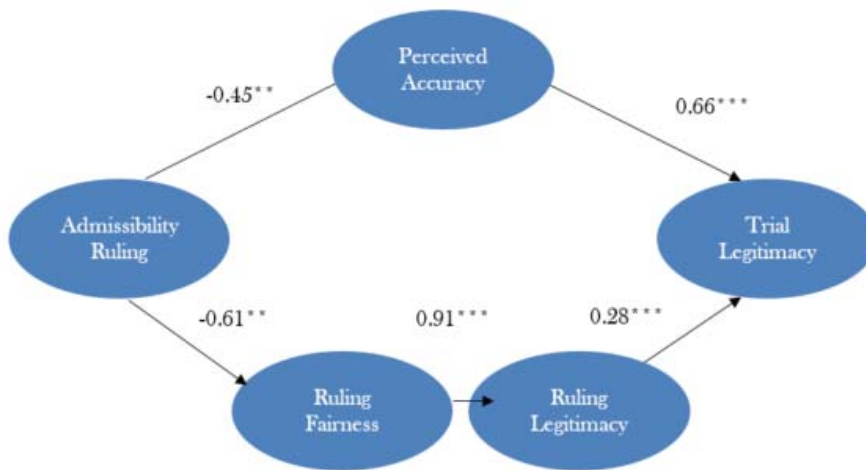
¹⁸¹ As in the serial mediation in Study 2, the path analysis proceeds in a series of regressions, which will show that the judge's admissibility decision affects people's willingness to legitimize the trial, but that it does so indirectly through two pathways: the procedural justice of the evidentiary decision and its effect on the court's ability to reach an accurate judgment.

¹⁸² The beta weights associated with each regression, and the statistical significance of the coefficients, appear in the figure.

an accurate verdict would be positively associated with their willingness to legitimize the trial verdict.

Our indirect path analysis confirmed our hypotheses. We illustrate below a more complete model that includes two indirect pathways between the trial court's evidentiary ruling regarding the character witness's testimony and our participants' willingness to legitimize the trial verdict: (1) an indirect pathway in which the decision affects their perception of the court's ability to reach an accurate verdict; and (2) an indirect pathway in which the fairness of that decision affects their views of the procedural justice of the fact-finding process. A model that combines these pathways appears below, along with the corresponding regression coefficients.

Figure 11. Path Analysis Examining Indirect Routes from the Admissibility of Character Evidence to Perceptions of Fact Finder Legitimacy



3. Discussion of Results in Study 3

Study 3 provides important, counterintuitive insights regarding the acceptability to the public of trials that include propensity evidence. Regardless of the party that offered it, participants perceived the admission of propensity evidence as increasing the trial court's ability to reach an accurate verdict, increasing the perceived fairness of the fact-gathering process that the court used, and increasing the court's ultimate legitimacy. As in Studies 1 and 2, these effects were robust; they remain statistically significant even when we added

thirteen relevant control variables to our models of perceived decisional accuracy, evidentiary fairness, and procedural legitimacy.

Moreover, two additional analyses explained the pathway by which propensity evidence influences the public's willingness to legitimize the courts. Not only is the public more willing to legitimize trial verdicts because they believe that propensity evidence increases the court's decisional accuracy, they also are more willing to legitimize verdicts because they believe that admitting propensity evidence increases the fairness of the process by which the court gathers its facts.

IV. IMPLICATIONS AND OBJECTIONS

The Federal Rules of Evidence deem propensity evidence an illegitimate source of proof in legal fact-finding. The basis for the perceived illegitimacy is two-fold. The Advisory Committee to the Federal Rules of Evidence posits that that the use of propensity evidence in legal fact-finding will unacceptably raise the risk of incorrect verdicts, and the public will perceive such evidence as procedurally unfair.¹⁸³ The findings from the literature on person perception and procedural justice, as well as the results from the three original experiments reported in this Article, suggest that the Advisory Committee's justifications may be incorrect.

The results from our first study do, however, support one aspect of the Advisory Committee's view of propensity evidence: we found that propensity evidence—both proffered by and against the defendant—had a meaningful effect on our mock jurors' verdicts. This was true regardless of the legal setting and across different types of cases. But just because jurors did not *ignore* the evidence does not mean that they weigh character evidence with an eye toward punishing defendants for past indiscretions. In fact, the propensity evidence in our study moved the percentage of participants willing to find the defendant liable only between five and fifteen points in either direction.¹⁸⁴

¹⁸³ See MUELLER & KIRKPATRICK, *supra* note 58, at § 4:22; see also FED. R. EVID. 404(a) advisory committee's notes ("Character evidence is of slight probative value and may be very prejudicial. It tends to distract the trier of fact from the main question of what actually happened on the particular occasion. It subtly permits the trier of fact to reward the good man to punish the bad man because of their respective characters despite what the evidence in the case shows actually happened.").

¹⁸⁴ This was not because the character evidence was insufficiently strong. Follow-up analyses indicated that our mock jurors believed the character evidence was one of the strongest pieces of evidence at the trial (likely because the testimony was given by a friend of the defendant who had known the defendant for several years). Nonetheless, the vast majority of our participants ranked the character evidence as significantly less

Our second study found that, consistent with the interactionist model of person perception, jurors display marked sensitivity to diagnostic features of propensity evidence on three distinct dimensions: the frequency of the predicate act, how long ago it occurred, and the similarity between the predicate act and the act that formed the basis of the current accusation.¹⁸⁵ These findings suggest that jurors make sensible decisions about the strengths and weaknesses of propensity evidence. They further suggest that the Advisory Committee's fears about the risks to the courts' ability to reach accurate verdicts may be misguided.

Our final study found that including character evidence at trial *increases* the public's willingness to legitimize verdicts, and it does so in two ways. First, the public associates such trials with more accurate verdicts. Second, notions of procedural justice—the fairness of the process by which the trial court collects its evidence—*increase* when propensity evidence is admitted. The public apparently believes that jurors should receive this evidence and weigh it as they see fit.¹⁸⁶ Several implications flow from these research findings for the courts, the Federal Rules of Evidence, and attorneys who make ground-level decisions under the current rules.

Our experimental findings suggest that the Advisory Committee's rationale for barring propensity evidence sits atop a shaky house of cards, and each empirical gust of wind shakes the rule's foundation further.¹⁸⁷ Ours is not the first or loudest call for lifting the prohibition on propensity evidence,¹⁸⁸ but it is the first to be supported by empirical data that speaks to both aspects of the propensity rule's legitimacy.

Eliminating the propensity bar will create doctrinal coherence that has eluded the current rule.¹⁸⁹ For example, it eliminates the controversy surrounding Federal Rules of Evidence 413, 414, and 415, which allow the

important to their verdicts than the police officer's testimony and the forensic evidence. Indeed, our mock jurors ranked only the defendant's self-serving alibi testimony as *less* important than the propensity evidence.

¹⁸⁵ See *supra* Section II.B.3.a. (discussing the results).

¹⁸⁶ See *supra* Section III.C.3.a. (discussing the results).

¹⁸⁷ Other evidentiary rules that have been questioned empirically include the hearsay rule under FRE 801, the limiting instruction under FRE 105, and the use of prior convictions for purposes of witness impeachment. See, e.g., Theodore Eisenberg & Valerie P. Hans, *Taking a Stand on Taking the Stand: The Effect of a Prior Criminal Record on the Decision to Testify and on Trial Outcomes*, 94 CORNELL L. REV. 1353, 1354–55 (2009); Justin Sevier, *Testing Tribe's Triangle: Juries, Hearsay, and Psychological Distance*, 103 GEO. L. J. 879, 886 (2015); Nancy Steblay et al., *The Impact on Juror Verdicts of Judicial Instruction to Disregard Inadmissible Evidence: A Meta-Analysis*, 30 LAW & HUM. BEHAV. 469, 469–70 (2006).

¹⁸⁸ Uviller, *supra* note 17.

¹⁸⁹ See *Michelson v. United States*, 335 U.S. 469, 486 (1948) (lamenting the lack of coherence in the doctrine).

prosecutor or plaintiff to proffer evidence of a defendant's propensity for sexual misconduct. Under the current regime, proponents of the rule have had difficulty justifying why the concerns surrounding propensity evidence generally (with respect to its impact on decisional accuracy and procedural legitimacy) do not apply to a defendant's propensity for sexual misconduct.¹⁹⁰ Under a regime in which propensity evidence is admissible by default, all propensity evidence would start with a presumption of admissibility that must be overcome, like all other evidence, by a showing that the propensity evidence at issue is substantially more prejudicial than it is probative.

Lifting the propensity ban also will improve judicial economy. If propensity evidence is admissible, fewer pretrial hearings would be necessary to determine whether a party's evidence is admissible as circumstantial evidence of another relevant fact, pursuant to FRE 404(b), or if the evidence is inadmissible propensity evidence under FRE 404(a).¹⁹¹ Recall that evidence that is admissible for the purpose of showing a party's intent, identity, scheme or plan, or opportunity to commit an act often appears, at first glance, to be inadmissible propensity evidence.¹⁹² Moreover, even when a party proffers evidence pursuant to FRE 404(b), if it might also be used as inadmissible propensity evidence, the court must evaluate the prejudicial effect and the probative value of the evidence under FRE 403, necessitating a hearing.¹⁹³ The prevalence of these time-consuming, expensive proceedings could be decreased if propensity evidence is admissible by default. Such a regime might, in some circumstances, obviate the need to determine whether the character evidence has a propensity or non-propensity purpose.

Moreover, lifting the propensity ban would eliminate the current disincentive for defendants to testify under the Federal Rules of Evidence. Defendants frequently must decide whether to testify—and risk evidence of their character for dishonesty used against them pursuant to FRE 609—or to remain

¹⁹⁰ See, e.g., Ellis, *supra* note 71, at 961–62, 972 (discussing the problem in detail).

¹⁹¹ This is currently conjecture, insofar as empirical data regarding motions in limine are not readily available in most jurisdictions. Collecting such data may be a worthwhile project for other empirical researchers.

¹⁹² See FED. R. EVID. 404(b) advisory committee's notes (“[E]vidence of other crimes, wrongs, or acts is not admissible to prove character as a basis for suggesting the inference that conduct on a particular occasion was in conformity with it. However, the evidence may be offered for another purpose, such as proof of motive, opportunity, and so on, which does not fall within the prohibition.”).

¹⁹³ See *id.* (“In this situation the rule does not require that the evidence be excluded. No mechanical solution is offered. The determination must be made whether the danger of undue prejudice outweighs the probative value of the evidence in view of the availability of other means of proof and other factors appropriate for making decisions of this kind under Rule 403.”).

silent to avoid such character attacks by the prosecution.¹⁹⁴ If the propensity ban is lifted, such evidence is presumptively admissible, and so the defendant's decision to testify would not be distorted by the prospect of opening the door to character evidence. This framework has the benefit of providing the jury with additional information, from both the prosecution and the defendant, on which to render its verdict.

If the bar on propensity evidence is lifted, the interactionist model provides guidance for the process by which such evidence should be admitted. Under the current regime, unless the evidence is used for a non-propensity purpose, only testimony in the form of reputation or opinion is admissible on direct examination when propensity evidence is admissible. The specific acts that form the basis of the pertinent personality trait are admissible only on cross-examination. The interactionist model, however, suggests that this is not the optimal way to present admissible propensity evidence. Jurors pay careful attention to factors attendant to the acts that underlie propensity evidence: their frequency, their age, and their similarity to the conduct at issue. The goal of decisional accuracy would be better served if specific act testimony—as well as the attendant circumstances surrounding the acts that form the basis of the propensity testimony—are admissible on direct examination in addition to cross-examination. This minor procedural reform should not prove controversial; the claim that admitting specific acts on direct examination would be overly burdensome¹⁹⁵ does not currently have any empirical support, and whatever minor delays such testimony might create might be outweighed by the legitimacy gains that propensity evidence offers.

Some critics may raise practical concerns about the admission of propensity evidence. For example, in trials with gruesome subject matter, courts might encounter forms of propensity evidence that are particularly inflammatory or otherwise problematic.¹⁹⁶ Some policymakers might object to fashioning a rule in which such evidence is presumptively admissible. It is important to note, however, that lifting the bar on propensity evidence does not eliminate a judge's discretion to preclude otherwise admissible evidence that is substantially more

¹⁹⁴ See Jeffrey Bellin, *The Silence Penalty*, 103 IOWA L. REV. 395, 407–10 (2018).

¹⁹⁵ See FED. R. EVID. 405 advisory committee's notes ("Of the three methods of proving character provided by the rule, evidence of specific instances of conduct is the most convincing. At the same time it possesses the greatest capacity to arouse prejudice, to confuse, to surprise, and to consume time.")

¹⁹⁶ For example, imagine that a defendant stands accused of a series of murders in a small town. Further imagine that the prosecution desires to put forth evidence of the defendant's prior killing of animals to prove (1) his propensity to be a serial killer and (2) that he therefore committed the murders.

prejudicial than probative.¹⁹⁷ The judge would still retain authority, in the interest of justice, to preclude propensity evidence that is particularly enflaming or dilatory.

Still other critics may raise concerns regarding the theoretical basis for lifting the propensity bar. Although the data suggests that lifting the propensity bar will raise—not lower—the courts’ legitimacy, critics might resist “democratizing” evidentiary rules in this manner, particularly if the public does not appreciate (to the extent that legal experts do) the nuances and implications of evidential rules.¹⁹⁸ The point is fair and important. Nonetheless, there are several areas in the law—evidence, criminal law, business law, and torts, for example—where rules have been modified to align with public conceptions of justice under the law.¹⁹⁹ It is possible that popular legitimacy has such a “darker side,”²⁰⁰ but until those arguments are sufficiently articulated and supported with empirical data, it may be preferable for the rule to align with popular opinion as a default. Policymakers can adjust the rule (by barring certain types of propensity evidence) when other important policy concerns override the default presumption.

Finally, critics may raise methodological concerns. The first involves what this author has deemed elsewhere “the measurement problem.”²⁰¹ In earlier work examining the rationale for the rule barring hearsay,²⁰² this author noted the difficulty of demonstrating that jurors discount hearsay evidence “appropriately” because such a claim

presupposes that the meaning—and probativeness—of a piece of evidence has a fixed value that can be measured reliably. Unfortunately, assessing the probative value of evidence is a topic that has vexed legal scholars for decades; there is currently no prevailing theory of how to appropriately measure various pieces of evidence, nor

¹⁹⁷ See, e.g., FED. R. EVID. 403. Moreover, the judge’s ruling would be subject to the lenient abuse of discretion standard. See *Michelson v. United States*, 335 U.S. 469, 480 (1948) (discussing abuse of discretion standard).

¹⁹⁸ For a review of the benefits and drawbacks of “democratizing” the criminal law, see, for example, Paul Robinson, *Democratizing Criminal Law: Feasibility, Utility, and the Challenge of Social Change*, 111 NW. U. L. REV. 1565, 1566–67 (2017).

¹⁹⁹ *Id.* at 1593–94; see also Sevier, *supra* note 108, at 664.

²⁰⁰ See Robert J. MacCoun, Voice, Control, and Belonging: The Double-Edged Sword of Procedural Fairness, 1 ANN. REV. L. & SOC. SCI. 171, 190 (2005).

²⁰¹ See Sevier, *supra* note 108, at 653–55.

²⁰² Hearsay is an out-of-court statement that a party attempts to enter into evidence for the purpose of demonstrating that the substance of the statement is true. See FED. R. EVID. 801(c). Such statements are excluded from evidence, subject to a wealth of exceptions. See FED. R. EVID. 802–07.

is there an agreed-upon manner to assess how closely legal decision makers adhere to that measurement.²⁰³

In this respect, propensity evidence is no different from hearsay. The studies reported here do not suggest that jurors give propensity evidence “appropriate weight” and that considering such evidence must therefore increase a tribunal’s decisional accuracy. As with hearsay, we have no way to objectively measure the weight that fact finders should give to evidence such as a defendant’s prior acts indicative of character. Instead, these studies support the view that—as with hearsay—jurors make defensible decisions regarding when to credit or discount propensity evidence, and the public finds tribunals that allow the jury to consider propensity evidence more legitimate.

The second methodological concern involves the use of empirical evidence in policy debates more generally. The judiciary has historically had a complex relationship with social science in shaping legal policy. This author has written elsewhere regarding the limitations and benefits of using experimental data to shape public policy.²⁰⁴ Empirical studies have shaped legal policy in a variety of areas, including eyewitness identification, false confessions, the size and shape of juries, the manner of proving discrimination, the regulation of corporate behavior, and the implementation of the death penalty.²⁰⁵ It is, of course, important not to overstate the implications of any one empirical study. But it is also important to situate empirical studies within the literature on which they are based to draw appropriate and measured conclusions about their findings.

CONCLUSION

One thought likely struck Marcus Caelius Rufus as he left the *Quaestio de vi* in the wake of his acquittal for the murder of the Alexandrian ambassador. Cicero’s *Pro Caelio* speech—and the propensity evidence that pervaded it—likely saved Caelius from certain execution for a crime he did not commit. This historical episode cuts against the prevailing narrative surrounding the use of propensity evidence at trial, in which incompetent jurors bungle the probative

²⁰³ Sevier, *supra* note 108, at 653–54 (citing, among other scholarly work, Edward J. Imwinkelried, *The Meaning of Probative Value and Prejudice in Federal Rule of Evidence 403: Can Rule 403 Be Used to Resurrect the Common Law of Evidence?*, 41 VAND. L. REV. 879 (1998)) (discussing probative value generally and in the context of FRE 403).

²⁰⁴ See Justin Sevier, *Vicarious Windfalls*, 102 IOWA L. REV. 651, 705–07 (2017).

²⁰⁵ See generally 3 ADVANCES IN PSYCHOLOGY AND LAW (Monica K. Miller & Brian H. Bornstein eds., 2018).

weight of the evidence in trials that the public would perceive as unfair and illegitimate.

The available empirical evidence suggests that we should allow modern jurors to do now what Caelius's triers did then: evaluate the relevant character evidence against the parties and decide how much (or how little) to credit it. The experimental data converge on the conclusion that jurors make reasonable judgments about the probative weight to attach to propensity evidence, and the public views the introduction of propensity evidence as consistent with notions of fair process. Thus, legitimizing the use of character evidence at trial will have beneficial effects not only for the perceived accuracy and fairness of American trial courts, but also for the public citizens who rely upon—and legitimize—the legal system.