

Southern Methodist University

SMU Scholar

Faculty Journal Articles and Book Chapters

Faculty Scholarship

2023

Character Evidence as a Conduit for Implicit Bias

Hillel J. Bavli

Southern Methodist University, Dedman School of Law

Author ORCID Identifier:

 <https://orcid.org/0000-0003-4888-2842>

Recommended Citation

Hillel J. Bavli, Character Evidence as a Conduit for Implicit Bias, 56 U.C. Davis L. Rev. 1019

This document is brought to you for free and open access by the Faculty Scholarship at SMU Scholar. It has been accepted for inclusion in Faculty Journal Articles and Book Chapters by an authorized administrator of SMU Scholar. For more information, please visit <http://digitalrepository.smu.edu>.

Character Evidence as a Conduit for Implicit Bias

Hillel J. Bavli*

The Federal Rules of Evidence purport to prohibit character evidence, or evidence regarding a defendant's past bad acts or propensities offered to suggest that the defendant acted in accordance with a certain character trait on the occasion in question. However, courts regularly admit character evidence through an expanding set of legislative and judicial exceptions that have all but swallowed the rule. In the usual narrative, character evidence is problematic because jurors place excessive weight on it or punish the defendant for past behavior. Lawmakers rely on this narrative when they create exceptions. However, this account arguably misses a highly troublesome feature of character evidence and far understates its pernicious effects.

In this Article, I develop a new model of character evidence that refocuses the debate on the distortions associated with the prior beliefs and prejudices inherent in a juror's perception of character evidence. Specifically, I draw on disciplines outside of law — including Bayesian statistics and cognitive psychology — to explain how jurors use character evidence to arrive at a verdict. I then apply this framework to show that when a court admits

* Copyright © 2023 Hillel J. Bavli. Assistant Professor of Law, SMU Dedman School of Law. For helpful comments and exchanges, I thank Edward Cheng, Eric Ruben, Bennett Capers, Richard Clary, Michael Saks, John Kenneth Felter, Anna Roberts, Liesa Richter, Edward Imwinkelried, Natalie Nanasi, Teneille Brown, Julia Simon-Kerr, Jonathan Koehler, Paul Rothstein, Ronald Allen, Barry Goldstein, George Fisher, Daniel Capra, and the participants of the 2022 *Harvard/Yale/Stanford Junior Faculty Forum*, the 2022 *Evidence Summer Workshop* at Vanderbilt Law School, the 2022 *Center for Law & Social Science Workshop* at USC Gould School of Law, and the 2022 *Annual Conference of the Berkeley Center on Comparative Equality & Anti-Discrimination Law*. I also thank Hyewon Choi and Benjamin Rice for their outstanding research assistance. I am grateful for generous financial support provided by SMU Dedman School of Law and the WWB Law Professors Fund.

character evidence through exceptions, judgments based on character evidence are inherently biased against certain groups of people based on their race, sex, appearance, accent, education, economic status, and other personal characteristics. I argue that exceptions to the rule against character evidence therefore drive inequality in the U.S. legal system, and that this provides a strong reason to limit such exceptions and to reverse the current trend toward a more permissive rule.

TABLE OF CONTENTS

INTRODUCTION	1021
I. CHARACTER EVIDENCE: THE RULE AND ITS EROSION	1029
A. <i>The Rule Against Character Evidence and its Underlying Policies</i>	1029
B. <i>The Rule's Erosion</i>	1032
II. THE REFLECTIVE MODEL OF CHARACTER EVIDENCE	1039
A. <i>The Probative Value and Influence of Character Evidence</i>	1041
1. Improving Accuracy with Context: Trading for Precision	1041
2. Character Evidence: A Bias-Variance Tradeoff.....	1045
B. <i>The Shrinkage Principle</i>	1053
C. <i>The Influence of Prior Beliefs</i>	1055
D. <i>The Psychology of Juror Decision-Making</i>	1065
III. DISCUSSION AND IMPLICATIONS.....	1070
A. <i>The Interaction Between Character Evidence and Implicit Bias</i>	1070
B. <i>Formal Exceptions to the Rule against Character Evidence</i> ...	1073
C. <i>The Trend Toward Broad Admissibility for Character Evidence</i>	1078
D. <i>Developing Sound Empirical Evidence</i>	1082
CONCLUSION	1084

*Origin of the word character: kharaktēr (Greek), a tool for stamping, from kharassein (Greek), meaning to engrave or stamp.*¹

INTRODUCTION

It is fitting that the word character has its origin in the Greek word *kharassein*, meaning to engrave or stamp:² the image of one's character is as much a creation of the observer as a creation of the observed. In spite of this, courts routinely invite jurors to make judgments based on their perceptions regarding an individual's character. Courts thereby sanction judgments that are based on the subjective prior beliefs and prejudices of jurors rather than on the evidence in a case. Consequently, courts invite jurors to rely on their priors — including their implicit biases — regarding an individual's race, sex, appearance, accent, education, economic status, and other background characteristics when determining a verdict.

The admissibility of evidence regarding a criminal defendant's³ prior bad acts has been described as “the single most important issue in contemporary criminal evidence law.”⁴ Indeed, in a large proportion of cases — including cases like the recent Chauvin trial, Rittenhouse trial, and Cosby trial — the court must decide whether to admit evidence regarding the prior acts of the defendant or a victim,⁵ and “[t]he stakes

¹ *Character*, DICTIONARY.COM, <https://www.dictionary.com/browse/character> (last visited Aug. 12, 2021) [<https://perma.cc/WUD7-JUEN>]; see also *The Characteristics of 'Character'*, MERRIAM-WEBSTER, <https://www.merriam-webster.com/words-at-play/word-history-of-character-origins> (last visited Aug. 12, 2021) [<https://perma.cc/VBL9-ZF2R>].

² *Character*, *supra* note 1.

³ For clarity, throughout this Article, I frequently refer to a criminal defendant as the individual about whom character evidence is offered. However, the discussion herein applies to civil and criminal cases, and to any individual about whom character evidence is offered.

⁴ Edward J. Imwinkelried, *The Use of Evidence of an Accused's Uncharged Misconduct to Prove Mens Rea: The Doctrines Which Threaten to Engulf the Character Evidence Prohibition*, 51 OHIO ST. L.J. 575, 576 (1990).

⁵ *State v. Chauvin*, No. 27-CR-20-12646, 2021 WL 252713, at *4 (D. Minn. Jan. 26, 2021); *Commonwealth v. Cosby*, 252 A.3d 1092, 1121-26 (Pa. 2021); NBC Chicago, *Full Video: Prosecutors Cross-Examine Kyle Rittenhouse*, YOUTUBE, at 32:00 (Nov. 11, 2021), <https://www.youtube.com/watch?v=JG8PhtFrOoY&t=1964s> [<https://perma.cc/9LXA-5LHS>]; see *United States v. Davis*, 726 F.3d 434, 441 (3d Cir. 2013) (“Rule 404(b) has become the most cited evidentiary rule on appeal.”); Jeffrey Bellin, *The Evidence Rules that Convict the*

for the prosecution and the defense are enormous.”⁶ This decision is often critical to — if not altogether determinative of — a case’s outcome.⁷

Rule 404 of the Federal Rules of Evidence (“FRE”)⁸ prohibits evidence of an individual’s character traits or past acts to prove that the individual acted in accordance with a certain character or propensity on the occasion in question.⁹ It reflects the principle that individuals should be judged legally based on what they did rather than who they are.¹⁰ It is not grounded in an absence of probative value.¹¹ Rather, the ban on character evidence is based on the unfair prejudice that it causes.¹² Under Rule 403, courts “exclude relevant evidence if its probative value is substantially outweighed by a danger of . . . unfair prejudice” and other risks, such as

Innocent, 106 CORNELL L. REV. 305, 343 (2021) (“Some of the most heated controversies in the modern evidence landscape involve evidence of uncharged crimes admitted under Federal Rule of Evidence 404(b) or innovative variants like Rules 413 and 414.”); Imwinkelried, *supra* note 4, at 577 (“Rule 404(b) has generated more published opinions than any other subsection of the Federal Rules.”).

⁶ Paul S. Milich, *The Degrading Character Rule in American Criminal Trials*, 47 GA. L. REV. 775, 780 (2013) (“Once the jury learns that the defendant has a criminal past, the odds of conviction skyrocket.”).

⁷ See Daniel J. Capra & Liesa L. Richter, *Character Assassination: Amending Federal Rule of Evidence 404(b) to Protect Criminal Defendants*, 118 COLUM. L. REV. 769, 772 (2018) (“Proof of a criminal defendant’s past crimes has a dramatic effect on a jury, almost guaranteeing conviction.”); see also *United States v. Burkhardt*, 458 F.2d 201, 204-05 (10th Cir. 1972) (“[A]n obvious truth is that once prior convictions are introduced the trial is, for all practical purposes, completed and the guilty outcome follows as a mere formality. This is true regardless of the care and caution employed by the court in instructing the jury.”).

⁸ Throughout this Article, I refer to Rule 404 of the Federal Rules of Evidence. However, the discussion herein applies similarly to state evidentiary rules regarding character evidence.

⁹ FED. R. EVID. 404.

¹⁰ See *United States v. Gomez*, 763 F.3d 845, 861 (7th Cir. 2014) (“[W]e try cases, rather than persons.” (quoting *People v. Allen*, 420 N.W.2d 499, 504 (Mich. 1988))).

¹¹ See *People v. Zackowitz*, 172 N.E. 466, 468 (N.Y. 1930); GEORGE FISHER, EVIDENCE 153-54 (3d ed. 2013); see also *infra* Section I.A.

¹² See *Burkhardt*, 458 F.2d at 204-05 (discussing the prejudicial effects of character evidence and concluding that “it is clear that the problem is not a simple evidentiary one, but rather goes to the fundamental fairness and justice of the trial itself”); see also *supra* note 7 and accompanying text; discussion *infra* Section I.A.

causing confusion and wasting time.¹³ But Rule 404 replaces Rule 403's ordinary balancing analysis for character evidence: it "reflects the judgment of Congress that as a matter of law the probative value of propensity evidence is substantially outweighed by the risk it poses of unfair prejudice, juror confusion, and waste of time."¹⁴

However, notwithstanding the rule against character evidence, courts routinely admit other-acts character evidence through an expanding set of exceptions and ad hoc departures.¹⁵ Some exceptions are legislative. For example, under Rule 413 of the FRE, in a criminal sexual-assault case, "the court may admit evidence that the defendant committed any other sexual assault" to prove that the defendant has a propensity to commit such acts and is therefore likely to have committed the crime in question.¹⁶ Other departures from the rule against character evidence have developed by common law, often through misinterpretations of Rule 404(b)(2).

Rule 404(b)(2) clarifies that other-acts evidence (as distinct from other-acts character evidence) may be admissible if it is offered for a non-character purpose — that is, for a purpose that does not require propensity reasoning, or inferences based on an individual's character or propensity — such as to prove motive, intent, or knowledge.¹⁷ This is important because other-acts evidence may be highly probative for non-character purposes without giving rise to the concerns associated with

¹³ FED. R. EVID. 403. Note that the term "prejudice" in Rule 403 is broader than the term as I use it to describe a juror's "prior beliefs and prejudices." The latter phrase, discussed *infra* Section II.C, refers to a juror's preconceptions, including the juror's preexisting beliefs, stereotypes, and generalizations, regarding a defendant's background characteristics. In Rule 403, this term — and the term "unfair prejudice" in particular — refers more broadly to "an undue tendency to suggest decision on an improper basis, commonly, though not necessarily, an emotional one." FED. R. EVID. 403 advisory committee's note to 1972 proposed rule.

¹⁴ FISHER, *supra* note 11, at 154.

¹⁵ See Milich, *supra* note 6, at 776-78 ("The American rule barring character evidence in criminal cases is degrading in every sense of the word. The rule's vitality has degraded as courts and legislatures expand existing exceptions and add new ones.").

¹⁶ FED. R. EVID. 413.

¹⁷ FED. R. EVID. 404(b)(2).

other-acts character evidence.¹⁸ However, courts also regularly misapply Rule 404(b)(2) by permitting other-acts evidence that, although relevant to a Rule 404(b)(2) purpose such as intent or motive, indeed requires propensity reasoning. For example, courts regularly admit evidence that a defendant previously engaged in drug trafficking to prove that the defendant has a propensity to commit such acts and is therefore likely to have had the requisite knowledge and intent to distribute drugs in the incident in question.¹⁹ Indeed, some of these misapplications are so deeply ingrained in the common law that courts do not even acknowledge the rule when departing from it.²⁰

In summary, the rule against character evidence is not a firm rule. It is riddled with exceptions, and the exclusion of even basic forms of character evidence is at best unpredictable.²¹ Moreover, these exceptions are only expanding, and courts and legislatures have grown increasingly permissive of character evidence.²²

However, in creating exceptions to the rule against character evidence, lawmakers frequently rely on a misleading account of the harms associated with this evidence. Specifically, in the usual narrative, repeated regularly by courts and scholars, the rule against character evidence is based on two forms of unfair prejudice: (1) a jury is likely to give excessive weight to other-acts character evidence, and (2) it may punish a defendant based on the defendant's prior bad acts rather than determine a verdict based on the act in question in the case.²³ Courts and

¹⁸ For example, in a murder trial, a defendant's possession of stolen personal items belonging to the murder victim may be offered to prove the defendant's motive and identity as the perpetrator — both non-propensity uses of the evidence.

¹⁹ See, e.g., *United States v. Smith*, 789 F.3d 923, 929-30 (8th Cir. 2015); *United States v. Williams*, 534 F.3d 980, 984 (8th Cir. 2008); *United States v. Manning*, 79 F.3d 212, 217 (1st Cir. 1996); see also *United States v. Ferrell*, 816 F.3d 433, 447 (7th Cir. 2015) (“We have . . . been mindful that loose policing of Rule 404(b)'s exceptions historically appears in drug cases.”); discussion *infra* notes 71-74 and accompanying text.

²⁰ See *infra* Section I.B.

²¹ See Milich, *supra* note 6, at 776-78 (“In sum, the character rule has become porous with exceptions and unpredictable in application.”); see also Dora W. Klein, *The (Mis)application of Rule 404(b) Heuristics*, 72 U. MIA. L. REV. 706, 713-51 (2018); discussion *infra* Section I.B.

²² See *infra* Section I.B.

²³ See *Old Chief v. United States*, 519 U.S. 172, 180-81 (1997); *Michelson v. United States*, 335 U.S. 469, 475-76 (1948); *People v. Zackowitz*, 172 N.E. 466, 468 (N.Y. 1930)

legislatures rely on this narrative when they create exceptions for circumstances in which character evidence seems highly probative relative to the unfair prejudice described by this narrative, or in which it seems possible to mitigate such unfair prejudice through counteracting measures such as providing juries with limiting instructions.²⁴

Unfortunately, this narrative arguably misses a highly troublesome feature of character evidence and far understates its pernicious effects. Moreover, these effects are intrinsic to character-propensity reasoning and are unlikely to be mitigated through counteracting measures. Relatedly, this narrative — firmly entrenched in caselaw and scholarship — has created a substantial blind spot for courts and scholars: while certain harmful effects of character evidence and Rule 404 erosion have been examined extensively,²⁵ the effect on inequality in the U.S. legal system has received relatively little attention.²⁶

Therefore, in this Article, I develop a new model of character evidence that refocuses the question of character evidence on the distortions associated with the prior beliefs and prejudices inherent in a juror's perception of character evidence. I draw on disciplines outside of law — including Bayesian statistics and cognitive psychology — to explain how jurors use character evidence to arrive at a verdict. I apply this framework to show that when a court admits character evidence through exceptions,

(citing 1 JOHN HENRY WIGMORE, WIGMORE ON EVIDENCE § 194 (1904)); FISHER, *supra* note 11, at 153-54; *see also infra* Section I.A.

²⁴ *See* Milich, *supra* note 6, at 781-84 (“[R]educing the ‘problem’ with character evidence to juror overvaluation of certain inferences tends to trivialize the character rule and invite the perspective that the ills of character evidence can be cured by instructing the jury and warning them away from the troublesome inferences.”). Under Rule 105 of the FRE, “If the court admits evidence that is admissible against a party or for a purpose — but not against another party or for another purpose — the court, on timely request, must restrict the evidence to its proper scope and instruct the jury accordingly.” FED. R. EVID. 105.

²⁵ *See, e.g.*, Capra & Richter, *supra* note 7, at 772, 776-802 (discussing counterproductive “practices [that] add up to a permissive culture of admissibility” under Rule 404(b)); Dora W. Klein, “Rule of Inclusion” Confusion, 58 SAN DIEGO L. REV. 379, 379-403 (2021) (discussing counterproductive interpretations and applications of Rule 404(b)); Klein, *supra* note 21, at 709-12 (discussing counterproductive analyses created by courts for applying Rule 404(b)).

²⁶ This marked gap in caselaw and scholarship surrounding character evidence is particularly surprising in light of the vast literature regarding the impact of implicit biases on case outcomes generally.

it invites jurors to rely on their prior beliefs and prejudices when determining a verdict, and that, consequently, judgments based on character evidence are inherently biased against certain groups of people based on their race, sex, appearance, accent, education, economic status, and other background characteristics.

To introduce what I believe to be a more accurate account of character evidence, I employ a set of basic principles from an area of statistics known as Bayesian inference. This type of reasoning sounds complex (and it can be), but its central principles are intuitive and provide a natural framework for understanding character evidence. At its center, Bayesian inference provides a formula — known as Bayes rule — for combining prior beliefs (a “prior”) with new evidence to arrive at a new belief or conclusion (a “posterior”) that accounts for the new evidence.²⁷ I employ this rule to explain how jurors use character evidence to arrive at a verdict.

Specifically, I consolidate principles of Bayesian inference into a single intuitive rule — which I refer to as the “shrinkage principle” — that tells us how much weight to afford a prior belief relative to new evidence when such evidence becomes available.²⁸ In short, the principle states that when combining prior beliefs with new evidence, each should be weighted in proportion to its relative precision (the inverse of uncertainty) with respect to the matter in question. In particular, a prior belief will be most valuable and will be relied on most heavily when it provides precise information regarding a matter for which there is poor or imprecise new evidence.²⁹ Conversely, a prior belief will be less valuable and will be relied on less heavily when either it provides imprecise information regarding the matter that it informs, or there is precise new evidence regarding that matter.³⁰

I apply this Bayesian framework to describe a process whereby a juror incorporates character evidence into a verdict in two stages, each of which can be understood as an application of the shrinkage principle to

²⁷ See *infra* note 127; see also PETER D. HOFF, A FIRST COURSE IN BAYESIAN STATISTICAL METHODS 1 (2009).

²⁸ See *infra* Sections II.A–B.

²⁹ See Hillel J. Bavli, *An Aggregation Theory of Character Evidence*, 51 J. LEGAL STUD. 39, 48–49, 57–58 (2022) [hereinafter *An Aggregation Theory of Character Evidence*].

³⁰ See *id.*

combine prior beliefs with new evidence. First, the juror combines her implicit biases regarding the defendant's background characteristics with court-sanctioned character evidence to formulate an assessment of the defendant's character. Second, the juror combines her character assessment with other (non-character) evidence in the case to arrive at a verdict.³¹

I argue that, based on this model — which I call the “reflective model” of character evidence — character evidence in a sense activates a juror's implicit biases regarding a defendant's race, sex, appearance, and other background characteristics. Specifically, I apply the shrinkage principle to argue that a juror's implicit biases — her priors regarding the defendant — remain relatively subdued in their effect when the juror is asked to make non-character judgments, such as whether a defendant committed the act in question; on the other hand, when a court admits character evidence, and a juror is asked (implicitly or explicitly) to engage in character reasoning, this activates the full force and effect of the juror's priors. Indeed, I argue that character evidence inherently requires a juror to rely on her prior beliefs and prejudices. Meanwhile, from the start of the trial, the juror has been developing an extensive set of prior beliefs and prejudices based on the defendant's background characteristics and regarding precisely what the juror is now asked to assess: the defendant's character.

Importantly, although the model proposed herein is grounded in a theory of information optimization based on Bayesian principles, it finds substantial support in the psychology of juror decision-making. Specifically, I draw on scholarship surrounding the concepts of “stereotype activation,” “confirmation bias,” and other concepts in cognitive psychology to demonstrate that the behavioral assumptions on which the reflective model relies are well grounded in empirical research regarding how people — and jurors in particular — process information.³²

Finally, I argue that the model proposed herein has critical implications for policy surrounding the admissibility of character evidence — policy that is currently premised on a narrative that does not adequately account for the pernicious nature and effects of character

³¹ See *infra* Sections II.A–C.

³² See *infra* Section II.D.

evidence. In particular, I examine implications for exceptions to the rule against character evidence and for the courts' increasingly permissive approach to character evidence. I also highlight the importance of developing sound empirical research regarding the interaction between character evidence and implicit bias, and I describe a basic approach to empirically test the proposed model.

Ultimately, I argue that, based on my analysis herein, a stricter rule against character evidence is necessary. In particular, courts should read Rule 404(b) as a rule of exclusion, interpreting Rule 404(b)(2) as providing for permissible non-propensity uses of other-acts evidence rather than providing for exceptions to the ban on character evidence. Moreover, although there may be extremely important concerns underlying some formal exceptions to the rule against character evidence — such as the exception for other-acts evidence in sexual-assault cases — these exceptions should be evaluated in light of their inherent relationship with jurors' prior beliefs and prejudices regarding a defendant's background characteristics, and the critical aims of fairness and equality should be protected along with those underlying such exceptions.³³

I proceed as follows: In Part I, I discuss the rule against character evidence, its underlying policy concerns according to the traditional account, and its many exceptions. In Part II, I develop the reflective model of character evidence. In particular, I apply basic concepts in Bayesian inference to describe how jurors use character evidence to arrive at a verdict, and to explain the shrinkage principle in simple and intuitive terms. I then apply this principle to demonstrate how character evidence may activate a juror's prior beliefs and prejudices and invite jurors to rely on their implicit biases regarding a defendant in determining a verdict. In Part III, I discuss implications of the reflective model for evidence law — including implications for exceptions to the

³³ Note that jury limiting instructions under Rule 105 aimed at preventing impermissible character reasoning do not alleviate the concern that I address herein, which involves circumstances in which courts *permit* character reasoning rather than seek to limit it via a limiting instruction. FED. R. EVID. 105. Moreover, for reasons discussed in detail below, a limiting instruction regarding a juror's prior beliefs in particular — although perhaps helpful in some circumstances — is unlikely to be effective in mitigating a juror's reliance on prior beliefs when determining a judgment based on character evidence. See *infra* notes 193, 199; discussion *infra* Section III.A.

rule against character evidence and for the trend toward a more permissive approach to character evidence. Finally, I discuss implications for empirically examining the interaction between character evidence and implicit bias with respect to case outcomes. In Part IV, I conclude.

I. CHARACTER EVIDENCE: THE RULE AND ITS EROSION

In this Part, I discuss the rule against character evidence and the policy concerns underlying it according to the traditional account. I then examine the state of erosion and disarray surrounding the rule against character evidence, and the general trend toward a more permissive rule.

A. *The Rule Against Character Evidence and its Underlying Policies*

Rule 404(a) of the FRE states that “[e]vidence 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.”³⁴ Rule 404(b) provides for a particularly important application of this rule: “Evidence of any other crime, wrong, or act is not admissible to prove a person’s character in order to show that on a particular occasion the person acted in accordance with the character.”³⁵

Rule 404 is often understood as a particular application of Rule 403, which excludes evidence “if its probative value is substantially outweighed by a danger of one or more of the following: unfair prejudice, confusing the issues, misleading the jury, undue delay, wasting time, or needlessly presenting cumulative evidence.”³⁶ Rule 404 preempts Rule 403 by substituting a case-by-case balancing with a blanket rule that excludes character evidence.³⁷

It is well-recognized that character evidence — and particularly other-acts character evidence — can be very probative.³⁸ As Justice Cardozo

³⁴ FED. R. EVID. 404(a).

³⁵ FED. R. EVID. 404(b).

³⁶ FED. R. EVID. 403.

³⁷ See FISHER, *supra* note 11, at 154 (“Rule 404 reflects the judgment of Congress that *as a matter of law* the probative value of propensity evidence is substantially outweighed by the risk it poses of unfair prejudice, juror confusion, and waste of time.”).

³⁸ See Steven Goode, *It’s Time to Put Character Back into the Character-Evidence Rule*, 104 MARQ. L. REV. 709, 711-19 (2021).

famously stated in *People v. Zackowitz*, “There may be cogency in the argument that a quarrelsome defendant is more likely to start a quarrel than one of milder type, a man of dangerous mode of life more likely than a shy recluse.”³⁹ Indeed, there is a large literature regarding the high degree of consistency in an individual’s behavior given a type of situation.⁴⁰

Moreover, we undoubtedly rely on such consistency in everyday life. For example, we rely on a contractor’s reputation before hiring her to renovate a house; we ask for recommendations and references before hiring an employee; we make predictions regarding an employee’s (or a friend’s) punctuality based on whether they have been punctual in the past; and we assume that a violent armed robbery is more likely to have been committed by someone who has been convicted of numerous violent crimes in the past than someone who has a spotless criminal record. These inferences are logical, if only approximations based on imperfect information.

On the other hand, character evidence can be highly prejudicial.⁴¹ In the usual narrative, Rule 404 aims to avoid two forms of unfair prejudice: (1) the risk that a factfinder will afford excessive weight to character evidence; and (2) the risk that a factfinder will punish a party based on their character or prior acts rather than on the act at issue in a case.⁴² For example, if a prosecutor is permitted to introduce evidence in a drug-trafficking case that the defendant has multiple prior convictions for

³⁹ *People v. Zackowitz*, 172 N.E. 466, 468 (N.Y. 1930).

⁴⁰ See generally Edward J. Imwinkelried, *Reshaping the “Grotesque” Doctrine of Character Evidence: The Reform Implications of the Most Recent Psychological Research*, 36 Sw. U. L. REV. 741, 747-59 (2008) (summarizing relevant psychology literature).

⁴¹ See *United States v. Burkhart*, 458 F.2d 201, 204-05 (10th Cir. 1972); Milich, *supra* note 6, at 776-78. Although I sometimes use the term “unfair prejudice” in relation to character evidence, I generally use the term “prejudice” or “prejudicial” in this context to include unfairness. See generally FED. R. EVID. 403 (excluding relevant evidence based on danger of unfair prejudice and other factors).

⁴² See *Old Chief v. United States*, 519 U.S. 172, 180-81 (1997); *Michelson v. United States*, 335 U.S. 469, 475-76 (1948); *Zackowitz*, 172 N.E. at 468 (citing 1 JOHN HENRY WIGMORE, WIGMORE ON EVIDENCE § 194 (1904)); FISHER, *supra* note 11, at 153; see also Goode, *supra* note 38, at 721 (“[Character] evidence is likely to have a prejudicial effect. For various reasons, jurors may tend to overvalue such evidence or to convict a defendant even if they are not convinced beyond a reasonable doubt that he is guilty of the current charge.”).

drug-trafficking, there is a high risk that the jury will make heavily weighted character inferences — that the defendant is a drug trafficker (i.e., has a character for committing this type of act) and is therefore likely to have trafficked drugs on the occasion in question — or that it will punish the defendant for his character or prior acts.⁴³ A jury may punish a defendant for prior acts, for example, by lowering its threshold for a conviction, or even more explicitly, by convicting the defendant based on his criminal history and failure to “change his ways” rather than on evidence regarding the act in question.⁴⁴ Additionally, jurors may unfairly incorporate character evidence through confusion — e.g., by judging the defendant based on the juror’s image of him, or by losing sight of the act at issue.⁴⁵

These rationales are arguably grounded in an overarching goal of accuracy — of arriving at a correct determination.⁴⁶ Indeed, accuracy is the central aim of the FRE.⁴⁷ As stated in Rule 102 of the FRE: “These rules should be construed so as to administer every proceeding fairly, eliminate unjustifiable expense and delay, and promote the development of evidence law, to the end of ascertaining the truth and securing a just determination.”⁴⁸

⁴³ See *Michelson*, 335 U.S. at 475-76 (“[Character] is said to weigh too much with the jury and to so overpersuade them as to prejudice one with a bad general record and deny him a fair opportunity to defend against a particular charge.”); discussion *infra* notes 72-74 and accompanying text; see also MICHAEL J. SAKS & BARBARA A. SPELLMAN, *THE PSYCHOLOGICAL FOUNDATIONS OF EVIDENCE LAW* 142-81 (2016) (“Psychological research on the predictive connection between personality and behavior, plus how people perceive character and its relationship to behavior, lends considerable support to the conclusions of the common law rulemakers. Indeed, it is possible that the rulemakers are standing on firmer ground than they realized.”).

⁴⁴ See *Zackowitz*, 172 N.E. at 468 (“The natural and inevitable tendency of the tribunal — whether judge or jury — is to give excessive weight to the vicious record of crime thus exhibited, and either to allow it to bear too strongly on the present charge, or to take the proof of it as justifying a condemnation irrespective of guilt of the present charge.” (internal quotation marks omitted) (quoting 1 JOHN HENRY WIGMORE, *WIGMORE ON EVIDENCE* § 194 (1904))).

⁴⁵ See FISHER, *supra* note 11, at 154.

⁴⁶ See Hillel J. Bavli, *An Objective-Chance Exception to the Rule Against Character Evidence*, 74 ALA. L. REV. 121, 127-30 (2022).

⁴⁷ See FISHER, *supra* note 11, at 1 (“We want juries to return the right verdict, and by that we may mean the *truthful* verdict, the one that accords with *what happened*.”).

⁴⁸ FED. R. EVID. 102.

Moreover, Congress's decision to replace Rule 403's balancing test with a rule forbidding character evidence regardless of the specifics of the probative value and prejudicial effect in a particular case also suggests a policy aim other than accuracy.⁴⁹ Specifically, Rule 404 is also grounded in a policy of affording parties a "blank slate" — that is, giving them the benefit of the doubt by assuming that if they acted a certain way in the past, they may well have changed their pattern of behavior.⁵⁰ In simple terms, Rule 404 reflects the "action-not-person principle"⁵¹ — that, at trial, people should be judged based on what they did and not based on who they are.⁵²

In summary, in the usual narrative, the rule against character evidence is grounded in a fear of jurors misusing character evidence by affording it excessive weight and by punishing a defendant for past bad acts rather than determining a verdict based on the act at issue in a case. The fear of introducing these forms of unfair prejudice is well founded. Indeed, it is highly likely that such misuses of character evidence result in false convictions and other unjust determinations.⁵³ However, as discussed in Parts II–III, this characterization is likely to far underestimate the pernicious nature and effects of character evidence.

B. *The Rule's Erosion*

The rule against character evidence does not live up to its goals. This is because it has been drastically eroded through a large set of legislative

⁴⁹ See FED. R. EVID. 403; *supra* note 36 and accompanying text.

⁵⁰ See *People v. Zackowitz*, 172 N.E. 466, 468 (N.Y. 1930) ("The law . . . is not blind to the peril to the innocent if character is accepted as probative of crime."); Capra & Richter, *supra* note 7, at 771 ("The prohibition on character evidence is a time-honored tenet of evidence law. The American adversary system was designed to convict defendants based upon their conduct and not based on their general character or past misdeeds.").

⁵¹ Bavli, *supra* note 46, at 127–30.

⁵² See *United States v. Gomez*, 763 F.3d 845, 861 (7th Cir. 2014); *United States v. Brown*, 597 F.3d 399, 404 (D.C. Cir. 2010). For a detailed discussion of the action-not-person principle and non-accuracy components of Rule 404, see Bavli, *supra* note 46, at 127–30.

⁵³ See *supra* notes 3–7 and accompanying text.

and judicial exceptions, ad hoc departures, and misinterpretations.⁵⁴ The rule's erosion has resulted in substantial unpredictability, which means that defendants will take unfair plea agreements in order to avoid risking the introduction of other-acts character evidence; defendants will detrimentally change their trial strategy in order to avoid risking a verdict based on character evidence; and, of course, defendants will suffer from verdicts that are based on their prior acts — or, rather, the jury's perception of their prior acts — rather than on the behavior at issue in a case.⁵⁵

First, there are numerous legislative exceptions to the rule against character evidence. For example, Rule 413 of the FRE provides that “[i]n a criminal case in which a defendant is accused of a sexual assault, the court may admit evidence that the defendant committed any other sexual assault. The evidence may be considered on any matter to which it is relevant.”⁵⁶ This rule, enacted in 1994, replaces Rule 404 in sexual-assault cases and allows a court to admit evidence of prior acts of sexual assault broadly for “any matter to which it is relevant.”⁵⁷ This includes evidence of prior sexual-assault claims and acts for which the defendant was never charged, prior acts for which the defendant was charged but not convicted, and prior convictions.⁵⁸ Moreover, Rule 413 is not like Rule 404(b)(2), which permits other-acts evidence for a non-character purpose, such as to prove knowledge or motive.⁵⁹ Instead, it permits evidence of prior acts of sexual assault for any relevant purpose, including for propensity purposes.⁶⁰ That is, such evidence may be admitted explicitly to invite the inference that the defendant has a character or propensity for committing sexual assault and is therefore more likely to have committed the act in question.⁶¹

⁵⁴ See Capra & Richter, *supra* note 7, at 770-76; Klein, *supra* note 25, at 379-403; Klein, *supra* note 21, at 713-51; Milich, *supra* note 6, at 776-78; see also Goode, *supra* note 38, at 711-23.

⁵⁵ See Capra & Richter, *supra* note 7, at 770-76; Milich, *supra* note 6, at 776-78; discussion *supra* notes 41-52 and accompanying text.

⁵⁶ FED. R. EVID. 413; see also FED. R. EVID. 414-15.

⁵⁷ FED. R. EVID. 413; see FISHER, *supra* note 11, at 207-08, 218.

⁵⁸ See FED. R. EVID. 413.

⁵⁹ FED. R. EVID. 404(b)(2).

⁶⁰ FED. R. EVID. 413; see FISHER, *supra* note 11, at 218-26.

⁶¹ See FED. R. EVID. 413; FISHER, *supra* note 11, at 218-26.

Many states have adopted evidentiary rules analogous to Rule 413.⁶² Additionally, some states have extended the reasoning in Rule 413 beyond allegations of sexual assault, to allegations of domestic violence. For example, Section 1109 of the California Evidence Code provides for a broad exception to California's rule against character evidence for "evidence of the defendant's commission of other domestic violence" in domestic-violence cases.⁶³

These exceptions have been criticized as giving rise to the precise problems that Rule 404 seeks to prevent.⁶⁴ At the same time, however, the policy objectives surrounding these exceptions and others are extremely important. They include enabling authorities to more effectively prosecute cases notwithstanding evidentiary problems that arise uniquely in these cases.⁶⁵

Second, the courts have carved out a wide range of exceptions to the rule against character evidence. Many exceptions have developed over time due to the unique probative value of some categories of character evidence.⁶⁶ Additionally, courts regularly and explicitly interpret Rule 404(b)(2) as providing for exceptions to Rule 404(b)(1) rather than a clarification of it — in other words, as permitting other-acts evidence that requires propensity reasoning if it falls within one of the Rule 404(b)(2) categories (such as to prove intent or motive) rather than as merely clarifying that other-acts evidence is permissible so long as it does not involve propensity reasoning.⁶⁷

Many scholars have argued that this reading misinterprets Rule 404(b)(2).⁶⁸ Specifically, Rule 404(b)(2) clarifies Rule 404(b)(1) by

⁶² See FISHER, *supra* note 11, at 218-19.

⁶³ CAL. EVID. CODE § 1109(a) (2022).

⁶⁴ See, e.g., Katharine K. Baker, *Once a Rapist? Motivational Evidence and Relevancy in Rape Law*, 110 HARV. L. REV. 563 (1997) (arguing that Rule 413 is misguided and grounded in incorrect assumptions); see also FISHER, *supra* note 11, at 218-34.

⁶⁵ See FISHER, *supra* note 11, at 218-26.

⁶⁶ See *An Aggregation Theory of Character Evidence*, *supra* note 29, at 39-44; Goode, *supra* note 38, at 711-60.

⁶⁷ See Klein, *supra* note 25, at 389-412 (reviewing cases); Capra & Richter, *supra* note 7, at 778-87 (reviewing cases).

⁶⁸ See FISHER, *supra* note 11, at 157-58 (“[T]he permitted purposes listed in Rule 404(b)(2) are not ‘exceptions’ to Rule 404(a)(1). They are merely possible uses of other-acts evidence *not banned* by Rule 404(a)(1) . . . to build a chain of inferences that goes

indicating that other-acts evidence may be admitted for other, non-propensity purposes. It provides: “This evidence may be admissible for another purpose, such as proving motive, opportunity, intent, preparation, plan, knowledge, identity, absence of mistake, or lack of accident.”⁶⁹ The rule thus provides examples of non-propensity uses of other-acts evidence, and it emphasizes that these uses may be permissible.⁷⁰ For example, if a burglar bypassed a complex alarm system in a previous burglary, evidence of the crime may be admissible in a current burglary case involving a similar alarm system to prove that the defendant knew how to bypass the complex alarm system. The relevance of this evidence does not rely on propensity reasoning but rather only on the reasoning that the prior act demonstrates the defendant’s knowledge of how to bypass a complex alarm system and that the defendant is therefore more likely to have committed the burglary in question.

However, courts regularly misapply Rule 404(b)(2) to admit character evidence. For example, evidence of prior drug-trafficking crimes is regularly admitted to prove intent or knowledge.⁷¹ As the First Circuit explained in *United States v. Manning*, “The evidence that [the defendant] had previously sold cocaine makes it more likely both that he was aware of the contents of the plastic bags in the briefcase and that he intended to distribute the two bags of cocaine.”⁷² However, this reasoning clearly

around the propensity box.” (“It is true that judges often *call* the permitted purposes listed in FRE 404(b)(2) ‘exceptions’ to Rule 404(a)(1) [and Rule 404(b)(1)]. But all such references are in error.”); *see also* Klein, *supra* note 25, at 381, 384-89 (“[C]ourts use the inclusive structure of Rule 404(b) to create unwarranted and erroneous presumptions about the admissibility of other acts evidence.”) (“The current approach of most federal circuit courts of appeals to the application of Rule 404(b) is flawed because these courts have used the inclusive structure of the rule as a basis for creating a presumption that other acts evidence is admissible.”); Capra & Richter, *supra* note 7, at 771-73 (“Notwithstanding its origins as part of a rule with an exclusionary purpose, Rule 404(b) has been characterized by many federal circuit courts as a rule of inclusion.”). *See generally* Paul F. Rothstein, *Intellectual Coherence in an Evidence Code*, 28 LOY. L.A. L. REV. 1259, 1259-65 (1995) (discussing an inconsistency in Rule 404(b)).

⁶⁹ FED. R. EVID. 404(b)(2).

⁷⁰ *See id.*

⁷¹ *See, e.g.,* *United States v. Smith*, 789 F.3d 923, 929-30 (8th Cir. 2015); *United States v. Williams*, 534 F.3d 980, 984 (8th Cir. 2008); *United States v. Manning*, 79 F.3d 212, 217 (1st Cir. 1996).

⁷² *Manning*, 79 F.3d at 217.

relies on propensity inferences. It suggests that the defendant had such knowledge and intent because the defendant has previously engaged in drug distribution and has a propensity to commit such acts. Many courts thus explicitly read Rule 404(b)(2) as allowing propensity reasoning so long as the evidence is offered for one of the enumerated purposes.⁷³ As stated in *United States v. Sterling*, “Rule 404(b)(1) generally prohibits the introduction of propensity evidence at trial. Rule 404(b)(2), however, provides an exception to this general rule for evidence that is also probative for some other purpose, ‘such as proving motive, opportunity, intent, preparation, plan, knowledge, identity, absence of mistake, or lack of accident.’”⁷⁴

Courts are split in their approach to evidence under Rule 404(b)(2), with some courts admitting Rule 404(b)(2) evidence only when it does not rely on propensity reasoning,⁷⁵ while many other courts treat Rule 404(b) as a “[rule] of inclusion,” reading Rule 404(b)(2) as providing for exceptions to the ban on character evidence, and holding that evidence offered for a 404(b)(2) purpose is, e.g., “presumed admissible absent a contrary determination.”⁷⁶ However, as many scholars have emphasized, reading Rule 404(b)(2) as providing for a presumption of admissibility or for exceptions to the rule against character evidence is incorrect,⁷⁷ and “[s]ometimes the error of calling these other purposes ‘exceptions’ to the propensity evidence ban can lead a court to the wrong result.”⁷⁸

Moreover, the trend away from a strict ban on character evidence is only accelerating.⁷⁹ In response to the courts’ increasingly loose treatment of character evidence, the Advisory Committee considered

⁷³ See Klein, *supra* note 25, at 389-412 (reviewing cases); Capra & Richter, *supra* note 7, at 778-87 (reviewing cases).

⁷⁴ *United States v. Sterling*, 738 F.3d 228, 237 (11th Cir. 2013) (quoting FED. R. EVID. 404(b)(2)); see also *United States v. Contreras*, 816 F.3d 502, 510-12 (8th Cir. 2016) (involving evidence of prior spanking incident); *United States v. Smith*, 383 F.3d 700, 706 (8th Cir. 2004) (involving evidence of prior drug transactions); *Young v. Rabideau*, 821 F.2d 373, 379 (7th Cir. 1987) (involving claims of excessive force).

⁷⁵ See Capra & Richter, *supra* note 7, at 787-95 (reviewing cases).

⁷⁶ *United States v. Williams*, 796 F.3d 951, 958 (8th Cir. 2015) (internal quotation marks and citations omitted).

⁷⁷ See *supra* note 68.

⁷⁸ FISHER, *supra* note 11, at 158.

⁷⁹ See Milich, *supra* note 6, at 776.

recommendations to reaffirm the exclusionary nature of Rule 404(b); however, it “determined that it would not propose substantive amendments to Rule 404(b), because they would make the Rule more complex without rendering substantial improvement.”⁸⁰ According to the Committee:

[A]n attempt to require the court to establish the probative value of a bad act by a chain of inferences that did not involve propensity would add substantial complexity, while ignoring that in some cases, a bad act is legitimately offered for a proper purpose but is nonetheless bound up with a propensity inference⁸¹

Thus, the “degrad[ation]” of the rule against character evidence continues.⁸² As one author recently stated:

Rule 404(b)(2)’s laundry list of permissible uses for other-acts evidence supplies courts with an easy way to avoid scrutinizing how the evidence is probative. Rather than asking whether the evidence’s probative value flows only through a character-propensity inference, courts often merely point to one or more of the listed permissible uses and state that Rule 404(b) authorizes its admission. Sometimes courts address the matter in more detail and strenuously maintain that the evidence is probative without any character-propensity inference, even though careful analysis shows that such an inference is required. Typically, they then perform a Rule 403 balancing to determine whether the danger that the jury will use the evidence for an impermissible character-propensity inference substantially outweighs the probative value of the evidence for its permissible purpose. But this Rule 403 balancing is hopelessly skewed because courts consider the (unrecognized) character-

⁸⁰ Advisory Comm. on Evidence Rules, Report of the Advisory Committee on Evidence Rules, at 4-5 (May 14, 2018) [hereinafter Advisory Committee Report]; see *United States v. Thorne*, No. 18-389, 2020 WL 122985, at *5 n.4 (D.D.C. Jan. 10, 2020) (noting the Advisory Committee’s consideration of changes to Rule 404(b)).

⁸¹ Advisory Committee Report, *supra* note 80, at 5; see *Thorne*, 2020 WL 122985, at *5 n.4.

⁸² Milich, *supra* note 6, at 776.

propensity-based inference as proper, rather than improper, and so place it on the probative-value side of the scale and not on the unfair-prejudice side.⁸³

Thus, although the rule against character evidence reflects a long tradition of ensuring a verdict based on the act at issue rather than a party's character or previous acts, the rule is very heavily eroded — so much so, that except in the most extreme cases, a party cannot predict with any certainty whether a court will exclude character evidence.⁸⁴ To the contrary, character evidence is routinely admitted, and verdicts are regularly decided based on prior acts and the jury's perception of the defendant's character.

Moreover, legislative and judicial exceptions to the rule against character evidence are frequently premised on the traditional account of the harms associated with character evidence, which leads legislatures and courts to create exceptions when a type of character evidence seems to involve a high degree of probative value relative to the prejudicial effects described in this account, or when circumstances seem to permit a court to mitigate a juror's tendency to afford excessive weight to, or to punish a defendant for, past bad acts.⁸⁵ However, as discussed below, this account fails to capture a broad category of distortionary effects associated with the prior beliefs and prejudices inherent in a juror's perception of character evidence. These effects are particularly harmful with respect to the goals of accuracy and equal treatment, and it is

⁸³ Goode, *supra* note 38, at 723-24 (footnotes omitted).

⁸⁴ See Capra & Richter, *supra* note 7, at 771 (“The American adversary system was designed to convict defendants based upon their conduct and not based on their general character or past misdeeds.”). The unpredictability with which the rule is applied means that the admissibility of other-acts character evidence — widely recognized as highly impactful on case outcomes — is left to chance. This leads to arbitrary verdicts and unfair plea bargains and settlements. See generally *id.* at 770-802 (discussing counterproductive “practices [that] add up to a permissive culture of admissibility” under Rule 404(b)); Milich, *supra* note 6, at 776-91 (highlighting both the substantial impact and the unpredictability of admissibility decisions surrounding character evidence).

⁸⁵ See generally Milich, *supra* note 6, at 781-84 (“[R]educing the ‘problem’ with character evidence to juror overvaluation of certain inferences tends to trivialize the character rule and invite the perspective that the ills of character evidence can be cured by instructing the jury and warning them away from the troublesome inferences.”).

unlikely that they can be resolved through limiting instructions and other counteracting measures.

II. THE REFLECTIVE MODEL OF CHARACTER EVIDENCE

There has been relatively little in-depth analysis regarding precisely how jurors use character evidence to arrive at a verdict. It is well-recognized that character evidence — and particularly other-acts character evidence — can be extremely impactful on, and even determinative of, a case's outcome. However, how a juror uses character evidence is a black box — it is unknown. Moreover, notwithstanding a large body of scholarship surrounding character evidence, and a separate large body of scholarship regarding the harmful effects of implicit bias on case outcomes, the literature contains no general theory for how character evidence interacts with a defendant's background characteristics or a juror's prior beliefs regarding these background characteristics — and, in particular, for how the rule against character evidence may act as a barrier that prevents a juror's prior beliefs, stereotypes, and prejudices from entering the trial. The current Part aims to develop such a theory.

To begin, assume that a prosecutor in an assault trial offers evidence that, in a separate incident two years earlier, the defendant was accused of committing an assault of a similar nature. The evidence relates to an entirely distinct act that may or may not have occurred. However, the prosecutor offers this evidence to show that the defendant has a propensity to commit such acts and is therefore more likely to have committed the act in question. How should a juror process this information? For example, one juror (or jury) may view the prior alleged act as meriting little weight for deciding the current case. After all, it was an allegation regarding a distinct incident from two years earlier. At the same time, another juror (or jury) may weight the evidence heavily — interpreting the evidence as highly indicative of whether the defendant has committed the crime in question — and may even decide the case entirely based on the evidence.

Neither courts nor scholars know how jurors will use a particular piece of character evidence, how it will impact a case's outcome, or what factors determine the weight attached to the evidence. While courts may prohibit certain uses of evidence, they generally do not instruct jurors on

how to use character evidence or how much weight to give it. This is certainly not to say that other-acts character evidence entails no probative value. To the contrary, this evidence can be quite probative, and certainly very influential on the outcome of a case. But the probative value and the influence of the evidence are left entirely to the discretion of the jurors.

I begin the current Part by modeling the probative value of character evidence and by providing a conceptual background for the shrinkage principle, which I will use to explain the role of character evidence and of prior beliefs in a juror's judgment. I discuss how character evidence impacts accuracy through a concept in statistics known as a bias-variance tradeoff, and I explain the factors that determine its impact on accuracy. I then apply these concepts to introduce a framework for understanding how jurors use character evidence and how prior beliefs impact a judgment via character evidence — and ultimately for understanding why judgments based on character evidence are inherently biased against certain groups of people based on their race, sex, appearance, and other background characteristics.

The reflective model is grounded in basic concepts borrowed from Bayesian statistics. To make the model as intuitive and accessible as possible notwithstanding these technical concepts, I do three things: First, in Section A, I explain the model's conceptual foundation in simple (mostly qualitative) terms with many illustrations. Second, in Section B, I distill a fundamental principle from the analysis in Section A — the shrinkage principle — that will serve as the centerpiece of the reflective model and the primary principle that I apply to analyze the interaction between character evidence and implicit bias in Section C. I explain this principle in standalone, intuitive terms. Third, after establishing basic concepts in Section A, I generally restrict technical notation and terminology to parentheses so that readers can easily skip these descriptions if they choose to do so. Although the Bayesian concepts underlying the model are important for a complete understanding of the model's conceptual foundation, they are not necessary for an intuitive understanding of the model and its implications.

A. *The Probative Value and Influence of Character Evidence*

1. Improving Accuracy with Context: Trading for Precision

Assume that there is a correct judgment — denoted by α — that would result from perfect information regarding the facts surrounding a case. This judgment could be a verdict or an intermediate judgment, such as whether an element is fulfilled or whether a witness is testifying truthfully. Because jurors do not have perfect information, however, they must arrive at an estimate of the correct judgment — that is, they must arrive at an actual judgment, denoted by $\hat{\alpha}$.⁸⁶ We can then define error in terms of the difference between the actual judgment and the correct judgment, and we can define accuracy in terms of the proximity between the two.⁸⁷

Error, as frequently defined in statistics and the sciences, can be deconstructed into two components: variance and bias. Variance measures the randomness around a mean or “expected” value, whereas bias measures the difference between the expected value and the correct value.⁸⁸

More specifically, if an actual judgment ($\hat{\alpha}$) is different from a correct judgment (α), the difference can be due to bias, variance, or both.⁸⁹ Variance measures “judgment variability,” which is the randomness, unpredictability, or dispersion associated with a judgment.⁹⁰ If a case is adjudicated multiple times, each time involving a different jury, judge, presentation of the evidence, etc., judgment variability represents the

⁸⁶ *An Aggregation Theory of Character Evidence*, *supra* note 29, at 46. Throughout this Article, I use the term “judgment” in isolation to refer to an *actual* judgment, unless I specifically indicate otherwise.

⁸⁷ *Id.*

⁸⁸ *Id.* at 46-47. For example, if I use a scale to weigh an object that weighs 100 pounds, and the scale generates a result of 105 pounds, the error (105-100=5 pounds) may result from bias (the scale is not calibrated correctly and, on average, generates overestimates of the correct weight), variance (the scale is imprecise and generates estimates that are highly variable, even if correct on average), or both.

⁸⁹ *Id.*

⁹⁰ Hillel J. Bavli, *The Logic of Comparable-Case Guidance in the Determination of Awards for Pain and Suffering and Punitive Damages*, 85 U. CIN. L. REV. 1, 17-18 (2017) [hereinafter *The Logic of Comparable-Case Guidance*].

variation in judgments from adjudication to adjudication.⁹¹ In a civil case, one jury may award a plaintiff \$100,000 for his pain and suffering while another jury awards \$0; in a criminal case, one jury may decide a verdict of guilty while another decides a verdict of not guilty.⁹² “Precision” is the inverse of variability: the higher the judgment variability, the less precise the judgment is, and vice versa.

Bias, on the other hand, measures the difference between a judgment’s expectation — the mean value that would result from repeated adjudications — and the correct judgment (α). A judgment is “unbiased” if its expectation is equal to the correct judgment. If its expectation does not equal the correct judgment, then it is “biased,” and the bias associated with the judgment measures the degree of difference between the two.⁹³

Remember that while unbiasedness may be a desirable property, error reflects both bias and variance.⁹⁴ A judgment that is unbiased can still be highly inaccurate if it lacks precision — i.e., if it is subject to a high degree of judgment variability.⁹⁵ An accurate judgment entails both a high degree of precision and a low degree of bias.

An important concept in statistics (and estimation theory in particular) is the concept of a bias-variance tradeoff. This occurs when reducing error due to variance requires introducing some error due to bias.⁹⁶ For example, if a damages award for pain and suffering is highly unpredictable — such that repeatedly adjudicating the same case four times would result in awards of \$0, \$50,000, \$150,000, and \$200,000 — then we may wish to apply a procedure that reduces the unpredictability

⁹¹ *An Aggregation Theory of Character Evidence*, *supra* note 29, at 47-48; see also Edward K. Cheng, *When 10 Trials are Better than 1000: An Evidentiary Perspective on Trial Sampling*, 160 U. PA. L. REV. 955, 958-60 (2012); Michael J. Saks & Peter David Blanck, *Justice Improved: The Unrecognized Benefits of Aggregation and Sampling in the Trial of Mass Torts*, 44 STAN. L. REV. 815, 833-41 (1992).

⁹² *An Aggregation Theory of Character Evidence*, *supra* note 29, at 46-48.

⁹³ *Id.* at 46.

⁹⁴ *Id.*

⁹⁵ *Id.* at 47-48.

⁹⁶ *Id.* at 46-49.

of the award, even if doing so requires the introduction of some bias.⁹⁷ For example, we may wish to provide jurors with guidance in the form of information regarding pain-and-suffering awards in factually similar cases.⁹⁸ This guidance may introduce some degree of bias, since it involves awards from other cases; however, it is likely to substantially reduce unpredictability, and the accuracy gains from this increased precision (the inverse of variability) are likely to far offset any introduction of error due to bias.⁹⁹ If, for example, the correct damages award in a case is \$100,000, we may well prefer a trial process that, upon repeated adjudication, would generate the awards \$90,000, \$93,000, \$97,000, and \$100,000 than one that would generate the awards \$0, \$50,000, \$150,000, and \$200,000. This is because, even though the trial process that produces the former set of awards is biased whereas the trial process that produces the latter set of awards is unbiased, the former process involves only a slight bias but far less unpredictability than the latter process.¹⁰⁰

Now, armed with the concepts above, we are prepared to understand the statistical idea of shrinkage. Shrinkage is a method that aims to improve the accuracy of an estimation process by incorporating information regarding other events of the same general kind as the event in question.¹⁰¹ In short, it improves accuracy by contextualizing evidence and thereby capitalizing on the concept of a bias-variance tradeoff. More specifically, by incorporating information regarding other events of the same general kind as the event in question, shrinkage contextualizes evidence regarding the event in question; it thereby introduces some degree of bias to gain far more — in terms of accuracy — by reducing variance.¹⁰²

⁹⁷ See Hillel J. Bavli & Reagan Mozer, *The Effects of Comparable-Case Guidance on Awards for Pain and Suffering and Punitive Damages: Evidence from a Randomized Controlled Trial*, 37 *YALE L. & POL'Y REV.* 405, 416 (2019).

⁹⁸ See *id.* at 406-09 (proposing “comparable-case guidance” for awards for pain and suffering and punitive damages).

⁹⁹ *Id.* at 416.

¹⁰⁰ *Id.*

¹⁰¹ See Bradley Efron & Carl Morris, *Data Analysis Using Stein's Estimator and Its Generalizations*, 70 *J. AM. STAT. ASS'N* 311, 311-19 (1975); see also *An Aggregation Theory of Character Evidence*, *supra* note 29, at 48-49.

¹⁰² *An Aggregation Theory of Character Evidence*, *supra* note 29, at 48.

For example, if a sports statistician is halfway through a baseball season and is interested in predicting end-of-season batting averages for a group of, say, twenty baseball players, the statistician could use each player's half-season batting average to predict the player's end-of-season batting average. However, the statistician could improve her predictions — perhaps counterintuitively — by incorporating information regarding the half-season batting averages of the other nineteen baseball players into her predictions for each individual baseball player.¹⁰³ For example, to predict Roberto Clemente's end-of-season batting average from his batting average after his first 45 at-bats — e.g., a first-45 batting average of .400 in the 1970 season — rather than relying exclusively on Clemente's first-45 average to predict an end-of-season average of .400, the statistician would be better off combining Clemente's average of .400 with a value that reflects the batting averages of the other players (e.g., by computing a weighted mean of Clemente's average on the one hand and the mean of the other players' averages on the other hand).¹⁰⁴ This would improve the statistician's prediction by placing Clemente's batting average in context — by accounting for our expectations regarding batting averages more broadly.

As a second example, consider the pain-and-suffering award discussed above.¹⁰⁵ This type of damages award is notoriously unpredictable — i.e., it entails a high degree of judgment variability — such that three different juries hearing the same case may arrive at three very different awards.¹⁰⁶ However, as discussed above, one method for improving the accuracy of a pain-and-suffering award may involve providing jurors with guidance in the form of information regarding pain-and-suffering awards in factually similar cases.¹⁰⁷ This method provides jurors with greater context for their award. As such, it introduces some degree of bias by introducing numerical information regarding awards in other cases. However, any introduction of bias is likely to be far offset by the

¹⁰³ Efron & Morris, *supra* note 101, at 312-14.

¹⁰⁴ *Id.*

¹⁰⁵ See *supra* notes 97-100.

¹⁰⁶ Bavli & Mozer, *supra* note 97, at 406.

¹⁰⁷ *Id.* at 405-09 (reporting the results of a randomized experiment and concluding that there is strong evidence that guiding jurors with awards in factually similar cases improves the accuracy of awards for pain and suffering and punitive damages).

significant accuracy gains associated with the precision-enhancing benefits of the contextual information.¹⁰⁸

2. Character Evidence: A Bias-Variance Tradeoff

Arguably, we use shrinkage frequently, although implicitly, in legal settings.¹⁰⁹ Character evidence is one such setting.¹¹⁰ In a recent article, I define a category of evidence — called “aggregation evidence”¹¹¹ — that relies on principles of shrinkage estimation.¹¹² Specifically, aggregation evidence “involves inferring information about an event at issue in a case from information about distinct events of the same general kind.”¹¹³ Character evidence constitutes a particular form of aggregation evidence because it involves inferring information regarding an individual’s act from information regarding other acts committed by the same individual.¹¹⁴ For example, a prosecutor may offer evidence of prior assaults committed by the defendant in order to prove that the defendant committed the assault in question — that is, the prosecutor may offer evidence of distinct events of the same general kind as the event in question in order to allow jurors to make inferences from those events regarding the event in question.¹¹⁵

Like the baseball and pain-and-suffering examples above, character evidence involves a bias-variance tradeoff.¹¹⁶ In a sense, it contextualizes the event in question. It thereby simultaneously introduces error via bias and reduces error via variance. In particular, by introducing information regarding other acts, the evidence shifts the judgment in the direction of this broader context; however, the evidence simultaneously anchors the

¹⁰⁸ See *id.* at 416; *supra* notes 97–100.

¹⁰⁹ See *An Aggregation Theory of Character Evidence*, *supra* note 29, at Part 3.

¹¹⁰ *Id.*

¹¹¹ *Id.*

¹¹² *Id.* at 48.

¹¹³ *Id.* at 45.

¹¹⁴ *Id.* Character evidence in the form of opinion or reputation can similarly be understood as a form of aggregation evidence.

¹¹⁵ *Id.* Other examples of aggregation evidence include “comparables” evidence in damages assessments, takings cases, and other contexts; anecdotal evidence in discrimination cases; and certain forms of statistical evidence. *Id.*

¹¹⁶ See *id.* at 48–53.

judgment to a meaningful context — it has probative value — thereby reducing the variability of the judgment.¹¹⁷ Thus, like other forms of aggregation evidence, character evidence will improve accuracy when its precision-enhancing effect on a judgment outweighs its bias-introducing effect.¹¹⁸

However, because character evidence involves both a cost and a benefit — it involves introducing bias while reducing variability — an optimal use of this evidence with respect to accuracy will appropriately balance the two. This balance is based on the features of a case, and it determines how much weight to afford character evidence relative to other evidence in the case.¹¹⁹

Scientific principles of Bayesian inference, and shrinkage estimation in particular, tell us how much weight to afford a prior relative to new evidence in order to maximize accuracy. This provides a framework for determining how jurors use character evidence to determine a verdict.¹²⁰

To gain a deeper understanding of how shrinkage applies to character evidence, let us build on the concepts above to model a judgment using a “hierarchical” structure. Consider a criminal assault case. We can define the correct judgment — denoted by α_* — as a binary variable indicating whether the defendant in fact committed the assault ($\alpha_* = 1$) or did not commit the assault ($\alpha_* = 0$). The jury does not have perfect information regarding the facts surrounding the case. It therefore must arrive at an estimate — an actual judgment, denoted by $\hat{\alpha}_*$ — indicating whether the jury finds that the defendant committed the assault ($\hat{\alpha}_* = 1$) or did not commit the assault ($\hat{\alpha}_* = 0$).¹²¹ On average, the jury makes the correct judgment ($\hat{\alpha}_* = \alpha_*$), but not always, since the actual judgment ($\hat{\alpha}_*$) is variable (with variance = σ^2). In other words, if we adjudicated the case multiple times before different juries and judges, the various resulting judgments would be correct on average but would vary to some degree from adjudication to adjudication. This explains the lower level of the hierarchical structure.

¹¹⁷ *Id.*

¹¹⁸ *Id.* This tradeoff can arguably be expressed in terms of probative value and unfair prejudice. See generally FED. R. EVID. 403.

¹¹⁹ *An Aggregation Theory of Character Evidence*, *supra* note 29, at 49-53.

¹²⁰ I discuss behavioral assumptions underlying this framework *infra* Section II.D.

¹²¹ *An Aggregation Theory of Character Evidence*, *supra* note 29, at 46, 50-51.

For the upper level, imagine that there is some process that generates correct judgments — i.e., the true factual occurrences underlying the correct judgments — that are somehow bound together in some pattern. Consider, for example, the illustration above in which a statistician wishes to estimate the end-of-season batting averages for twenty players — and Roberto Clemente in particular — based on her knowledge of each player’s first 45 at-bats.¹²² The statistician’s estimate of Clemente’s end-of-season batting average is analogous to the actual judgment in the assault example and is described by the lower level of the hierarchical model.¹²³ The upper level of the model meanwhile tells us how the correct batting averages of the players (and Clemente’s average in particular) are bound together, or distributed (i.e., with a certain mean μ and a certain variance τ^2). Although each player has a different skill level, their batting averages are bound together by forces of nature that govern baseball players’ skill levels — even at the highest level of the game. The upper level of the hierarchical model tells us (via its parameters μ, τ^2) how these values are bound together.

Character evidence can similarly be described in terms of the upper level of the hierarchical model. After all, an individual’s character is a force that binds together the individual’s acts. Other-acts character evidence — and the factual occurrences (the other acts) underlying the evidence — can therefore be understood as arising from the same distribution as the act in question. In the assault example above, the upper level of the hierarchical model tells us (again, via its parameters μ, τ^2) how the individual’s acts are bound together, or distributed. Although different circumstances involve different behavioral responses, the individual’s acts are bound together (with some degree of consistency, represented by τ^2) by a certain force — that is, by the individual’s character. This is why other-acts character evidence is probative: it tells us about the force — an individual’s character — that gave rise not only to the other acts but also to the act in question.

More formally, assume that an actual judgment $\hat{\alpha}_*$ is distributed with mean α_* and variance σ^2 .¹²⁴ In other words, the actual judgment is “unbiased” — on average it will equal the correct judgment — but it

¹²² See *supra* notes 103–04 and accompanying text.

¹²³ See *The Logic of Comparable-Case Guidance*, *supra* note 90, at 15–19.

¹²⁴ *An Aggregation Theory of Character Evidence*, *supra* note 29, at 48.

entails some degree of randomness or variability around the correct judgment. The correct judgment α_* is, in turn, distributed with mean μ and variance τ^2 . Notationally, $\hat{\alpha}_* \sim (\alpha_*, \sigma^2)$, $\alpha_* \sim (\mu, \tau^2)$.¹²⁵ Further, denote the actual and correct judgments regarding other events (of the same general kind) offered as evidence as $\hat{\alpha}_i$ and α_i (respectively), where $\hat{\alpha}_i \sim (\alpha_i, \sigma^2)$, $\alpha_i \sim (\mu, \tau^2)$, for $i = 1 \dots n - 1$.¹²⁶ This simply means that the other events offered as evidence — for example, other-acts character evidence — arise from the same distribution as the event in question (e.g., with the same mean and variance), and that actual judgments regarding the other events (remember that a juror only hears evidence; she must still make fact judgments regarding the other events) are similarly unbiased with some degree of variance.

The actual judgments ($\hat{\alpha}_i$) regarding other events then provide information regarding the event in question (α_*) via information regarding the distribution that gave rise to it (and specifically, regarding μ and τ^2). In these terms, character evidence can be understood as providing information regarding the force or distribution (μ, τ^2) that binds together the defendant's acts ($\alpha_*, \alpha_1, \dots, \alpha_n$). I refer to this distribution as the propensity prior. Indeed, in Bayesian terms, character evidence can be understood as informing a prior that is then combined with evidence directly relevant to the event in question in order to facilitate a more accurate judgment.¹²⁷ This process is illustrated in Figure 1 below.

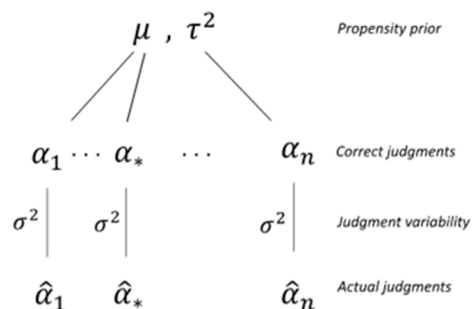
¹²⁵ *Id.*

¹²⁶ *See id.*

¹²⁷ Bayesian statistical analysis “uses the mathematical rules of probability to combine data with prior information to yield inferences which (if the model being used is correct) are more precise than would be obtained by either source of information alone.” Andrew Gelman, *Bayesian Statistics: What's It All About?*, STAT. MODELING, CAUSAL INFERENCE, & SOC. SCI. (Dec. 13, 2016), <https://statmodeling.stat.columbia.edu/2016/12/13/bayesian-statistics-whats/> [<https://perma.cc/5JEN-XKGU>]. As one author explained:

We often use probabilities informally to express our information and beliefs about unknown quantities. However, the use of probabilities to express information can be made formal: In a precise mathematical sense, it can be shown that probabilities can numerically represent a set of rational beliefs, that there is a relationship between probability and information, and that Bayes' rule provides a rational method for updating beliefs in light of new information.

Figure 1



Character evidence can be understood using a hierarchical model in which an individual's acts (and the correct judgments regarding those acts) are bound together by a central distribution that reflects the individual's character. I refer to this distribution as the "the propensity prior." Applying this model, other-acts character evidence can improve the accuracy of an actual judgment $\hat{\alpha}_*$ by providing information regarding the individual's other acts of a similar kind as the act at issue, and thereby providing information regarding the correct judgment α_* regarding the act at issue via information regarding the propensity prior — the probability distribution that gives rise to the act.

Thus, in the baseball example, when the statistician uses the other-player first-45 batting averages to improve her estimate of Clemente's end-of-season average,¹²⁸ she can be understood as viewing the upper level distribution (μ, τ^2) that reflects the other-player batting averages as a prior that she combines with new evidence generated by Clemente's first 45 at-bats to arrive at an accurate estimate of Clemente's end-of-season batting average.¹²⁹ In a sense, the prior represents the statistician's "prior beliefs" regarding the skill level of baseball players. It provides her with context regarding Clemente's first-45 average. After all, pursuant to the hierarchical model, this distribution is viewed as giving rise to Clemente's batting average as well as those of the other players — as reflecting the process that generates these averages and binds them together. Therefore, knowing information regarding this distribution

The process of inductive learning via Bayes' rule is referred to as *Bayesian inference*.

HOFF, *supra* note 27, at 1.

¹²⁸ See *supra* notes 103–04 and accompanying text.

¹²⁹ See *The Logic of Comparable-Case Guidance*, *supra* note 90, at 15-19.

indirectly provides information regarding Clemente's batting average. For example, it tells us how extreme Clemente's first-45 batting average is relative to that of other players.¹³⁰ If Clemente's average is extreme (which it is), then the statistician may attribute some of his average to chance, or "luck," and incorporate this into the end-of-season estimate.¹³¹

This is the idea behind shrinkage: instead of applying Clemente's first-45 average as the statistician's end-of-season estimate for Clemente, the statistician will "shrink" Clemente's end-of-season estimate in the direction of the mean of the other-player batting averages.¹³²

Similar to the baseball example, character evidence can be understood as informing a prior to be combined with new evidence. In a sense, character evidence provides some prior expectation regarding a defendant. Thus, assume that there is some probability (p) that the defendant in fact committed the assault (just as though there is some probability (p) that a coin flip will result in a "heads"). This probability value characterizes the upper level of the hierarchical model and the prior distribution — the propensity prior — that gives rise to the correct judgment and the underlying event at issue ($\alpha_* \sim (\mu, \tau^2)$, where $\mu = p$). The jury does not know the truth regarding whether the defendant committed the assault (i.e., the value of α_*), and it does not know the true nature of the defendant's character — including the probability with which the defendant committed the assault (the value of p). If the jury knew the nature of the defendant's character (including the value of p), it could not say with certainty whether or not the defendant committed the assault in question (the value of α_*), but this information could be helpful in arriving at a judgment (in estimating α_*) — just as knowing that a coin has an 80% probability of falling heads would be helpful for predicting a flip.

This is where character evidence comes in. In particular, other-acts character evidence ($\alpha_1 \dots \alpha_{n-1}$; $\hat{\alpha}_i \sim (\alpha_i, \sigma^2)$, $\alpha_i \sim (\mu, \tau^2)$) supplies the jury with information regarding the correct judgment (α_*) by supplying

¹³⁰ *Id.*

¹³¹ This can be stated in terms of the statistical concept of "regression toward the mean." *Id.* at 17.

¹³² See Efron & Morris, *supra* note 101, at 312-14. Note that shrinkage can also apply to move an estimate in the upward direction.

information regarding the propensity prior (μ, τ^2).¹³³ That is, it provides information regarding the defendant's character and thereby allows a better judgment by supplying additional information regarding the prior probability with which the defendant committed the assault in question.

More specifically, the parameters of the propensity prior (μ and τ^2) are fixed but unknown. Like the judgment itself (α_*), they must be estimated. Providing jurors with other-acts character evidence invites jurors to assess, or estimate, the defendant's character, and to use the other-acts character evidence to do so. In turn, jurors combine their character assessment with other evidence in a case.

To model how a juror combines her character assessment with other evidence in a case, let us now consider factors for determining how a prior is combined with new evidence to maximize accuracy in Bayesian statistics, and shrinkage estimation in particular. As suggested above, finding the optimal balance between character evidence and other evidence in a case depends on the bias-variance tradeoff. This balance can be stated in terms of two factors that essentially reduce to the simple idea of weighting character evidence in proportion to the value that it offers relative to the other evidence in a case.

First, the informational value of the other-acts character evidence is based on the precision of the character evidence (a reflection of τ^2 in the hierarchical model), which is based on the uniformity and the evidential strength of the evidence.¹³⁴ Greater uniformity means greater precision and therefore greater informational value: if the evidence involves other acts that have greater uniformity — or similarity, in terms of the acts and surrounding circumstances — with respect to one another and with respect to the act in question, the evidence provides sharper information regarding the defendant's character with respect to the act in question. On the other hand, a set of other acts that are factually dissimilar to one another or to the act in question is less valuable. Moreover, the value of the other-acts character evidence is based on the evidential strength of the evidence. Evidential strength depends on the number of "samples" of other acts and on the certainty that the evidence reflects the truth. For example, highly credible evidence involving three prior acts provides greater informational value about the act in question (α_*) via the

¹³³ See *An Aggregation Theory of Character Evidence*, *supra* note 29, at 49-53.

¹³⁴ *Id.* at 51.

defendant's character (and μ in particular) than either evidence involving poor credibility or a single prior act.¹³⁵ In turn, the higher the value of the character evidence, the more weight it will receive.

Consider the assault example discussed above. Character evidence that involves strong evidence regarding three prior acts that are similar to one another and similar to the act in question will have greater informational value — i.e., greater precision with respect to the judgment in question — than character evidence that involves either weaker evidence regarding the prior acts, fewer prior acts, or prior acts involving less uniformity (i.e., less similarity to one another or to the act in question). These features determine how the evidence is used to assess the defendant's character (i.e., to estimate the propensity prior) and, in turn, to arrive at a judgment regarding the act in question. In particular, the precision of the other-acts character evidence will determine how much weight to afford the character evidence. In the assault illustration immediately above, the precision of the character evidence will drive jurors to rely more heavily on the former (more precise) set of character evidence and less heavily on the latter (less precise) set of character evidence relative to the other evidence in the case.

However, the precision of the character evidence with respect to the judgment in question cannot be viewed in isolation. Its importance in determining the value of the character evidence — and the weight afforded to it — must be viewed in relation to the other evidence in the case. Therefore, the second factor is the precision of the other evidence in the case and, in turn, the precision of the judgment in the absence of the character evidence (a reflection of σ^2 in the hierarchical model). Specifically, the value of the character evidence depends on the unpredictability of the judgment (the inverse of the judgment's precision) — i.e., the judgment variability — in the absence of the character evidence.¹³⁶ This depends on the existing evidence in the case.¹³⁷

For example, if the evidence against a defendant is extremely strong, then ten out of ten juries would find the defendant guilty; however, if the evidence is mixed — in the sense that some evidence favors the

¹³⁵ *Id.* at 51-52.

¹³⁶ *Id.* at 52-53.

¹³⁷ *Id.*

defendant while some evidence favors the prosecution — or if there is a significant degree of uncertainty surrounding the evidence, then there will be a high degree of judgment variability and less consistency in the verdict.¹³⁸ Therefore, if in the absence of the character evidence, a case involves a high degree of judgment variability due to imprecise evidence, then character evidence can provide valuable guidance and is afforded more weight; if the case is clear cut (evidentiarily) without the character evidence, then the relative value of the character evidence — and the weight that is afforded to it — will be less.

Thus, character evidence involves benefits and costs. Its accuracy benefits rely on the principles of shrinkage estimation, and on the bias-variance tradeoff in particular. The value of the character evidence, and the weight that is optimally afforded to it, is based on both the precision of the character evidence and the judgment variability in the absence of the character evidence, which, in turn, is based on the existing (i.e., non-character) evidence in a case.

B. *The Shrinkage Principle*

In this Section, I consolidate the above factors to formulate a rule that I refer to as the shrinkage principle.

To summarize, if the existing evidence in a case is relatively weak and mixed, then judgment variability is likely high.¹³⁹ In this situation, character evidence is more valuable: it fills a substantial gap in the evidence, and it guides the jury to a more accurate judgment.¹⁴⁰ This is because the value of guiding the jury and thereby reducing judgment variability offsets any introduction of bias caused by the evidence.¹⁴¹ Moreover, if the character evidence is precise with respect to its evidential strength and its uniformity (internally and relative to the event in question), then it is more valuable: it provides a strong informational signal with respect to the event in question. The precision

¹³⁸ *Id.*

¹³⁹ *Id.*

¹⁴⁰ *Id.* at 53.

¹⁴¹ *Id.*

of the character evidence thus determines how well it can fill an evidentiary gap.¹⁴²

Therefore, character evidence will be more valuable and merit greater weight relative to the other evidence in a case if it is precise in terms of its uniformity and evidential strength, and if it relates to a matter that otherwise (i.e., in the absence of the character evidence) involves a high degree of variability.¹⁴³

Now, to state this conclusion more generally, recall that the discussion above applies principles of Bayesian inference by viewing character evidence in Bayesian terms as a prior: in a sense, as representing prior beliefs, or a prior expectation, regarding a defendant. This prior is combined with “new” evidence — that is, evidence regarding the act in question — to arrive at a judgment (or, in Bayesian terms, a posterior belief). Therefore, we can state the above conclusion in more general terms as follows:

The Shrinkage Principle: In combining a prior with new evidence, the prior should be weighted in proportion to its relative precision with respect to the matter in question. This means that the prior will be most valuable — and should therefore receive substantial weight — when it is highly precise with respect to the matter in question relative to the new evidence. Conversely, the prior will be least valuable — and should therefore receive little weight — when it is highly imprecise with respect to the matter in question relative to the new evidence.¹⁴⁴

Thus, pursuant to the shrinkage principle, character evidence will be most valuable, and will be relied on most heavily by an accuracy-

¹⁴² *Id.* at 51-53.

¹⁴³ *Id.* at 48-49.

¹⁴⁴ *Id.* We can formally define a judgment based on shrinkage, or a shrinkage “estimator” $\hat{\alpha}_s$, as a weighted average of the judgment that would emerge in the absence of any character evidence $\hat{\alpha}_{(*)}$ on the one hand, and μ on the other hand, weighted by the precision associated with each, $1/\sigma^2$ and $1/\tau^2$, respectively. We can therefore express the shrinkage estimator as follows: $\hat{\alpha}_s = \frac{\hat{\alpha}_{(*)} + \frac{\mu}{\tau^2}}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}$. *The Logic of Comparable-Case Guidance*, *supra* note 90, at 20-21 n.67. See generally Efron & Morris, *supra* note 101, 311-14 (discussing Stein’s paradox and methods of estimation in the context of predicting batting averages in baseball).

maximizing juror, when (1) it informs matters that involve a high degree of judgment variability due to poor existing evidence, and (2) it provides precise information regarding those matters.¹⁴⁵ Conversely, it will be less valuable, and will be relied on less heavily, when it informs matters that involve a low degree of judgment variability due to strong existing evidence in the case, or when it provides imprecise information regarding those matters.¹⁴⁶

C. The Influence of Prior Beliefs

In the previous Sections, I applied basic principles of Bayesian reasoning to introduce a model that describes the probative value of character evidence and the weight that a juror will afford to her assessment of the defendant's character relative to the other evidence in a case. I then stated a general rule for determining how heavily a prior should be weighted when combined with new evidence to arrive at a judgment. In the current Section, I build on this model to examine the role of a juror's prior beliefs and prejudices — and their relationship to character evidence — in the juror's determination of a judgment.

Let us begin by considering in greater detail how character evidence (evidence regarding other events $\alpha_1 \dots \alpha_{n-1}$) is incorporated in a juror's assessment of the defendant's character, and, in turn, the juror's estimate ($\hat{\alpha}_*$) of the correct judgment. We can view the defendant's character in relation to a specific trait (violence, truthfulness, etc.) as having two dimensions — one regarding the nature or direction (μ) of the trait (violent or non-violent, truthful or untruthful, etc.), and one regarding its consistency or predictability (τ^2). These dimensions characterize the propensity prior discussed above, and they provide information regarding how a defendant is likely to act in a certain situation and with what degree of consistency. Importantly, the propensity prior does not determine behavior; rather, it only predicts it probabilistically.

Previously, I showed how character evidence can be understood as informing a propensity-prior — reflecting the juror's perception of the defendant's propensity — that is combined with “new” evidence to arrive at a judgment. Now, however, I consider how exactly a juror uses

¹⁴⁵ *An Aggregation Theory of Character Evidence*, *supra* note 29, at 57-58.

¹⁴⁶ *See id.*

character evidence to inform this propensity-prior estimate. I do this by defining a prior on the propensity prior. Specifically, I define a prior for each of the parameters (or dimensions) associated with a juror's assessment of the defendant's character — a prior (θ) for the nature parameter (μ) and a prior (ψ) for the consistency parameter (τ^2). I refer to these priors as “propensity hyperpriors.”

This may seem complex at first glance, but it is not. The propensity hyperpriors (illustrated in Figure 2 below) simply indicate that the jurors have beliefs regarding the defendant's character prior to hearing character evidence. They reflect a juror's prior beliefs regarding the defendant's propensity — including the juror's preexisting prejudices and stereotypes, as well as her preexisting beliefs regarding the consistency of an individual's conduct.

Thus, to clarify the role of implicit bias in a judgment that is based on character evidence, I am proposing a model of character evidence that involves two applications of Bayesian reasoning, and the shrinkage principle in particular. First, a juror combines prior beliefs (the initial prior) — including her implicit biases — with character evidence (the initial “new” evidence) to form an assessment of the defendant's character (the posterior, or new set of beliefs); then, the juror combines her character assessment (the new prior) with the other evidence in the case to arrive at a judgment.

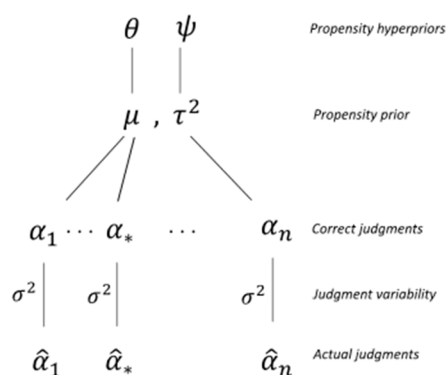
Immediately upon sitting for a trial, a juror begins to collect information regarding the defendant, and she begins to make judgments regarding the defendant's character.¹⁴⁷ It is well-accepted that we form impressions of others almost immediately, and that those initial — and perhaps instantaneous — impressions can be very influential on the opinion and image that we ultimately form regarding an individual.¹⁴⁸

¹⁴⁷ For a thorough discussion of how jurors form impressions of the defendant, see Teneille R. Brown, *The Content of Our Character*, 126 PENN. ST. L. REV. 1, 7-14 (2021).

¹⁴⁸ See *id.* at 6-7 (“Within 0.1 second of meeting anyone, we have already formed an impression of them — including when we are jurors listening to witness testimony. Almost immediately we have decided whether we think this person is honest, clever, likeable, or dangerous based on superficial information that is outside of their control.” (internal citations omitted)); see also Bennett Capers, *Evidence Without Rules*, 94 NOTRE DAME L. REV. 867, 867-71 (2019) (arguing that jurors rely on the appearance and background characteristics of a defendant, and that this “evidence” generally goes completely “unchecked”); discussion *infra* Section II.D.

Frequently, these impressions occur unintentionally, and often even subconsciously.¹⁴⁹ The juror observes the defendant’s appearance, demeanor, and background characteristics. These observations are combined with the juror’s preexisting general beliefs — conscious and unconscious — to form prior beliefs regarding the defendant,¹⁵⁰ or, in other words, to inform the juror’s propensity hyperpriors (θ and ψ). This often occurs implicitly and subtly rather than as an explicit consideration or examination.¹⁵¹

Figure 2



Character evidence can be understood using a hierarchical model in which an individual’s acts (and the correct judgments regarding those acts) are bound together by a central distribution that reflects the individual’s character — the propensity prior. Other-acts character evidence can improve the accuracy of a judgment by providing information regarding the propensity prior. However, a juror will assess the defendant’s character only by combining the character evidence with her prior beliefs and prejudices regarding the defendant —

¹⁴⁹ Brown, *supra* note 147, at 6-7 (“Humans are so motivated to infer people’s traits that, in the absence of information about how others have behaved, we instead rely on crude proxies such as race, dress, accent, and facial features to predict their personalities. We spontaneously infer whether someone is threatening, kind, intelligent or trustworthy based on social context and the superficial features of their face. We make these predictions immediately, unintentionally, and without even realizing we are doing so.” (internal citations omitted)).

¹⁵⁰ See *infra* Section II.D. The term “prior” suggests that these beliefs are formed before the juror even hears evidence regarding the defendant’s character.

¹⁵¹ See Brown, *supra* note 147, at 6-7.

reflected in the propensity hyperpriors θ and ψ . Based on the shrinkage principle, it is likely that these prior beliefs and prejudices play a relatively minor role in a judgment unless and until character evidence is introduced. Once a court permits the introduction of character evidence, however, the juror's prior beliefs and prejudices regarding the defendant play a major role in the judgment by heavily influencing the juror's assessment of the defendant's character.

The juror's prior beliefs (as reflected in the propensity hyperpriors) can impact a judgment or verdict. They can introduce unwanted and harmful bias into a verdict — bias based on race, sex, appearance, economic status, and other background characteristics. Indeed, various scholars have highlighted the significant effect of race, appearance, and other forms of “unregulated evidence” on trial outcomes.¹⁵²

However, to understand the interaction between character evidence and a juror's prior beliefs, we must first understand how a juror's prior beliefs impact a judgment in the absence of character evidence. I argue that the impact of a juror's prior beliefs and prejudices is generally significantly muted — it is not nearly at its full potential force — unless and until character evidence is introduced in the case. This conclusion follows from the shrinkage principle introduced in Sections II.A–B and is supported by the psychology literature discussed in Section II.D. Let us begin by understanding why it follows from the shrinkage principle.

The shrinkage principle states that a prior should be relied on most heavily when it informs a matter that involves a high degree of judgment variability due to poor existing evidence, and when it provides precise information regarding that matter; conversely, prior beliefs should not be relied on heavily when they relate to a matter for which existing evidence does or would provide a precise informational signal, or when the prior beliefs provide imprecise information regarding the matter.¹⁵³

Applying this principle, let us consider the precision of the prior relative to that of the evidence in the case. First, the juror's prior beliefs

¹⁵² Capers, *supra* note 148, at 867-71; *see also* Brown, *supra* note 147, at 1-14. Professor Brown has proposed that character evidence may in fact help to address a juror's reliance on impressions and unregulated evidence by, in a sense, counteracting unregulated evidence (or the juror's need for it in the first place) with sanctioned evidence regarding a defendant's character, and a defendant's good character in particular. *See* Brown, *supra* note 147, at 1-14.

¹⁵³ *See supra* Section II.B; *see also* *An Aggregation Theory of Character Evidence*, *supra* note 29, at 48-49, 58.

regarding a defendant are diffuse and imprecise with respect to the judgment in question. Remember, the juror is not yet asked to determine a character assessment. Rather, the juror is asked to determine a judgment regarding a matter in question in the case — e.g., whether the defendant committed a certain act or had a certain intent. The prior beliefs are therefore general and do not speak directly to the judgment in question. For example, if a juror is asked to determine whether a defendant committed an assault, the juror's prior beliefs and prejudices regarding the background characteristics of the defendant provide at most a highly vague and indistinct informational signal regarding the question of the assault in particular. Even if the juror holds strong implicit or explicit prior beliefs, these beliefs do not relate directly to the assault. Even from the juror's perspective, these prior beliefs are broad generalizations relative to the judgment in question. Moreover, the juror's prior beliefs based on the defendant's background characteristics are often implicit rather than overt, and the juror generally knows that it is impermissible to make judgments based on background characteristics such as race, sex, religion, and economic status, or based on character in the first instance.

Second, the juror's prior beliefs are even less precise — again, even from the juror's perspective — relative to the existing evidence (or lack thereof) in the case. For example, if a prosecutor brought a criminal action for assault, a juror would expect that the prosecutor will attempt to prove her case with concrete evidence. Generally, as a class of judgments, the question, did the defendant commit the assault, is not one for which there is particularly little or poor evidence.¹⁵⁴ Presumably, the prosecutor will present evidence in attempt to prove the government's case. If she does, it is likely that at least some evidence speaks directly to the assault — it is precise with respect to the judgment in question — and the juror's prior beliefs carry a highly imprecise informational signal relative to the concrete evidence in the case. On the other hand, if the prosecutor does not present evidence to prove the government's case, this also sends a relatively precise informational signal regarding the judgment in question.

¹⁵⁴ See *An Aggregation Theory of Character Evidence*, *supra* note 29, at 59-61.

The point is that, when a juror is asked to assess a non-character judgment (such as whether a defendant committed a certain act) based on concrete evidence, a broad impermissible generalization is likely to be overshadowed, or at least heavily muted, by the force of the actual evidence, or lack of evidence, directly related to the act in question. This is only accentuated when the court emphasizes the significance of the prosecutor's burden of proof and the importance of judging the defendant based on the evidence. Thus, pursuant to the shrinkage principle, the juror should place relatively little weight on the prior beliefs relative to the evidence.

Importantly, I do not mean to minimize the negative impact that implicit bias has on case outcomes. Rather, my intention is only to emphasize the very substantial role that implicit bias plays in case outcomes when the court allows the introduction of character evidence. Accordingly, the discussion herein refers to a muted effect of implicit bias only relative to its effect when a prosecutor introduces character evidence.

Now let us consider the role of a juror's prior beliefs and prejudices (and the propensity hyperpriors in particular) when character evidence is introduced. In this context the role of the juror's priors is very different than their role in the absence of character evidence. As before, a juror begins to form prior beliefs regarding the defendant's character as soon as the juror sits for trial, and perhaps even earlier.¹⁵⁵ The impact of these priors on the judgment is relatively muted until the prosecutor introduces character evidence against the defendant. At this point, however, the character evidence in a sense activates the juror's propensity hyperpriors, and, pursuant to the shrinkage principle, the juror's prior beliefs and prejudices then play a substantial role in the juror's judgment.

When a court permits a party to introduce character evidence, it asks jurors (implicitly or explicitly) to assess the defendant's character based on this evidence (i.e., to use the evidence to arrive at estimates $\hat{\mu}$ and $\hat{\tau}^2$ for μ and τ^2 , the parameters that characterize the nature and consistency of the defendant's character). A juror's character assessment is then explicitly and often sharply incorporated into the juror's judgment ($\hat{\alpha}_*$)

¹⁵⁵ See *supra* notes 149–51 and accompanying text.

— as sanctioned by the court via its admission of character evidence — pursuant to the shrinkage principle and well-accepted patterns of juror decision-making.¹⁵⁶

However, in arriving at her assessment of the defendant’s character, the juror will rely on both character evidence and her prior beliefs and prejudices regarding the defendant. Specifically, the juror will apply the shrinkage principle to combine her propensity hyperpriors (θ and ψ) with the character evidence offered in the case to determine an estimate of the defendant’s character (i.e., to determine $\hat{\mu}$ and $\hat{\tau}^2$).

Contrary to circumstances in which character evidence is not admitted in a case, when character evidence is admitted, the juror relies heavily on her prior beliefs and prejudices regarding the defendant in assessing the defendant’s character. In turn, the juror relies heavily on this character assessment in determining a judgment.

Specifically, in assessing the defendant’s character, pursuant to the shrinkage principle, the juror will weight her priors in proportion to their precision, with respect to the judgment in question, relative to the precision of the “new” evidence. In the context of a character assessment, this means that the juror will weight her prior beliefs and prejudices based on how precisely they inform the matter in question — the juror’s assessment of the defendant’s character — relative to the character evidence offered in the case.

Now — as opposed to the scenario above in which the court does not admit character evidence — the juror’s prior beliefs and prejudices regarding the defendant provide the juror with a precise informational signal with respect to the immediate matter in question, the defendant’s character.¹⁵⁷ The juror’s prior beliefs are perfectly symmetric to the

¹⁵⁶ See *supra* Section I.A; *infra* Section II.D.

¹⁵⁷ Remember that precision is distinct from accuracy. It reflects the factual uniformity and evidential certainty or strength (or absence of variability) surrounding the information. Information that is precise but biased can cause substantial inaccuracy. For example, if someone asks me for directions, and I indicate that I am certain about the directions, this signals informational precision. However, if I provide incorrect — although *precise* — directions, the recipient of the directions would likely have been better off had I conveyed imprecision and a weak informational signal — for example, by indicating that I am not quite sure of the correctness of my directions. The combination of precision and biasedness can lead to very incorrect conclusions and very harmful consequences.

judgment that the juror is tasked with making. That is, they speak precisely to the character assessment that the juror is asked to make. Although the juror's prior beliefs may arise from broad generalizations and inappropriate stereotypes and prejudices, they have likely developed and become ingrained over a long period of time, and they apply precisely to form inferences regarding the defendant's character.

Moreover, addressing the second component of the shrinkage principle, the juror's character judgment in the absence of her priors is highly imprecise. This is because, although the juror is asked to assess the defendant's character, she is provided with little information to do so. Jurors are generally asked to make character assessments and propensity inferences based on a small and biased sample of prior acts. Although prior acts indeed permit inferences regarding a defendant's character and later behavior, having a small and biased sample of prior acts detracts substantially from the informational value and the precision of the evidence.

Additionally, as suggested previously, when a juror hears character evidence, she will generally receive little or no information regarding the consistency of the defendant's behavior across changing circumstances. For example, jurors are generally not exposed to expert social science evidence regarding whether behavior in a certain context is dominated by character or context. Nor are jurors supplied with substantial evidence regarding the consistency of the defendant's behavior in particular. This absence of information regarding the consistency of the defendant's behavior leaves a large gap in the chain of inferences leading from character evidence to a character assessment and ultimately to an inference regarding the matter in question. This evidentiary gap represents a substantial source of imprecision in the character evidence with respect to the assessment that jurors are asked to make.

Thus, the precision of the juror's prior beliefs with respect to the juror's assessment of the defendant's character on the one hand, and the substantial imprecision of the character evidence on the other hand, result in a juror's heavy reliance on her prior beliefs and prejudices in assessing the defendant's character. By inviting a juror to assess the defendant's character and to make judgments based on her assessment of the defendant's character, while at the same time providing inadequate and certainly imprecise information for such character inferences, the court invites the juror to rely heavily on her prior beliefs

and prejudices in making judgments based on her assessment of the defendant's character.

Finally, once the juror arrives at her assessment of the defendant's character (i.e., her estimates $\hat{\mu}$ and $\hat{\tau}^2$ for μ and τ^2), then pursuant to the shrinkage principle and the discussion in Sections II.A–B, this character assessment is combined with the non-character evidence in the case to determine the judgment in question ($\hat{\alpha}_*$). As suggested in the Introduction and Part I, character evidence and the juror's assessment of a defendant's character in particular — and therefore her prior beliefs and prejudices regarding the defendant — can be extremely impactful on a juror's judgment and determination of a verdict.¹⁵⁸ This conclusion is well-supported in the case law and scholarship, and it can be explained by concepts such as confirmation bias — the tendency to interpret new evidence in a way that confirms prior beliefs — and other phenomena in the psychology of juror decision-making.¹⁵⁹

It is important to realize that the effects described in this Section occur when a court admits character evidence — regardless of whether a juror misuses the evidence by weighting it excessively or punishing a defendant based on past bad acts. Indeed, the court invites the juror's reliance on her prior beliefs. This contributes substantially to the perniciousness of these effects. Moreover, however, the risk that a juror will misuse character evidence in these ways only compounds the

¹⁵⁸ See *United States v. Burkhart*, 458 F.2d 201, 204–05 (10th Cir. 1972) (“[A]n obvious truth is that once prior convictions are introduced the trial is, for all practical purposes, completed and the guilty outcome follows as a mere formality. This is true regardless of the care and caution employed by the court in instructing the jury.”); Capra & Richter, *supra* note 7, at 770–87 (discussing the rule against character evidence and commenting regarding the “dramatic” impact that character evidence has on a jury); Milich, *supra* note 6, at 780 (“Once the jury learns that the defendant has a criminal past, the odds of conviction skyrocket.”); see also discussion *supra* Section I.A; discussion *infra* Section II.D.

¹⁵⁹ See Brown, *supra* note 147, at 31–36 (discussing “a person-centered account of blame” and concluding that “[m]ost of us judge people not for what they do or cause, but because of the kinds of people they are. We think that ‘good’ people deserve less punishment, and ‘bad’ people deserve greater punishment, for the same bad acts.” (citing sources)); Raymond S. Nickerson, *Confirmation Bias: A Ubiquitous Phenomenon in Many Guises*, 2 REV. GEN. PSYCH. 175, 175 (1998); discussion *infra* Section II.D.

harmful effects discussed herein.¹⁶⁰ Specifically, because the juror's character assessment relies heavily on her implicit biases, these prior beliefs and prejudices will be even more influential if jurors give undue weight to such character assessments or punish defendants directly based on such assessments rather than the act in question in the case.

Thus, let us summarize the reflective model of character evidence with the following sequence that begins with a juror's prior beliefs and prejudices regarding the defendant's character and ends with a judgment that is strongly influenced by these prior beliefs and prejudices:

- The juror observes the defendant and forms priors regarding the defendant's character. These priors inform the propensity hyperpriors (θ and ψ), which can be understood simply as the juror's beliefs regarding the probability that the defendant will commit a particular act in a certain circumstance.
- The juror observes the introduction of character evidence (as well as the defendant's demeanor, behavior, and testimony throughout the trial). The juror is asked, implicitly or explicitly, to assess the defendant's character and to make propensity inferences in determining a verdict.
- The juror combines her prior beliefs regarding the defendant's character (θ and ψ) with court-sanctioned character evidence to arrive at an assessment of the defendant's character ($\hat{\mu}$ and $\hat{\tau}^2$). Applying the shrinkage principle, the juror in a sense fills substantial evidentiary gaps with her prior beliefs and prejudices, which are relatively precise with respect to the matter at hand. The juror thereby places substantial weight on these priors in assessing the defendant's character.
- The juror applies her character assessment ($\hat{\mu}$, $\hat{\tau}^2$) to arrive at a judgment ($\hat{\alpha}_*$) — i.e., an estimate of the correct judgment (α_*). Specifically, pursuant to the shrinkage

¹⁶⁰ See *People v. Zackowitz*, 172 N.E. 466, 468 (N.Y. 1930) (citing 1 JOHN HENRY WIGMORE, WIGMORE ON EVIDENCE § 194 (1904)); FISHER, *supra* note 11, at 153; see also *supra* Section I.A.

principle, she combines her character assessment with the non-character evidence in the case, weighting it heavily to fill inevitable evidentiary gaps and arrive at a judgment ($\hat{\alpha}_*$). Accordingly, her judgment is heavily influenced by her prior beliefs and prejudices (θ and ψ).

D. *The Psychology of Juror Decision-Making*

The proposed model — the reflective model of character evidence — finds substantial support in the psychology of juror decision-making.

The goal of this Article is to propose a testable theory of how the introduction of character evidence causes jurors to rely heavily on their implicit biases in determining a verdict. This question has critical implications for policy surrounding the admissibility of character evidence. The reflective model suggests that character evidence in a sense activates a juror's prior beliefs and prompts their influence in a case. However, other behavioral models are possible — including some that suggest a reverse effect by which character evidence may reduce the influence of implicit bias.¹⁶¹ Ultimately, how character evidence interacts with a juror's implicit biases must be resolved empirically — perhaps with well-designed experimental studies that test the interaction effects described above. However, the reflective model has a strong theoretical foundation that finds substantial support in our understanding of juror decision-making.

First, as a starting point, the reflective model provides a natural and intuitive prediction of juror behavior. This is because its central assumption is that jurors are accuracy-maximizers in the sense that they at least seek to optimize the information provided to them to arrive at the correct conclusion.¹⁶² As explained in Sections II.A–B, the Bayesian principles on which the model relies aim to utilize different sources of information (i.e., prior beliefs and new evidence) in proportion to their relative value in order to maximize accuracy.

¹⁶¹ See Brown, *supra* note 147, at 11-12; discussion *infra* note 183; see also Evelyn M. Maeder & Jennifer S. Hunt, *Talking About a Black Man: The Influence of Defendant and Character Witness Race on Jurors' Use of Character Evidence*, 29 BEHAV. SCI. & L. 608, 608-18 (2011) (reporting empirical results and concluding that character evidence may be more influential when *inconsistent* with stereotypes based on race).

¹⁶² See *supra* Sections II.A–C.

This is not to say that jurors in fact behave like perfect statisticians. As we know from decades of research into the irrational quirks of human behavior, people are far from perfectly rational — and jurors are no exception.¹⁶³ However, unlike behavior in everyday life, jurors sit for trial with the purpose of finding the correct answer — of making accurate fact judgments and arriving at a correct verdict based on them.¹⁶⁴ To be sure, the jury has numerous important functions in our society; however, the most basic function of the jury is fact-finding, truth-seeking, and determining a correct verdict.¹⁶⁵ This goal is stated plainly to the jury, and it serves as the most basic underlying purpose of the trial and the rules of evidence.¹⁶⁶

Moreover, there is substantial support for the proposition that jurors take this task seriously — they carefully examine the evidence in a case to find the truth regarding what happened in the incident in question and to arrive at the correct verdict.¹⁶⁷ Additionally, although errors in jury

¹⁶³ See DAN ARIELY, *PREDICTABLY IRRATIONAL: THE HIDDEN FORCES THAT SHAPE OUR DECISIONS* (2008); DANIEL KAHNEMAN, *THINKING FAST AND SLOW* (2011); STEVEN PINKER, *RATIONALITY: WHAT IT IS, WHY IT SEEMS SCARCE, WHY IT MATTERS* (2021); RICHARD H. THALER, *MISBEHAVING: THE MAKING OF BEHAVIORAL ECONOMICS* (2015).

¹⁶⁴ See FED. R. EVID. 102 (“These [evidence] rules should be construed so as to administer every proceeding fairly, eliminate unjustifiable expense and delay, and promote the development of evidence law, to the end of ascertaining the truth and securing a just determination.”); *Estes v. Texas*, 381 U.S. 532, 540 (1965) (“Court proceedings are held for the solemn purpose of endeavoring to ascertain the truth . . .”); FISHER, *supra* note 11, at 1; LEONARD M. NIEHOFF, *EVIDENCE LAW 1* (2016).

¹⁶⁵ See FISHER, *supra* note 11, at 1; NIEHOFF, *supra* note 164, at 5-6.

¹⁶⁶ See FED. R. EVID. 102; *Estes*, 381 U.S. at 540; see also FED. JUD. CTR., *BENCHBOOK FOR U.S. DISTRICT COURT JUDGES* § 7.08, at 268-69 (6th ed. 2013) (providing juror oaths that include swearing or affirming that jurors will “render a true verdict according to the law and the evidence”); ADMIN. OFF. OF THE U.S. COURTS, *HANDBOOK FOR TRIAL JURORS SERVING IN THE UNITED STATES DISTRICT COURTS* 8-9.

¹⁶⁷ See Brian H. Bornstein & Edie Greene, *Jury Decision Making: Implications for and from Psychology*, 20 *CURRENT DIRECTIONS PSYCH. SCI.* 63, 65 (2011) (“[T]here is plenty of evidence that jurors also use careful, systematic processing strategies.”); Dennis J. Devine, Jennifer Buddenbaum, Stephanie Houp, Dennis P. Stolle & Nathan Studebaker, *Deliberation Quality: A Preliminary Examination in Criminal Juries*, 4 *J. EMPIRICAL LEGAL STUD.* 273, 300 (2007) (“Juries reported . . . thoroughly reviewing their evidence . . .”).

verdicts are well-documented, juries are generally thought to perform their task of accurate fact-finding reasonably well.¹⁶⁸

Second, there is empirical evidence to support the proposition that jurors are Bayesian thinkers and that Bayesian models perform well in predicting juror behavior.¹⁶⁹ Moreover, a recent experimental study provides evidence that explicitly supports a model of juror behavior grounded in shrinkage estimation.¹⁷⁰ In particular, although not a study pertaining to character evidence, a large randomized controlled trial revealed that mock jurors incorporate information regarding damage awards in prior cases — another form of aggregation evidence — in line with the predictions of the shrinkage principle discussed above in arriving at an award judgment.¹⁷¹ Importantly for purposes of the current Article, the study found that mock jurors even weighted this aggregation evidence (relative to other evidence) based on the relative precision of the aggregation evidence, as predicted by the shrinkage principle and the model herein.¹⁷²

Finally, in addition to the arguments above, there is a direct psychological basis for the model's prediction (based on the shrinkage principle) that character evidence in a sense activates a juror's priors

¹⁶⁸ Bornstein & Greene, *supra* note 167, at 65-66 (describing studies and suggesting that “jurors perform their duties reasonably well”).

¹⁶⁹ See Tamara Shengelia & David Lagnado, *Are Jurors Intuitive Statisticians? Bayesian Causal Reasoning in Legal Contexts*, *FRONTIERS PSYCH.* (Feb. 5, 2021), <https://doi.org/10.3389/fpsyg.2020.519262> [<https://perma.cc/FL2K-K6V5>] (reviewing literature, reporting experimental results, and concluding that “[o]verall the results indicate that people’s qualitative reasoning is mostly accurate and follows qualitative predictions of Bayesian models in predictive, diagnostic and explaining away inferences. These findings reinforce results from previous studies where Bayesian probabilistic reasoning was observed”). *But see* Paul Thagard, *Testimony, Credibility, and Explanatory Coherence*, 63 *ERKENNTNIS* 295, 295-96 (2005) (arguing that the “explanatory-coherence account is more plausible . . . than a Bayesian account”). *See generally* Jesse Bull & Joel Watson, *Statistical Evidence and the Problem of Robust Litigation*, 50 *RAND J. ECON.* 974, 974-82, 991 (2019) (applying a Bayesian model to describe a jury’s evaluation of hard evidence and characterizing Rule 404’s prohibition on character evidence as “stat[ing] that, in the language of [the authors’] model, the face-value signal of character evidence is so outweighed by the potential litigant-type signal as to make this kind of evidence unreliable for robust litigation”).

¹⁷⁰ *See* Bavli & Mozer, *supra* note 97, at 432-50.

¹⁷¹ *Id.*

¹⁷² *Id.* at 446-48.

such that the juror will rely heavily on her implicit biases when character evidence is introduced and less heavily when it is not.¹⁷³ In particular, the psychology literature regarding “category activation” — the triggering of “categorical representations to simplify and streamline the person perception process” — supplies substantial evidence that an individual’s implicit biases may remain relatively muted unless and until they are activated by certain stimuli.¹⁷⁴ As one article states: “According to recent thinking on the topic, mere exposure to a stereotyped target may be insufficient to trigger category activation. Instead, activation [and “associated cognitive contents”¹⁷⁵] may only occur under certain precipitating conditions.”¹⁷⁶ In turn, the empirical studies in this regard have evidenced that “two factors appear to play a prominent role in the regulation of category activation: perceivers’ temporary processing goals and their general attitudes (i.e., prejudice level) toward the members of the category in question.”¹⁷⁷

These findings are supportive of the reflective model and the factors of which the shrinkage principle is comprised. Specifically, the two factors that are found to regulate category activation can be interpreted as capturing the idea of relative precision with respect to the matter in question as the central element in determining the weight that is afforded to a prior belief regarding a category (e.g., race or gender) to which an individual is perceived to belong.¹⁷⁸ They support the idea that implicit biases regarding the background characteristics of a defendant remain relatively subdued when there is no character evidence and a juror’s prior beliefs are relatively imprecise with respect to the matter in question, or, in other terms, less relevant to the “temporary processing goals” of the juror — e.g., whether the defendant committed the act in question.¹⁷⁹ They also support the converse: when character evidence is introduced, and a juror’s prior beliefs are highly precise with respect to

¹⁷³ See *supra* Section II.C.

¹⁷⁴ C. Neil Macrae & Galen V. Bodenhausen, *Social Cognition: Thinking Categorically About Others*, 51 ANN. REV. PSYCH. 93, 96 (2000).

¹⁷⁵ *Id.*

¹⁷⁶ *Id.* at 98.

¹⁷⁷ *Id.* at 99 (citations omitted).

¹⁷⁸ *Id.*; see *supra* Sections II.B–C.

¹⁷⁹ Macrae & Bodenhausen, *supra* note 174, at 99-100 (citing multiple studies). A juror’s high “cognitive load” only further supports this point. *Id.* at 99.

the matter in question, or, in other terms, directly relevant to the “temporary processing goals” of the juror — i.e., the goal of assessing the defendant’s character — the juror’s implicit biases are activated and will have a substantial influence in the juror’s judgment.¹⁸⁰

Moreover, the psychology literature regarding “confirmation bias” provides additional support for the reflective model’s predictions. Confirmation bias is a well-studied behavioral phenomenon whereby observers — including jurors — tend to construe new evidence in accordance with, or in support of, their prior beliefs.¹⁸¹ The empirical scholarship in this regard, combined with the category-activation literature discussed above, further supports the model’s predictions that when — and particularly when — character evidence is introduced in a case (1) a juror’s prior beliefs will tend to be highly influential on her assessment of the defendant’s character, and (2) the juror’s assessment of the defendant’s character (which, in turn, relies heavily on her prior beliefs) will tend to be highly influential on her evaluation of the non-character evidence in the case.

In summary, although additional empirical examination is necessary to evaluate the hypothesis that character evidence activates the heavy influence of implicit bias in juror decision-making, the reflective model finds theoretical support in the psychology of human — and, in particular, juror — decision-making.

¹⁸⁰ *Id.* at 99-100; see also Ziva Kunda & Steven J. Spencer, *When Do Stereotypes Come to Mind and When Do They Color Judgment? A Goal-Based Theoretical Framework for Stereotype Activation and Application*, 129 *PSYCH. BULL.* 522, 525, 540 (2003) (“[W]hen one is motivated by comprehension goals to predict a person’s attributes, an applicable stereotype can contribute to this only if it contains information deemed relevant.”). See generally SAKS & SPELLMAN, *supra* note 43, at 36-39 (arguing that “[f]or the most part . . . evidence is what drives most jurors most of the time to make the decisions they do,” but highlighting exceptions in which “juror characteristics and predispositions become more important to the trial’s outcome” — including circumstances in which “personal attributes are made an issue in the case,” thereby “render[ing] jurors’ own group identities salient”); Jonathan J. Koehler, *The Base Rate Fallacy Reconsidered: Descriptive, Normative, and Methodological Challenges*, 19 *BEHAV. & BRAIN SCIS.* 1, 5-11 (1996) (describing when and to what extent base rates are used by decision makers); *id.* at 12-13 (describing the relationship between base rates and prior beliefs, or “prior probability estimates”).

¹⁸¹ Nickerson, *supra* note 159, at 175.

III. DISCUSSION AND IMPLICATIONS

A. *The Interaction Between Character Evidence and Implicit Bias*

Understanding how character evidence interacts with a juror's prior beliefs and prejudices is of critical importance. The model introduced in the previous Part implies that, based on the unique nature of character evidence, this type of evidence swings open the door to a juror's prior beliefs and prejudices that may well have otherwise remained relatively uninfluential in a case. In this Section, I highlight two features of character evidence that distinguish it from other types of evidence in terms of its tendency to invoke a juror's prior beliefs regarding a defendant as a central component of the juror's judgment.

First, as discussed above, character evidence invites jurors to make inferences regarding the defendant's character. Throughout the trial, and from the moment that jurors sit for trial, they form prior beliefs regarding the defendant's character via preexisting generalizations, stereotypes, and prejudices. As discussed above, this is inevitable.¹⁸² These inferences, although prejudicial and certainly problematic, remain relatively subdued when character evidence is not introduced. This is based on the shrinkage principle, as well as a number of other factors (including, e.g., the implicit nature of the priors, the court's instructions to the jury, and the impermissibility of character inferences).

However, when a court admits character evidence against a defendant, the prosecutor and court ask jurors to assess precisely what they have been forming prior beliefs and prejudices about — the defendant's character — and all while providing little other information to guide a juror's assessment. In other words, based on the shrinkage principle and psychology discussed in Part II, character evidence is unique in its ability to activate the influence of prior beliefs. This is due to the relative precision of a juror's prior beliefs with respect to the assessment in question when a court admits character evidence.¹⁸³

¹⁸² See *supra* notes 147–52 and accompanying text.

¹⁸³ Consider how the model and argument herein compare to Professor Brown's recent argument that, because jurors inevitably resort to instantaneous impressions based on appearance, the ban on character evidence should be reversed and replaced "with a rule that permits moral and neutral character evidence and presumes inadmissibility for character evidence indicating an immoral trait." Brown, *supra* note

147, at 50-51. According to Professor Brown, “The ban on character evidence only heightens our reliance on unreliable facial traits. If we deny jurors information about how an accused has behaved in the past, they will instead subconsciously rely on immutable facial characteristics rooted in race, class, or sex, that will be even more inaccurate and unfair.” *Id.* at 11. However, “the effect of automatic facial impressions may be mitigated by hearing about a witness’s past acts. Counterintuitively, to achieve the goals of the character evidence ban, we should permit more character evidence, rather than less.” *Id.* at 11-12 (citation omitted).

Professor Brown’s theory arrives at a different conclusion than the argument herein: it promotes a reversal of the ban on character evidence in favor of a more permissive approach, whereas the argument herein promotes a stricter rule against character evidence. Nevertheless, we agree on two important premises — that jurors form instantaneous impressions regarding a defendant’s character based on appearance, and that one effect of more *sanctioned* evidence regarding a defendant’s character can, at least under some circumstances, cause less reliance on instantaneous impressions by, for example, filling informational gaps with good-character evidence to replace a juror’s bad-character impressions based on appearance. Indeed, in some circumstances, this premise can find support in the shrinkage principle discussed in Part II.

However, my model quickly diverges from Professor Brown’s theory. Specifically, the model herein implies that, based on principles of Bayesian inference, character evidence (while providing some information) *invites and encourages* jurors to rely on their prior impressions, beliefs, and prejudices, and that, in most circumstances, any information-filling role of character evidence will be overshadowed by its effect on encouraging a juror’s reliance on prior beliefs and prejudices. This effect of character evidence — that it is the character evidence itself that invites the full force of prior beliefs and prejudices — is central to the reflective model proposed herein.

Note that I do not take a strong position in the current Article regarding circumstances in which a defendant elects to introduce good-character evidence. In this context, it is possible that sanctioned character evidence could play an important role in counteracting prior beliefs. However, contrary arguments exist, and this issue is complex when considered more broadly. *See infra* Section III.B. My focus in the current Article is primarily on other-acts character evidence, bad-character evidence offered against defendants and other parties, and more broadly, the many exceptions to, and frequent departures from, the rule against character evidence.

The central difference in assumptions between Professor Brown’s approach and the model herein results in divergent conclusions in many applications. For example, Professor Brown argues that, applying her proposal, “there will likely be more prosecutions and civil claims for sexual assault cases” and “[t]here will almost certainly be far fewer appeals based on the character evidence rules,” since judges will ultimately “have discretion in applying the rule.” Brown, *supra* note 147, at 55. However, as discussed in the following Section, although I agree fully with Professor Brown’s statements that “at present, rape is significantly under-prosecuted” and that “[p]rosecutors are reluctant to indict, and plaintiffs are reluctant to sue, when evidence of the defendant’s past sexual assaults cannot be heard by the jury,” — and, more broadly, that the evidentiary issues in

Second, more than just permitting a juror to rely on prior beliefs and prejudices, by inviting character inferences, the prosecutor and court invite — and, in a sense, require — jurors to rely on their priors by soliciting a character assessment while providing very limited information on which to base such an assessment. A juror is expected to arrive at a character assessment based on only a small and biased sample of prior acts. Moreover, contrary to other forms of evidence, the relevance of other-acts character evidence is based on particularly crucial but ambiguous assumptions regarding the consistency of a defendant's behavior with respect to the act in question. Jurors are generally not instructed on how to apply propensity reasoning: they are not provided with social-science evidence regarding an individual's tendency to act in accordance with a propensity notwithstanding changing circumstances, and they are generally not provided with a sufficient number of prior acts to make inferences with respect to such consistency. Instead, they hear evidence regarding other acts committed by a defendant and are left to their own devices — or their prior beliefs and prejudices — to determine what to do with it.¹⁸⁴

In summary, character evidence is not like other evidence. It invites jurors to make character inferences, and with them a range of inferences based on the jurors' prior beliefs and prejudices regarding a defendant. Moreover, these priors do not simply reflect biases that would otherwise permeate a trial. Rather, by inviting character reasoning, the prosecutor

sexual-assault cases provide a strong argument for the admissibility of character evidence — I argue that the admissibility of character evidence in sexual-assault cases will inevitably invite and encourage jurors to rely on their prior beliefs and prejudices. *Id.*; see *infra* Section III.B. It inevitably introduces substantial prejudice and inequality in the outcomes of these cases. Importantly, I do not take a normative position regarding character evidence in these cases. As Professor Brown highlights, this evidence often plays a crucial role in encouraging actions and permitting successful prosecutions. I only aim to highlight the implications for prejudice and inequality as a critical consideration. Future analysis may well produce a method for mitigating such effects while upholding the crucial objectives of the sexual-assault exceptions to the rule against character evidence.

¹⁸⁴ Contrary to other types of evidence, even if two jurors identically interpret the strength and nature of the character evidence presented to them, it is very possible that one juror will attach substantial importance to the evidence while the other juror attaches little or no importance to it.

and court may activate the influence of priors that would otherwise remain relatively subdued.

B. Formal Exceptions to the Rule Against Character Evidence

The analysis above implies that formal exceptions to the rule against character evidence may swing open the doors to a wide range of unwanted prior beliefs and prejudices held by jurors. These exceptions may thereby initiate or exacerbate unequal treatment based on race, sex, appearance, accent, education, economic status, and other background characteristics of a defendant. As highlighted in the reflective model described in Part II, the priors that give rise to this unequal treatment are likely inherent in formal exceptions to the rule against character evidence.

Some such exceptions — such as those permitting a criminal defendant to introduce good-character evidence — must be initiated by the defendant.¹⁸⁵ The reflective model implies that these exceptions may be used to some defendants' advantage more than to other defendants' advantage based on the background characteristics of the defendant. It is possible, for example, that a juror will weight good-character evidence heavily for a wealthy and well-educated defendant, and weakly for an impoverished and poorly-educated defendant, or that the juror will allow prior beliefs based on race or sex to dominate any good-character evidence. Moreover, once a defendant “opens the door” to character evidence, the prosecutor may take advantage of various avenues to introduce bad-character evidence against the defendant. A juror can similarly weight this evidence differently for different groups of defendants and may similarly allow prior beliefs to permeate or dominate her assessment of the defendant's character.

However, the exception for good-character evidence is ultimately available for the defendant's benefit should the defendant choose to take advantage of it notwithstanding its risks. I therefore put aside this category of evidence and reserve a more thorough analysis of the reflective model's implications for it for future work.

In the remainder of this Section, I focus on exceptions to the rule against character evidence that involve evidence of a defendant's bad

¹⁸⁵ See FED. R. EVID. 404(a)(2)(A).

character. I focus specifically on perhaps the most well-known of such exceptions in the criminal context — the exception for “[s]imilar [c]rimes in [s]exual-[a]ssault [c]ases” under Rule 413 of the FRE.¹⁸⁶ This rule provides: “In a criminal case in which a defendant is accused of a sexual assault, the court may admit evidence that the defendant committed any other sexual assault. The evidence may be considered on any matter to which it is relevant.”¹⁸⁷

Rule 413 provides for a broad exception to the ban on character evidence in sexual assault cases. Under this rule, jurors are invited to use prior allegations of sexual assault, or other evidence of prior sexual assault — whether or not the defendant was charged or convicted of a crime — for the purpose of making the inference that the defendant has a character for committing sexual assault and is likely to have acted in accordance with that character.¹⁸⁸ It is substantive evidence that, together with an allegation that the defendant has committed the sexual assault in question, is sufficient to form the basis of a conviction.

The primary rationales given for this exception are (1) that sexual assault crimes involve a higher rate of recidivism than other types of crime; and (2) that there is often little evidence in sexual assault cases, and character evidence makes it feasible to prosecute these cases.¹⁸⁹ Indeed, character evidence is frequently of critical importance in encouraging actions and permitting successful prosecutions.¹⁹⁰ The unique evidentiary issues and the absence of evidence in many sexual-assault cases provide strong arguments for the admissibility of character evidence in these cases — arguments that find some support in the Bayesian principles discussed in Part II.

My intention in this Section is not to evaluate these rationales or to provide a normative analysis regarding the sexual-assault exception to the rule against character evidence or the policy concerns surrounding it. Rather, it is only to examine one critical consideration implied by the model herein — the concern that the exception for character evidence in sexual-assault cases generates substantial prejudice and inequality in

¹⁸⁶ FED. R. EVID. 413.

¹⁸⁷ *Id.*; see also FED. R. EVID. 414 (“Similar Crimes in Child-Molestation Cases”).

¹⁸⁸ FED. R. EVID. 413.

¹⁸⁹ See FISHER, *supra* note 11, at 207-22.

¹⁹⁰ See Brown, *supra* note 147, at 55.

these cases by inviting and encouraging jurors to rely on their prior beliefs and prejudices based on the race, appearance, economic status, and other background characteristics of the defendant.

Let us examine how the reflective model would apply to character evidence in a sexual-assault trial. First, from the moment the jury sits for trial, the jurors know that the defendant is on trial, charged with sexual assault. Before any evidence is introduced, the jurors are already informing their propensity hyperpriors with prior beliefs and prejudices based on the defendant's appearance and background characteristics, including the defendant's race, sex, presumed economic status, etc. The jurors continue to inform these hyperpriors throughout the trial.

When jurors are asked to make non-character judgments, such as whether the defendant committed the act in question, then based on the shrinkage principle and the factors previously discussed, their prior beliefs and prejudices remain relatively subdued and uninfluential. However, in sexual assault cases, this is not what occurs. Instead, jurors may well be asked to arrive at a judgment by making character-propensity inferences.

When this happens, a juror's mode of reasoning changes. A character assessment is not like assessing whether there is evidence proving that the defendant committed the act in question. For the reasons discussed, character inferences inherently require a juror to rely on her prior beliefs and prejudices. Meanwhile, the juror has been developing an extensive set of prior beliefs regarding precisely what the juror is now asked to assess — the defendant's character.¹⁹¹

¹⁹¹ One way in which a juror may combine her prior beliefs and prejudices with sanctioned evidence in a case is by observing all evidence with reference to her prior beliefs, and applying the evidence to confirm these beliefs. This is a form of confirmation bias, discussed in Section II.D. See generally John Rafael Peña Perez, *Confronting the Forensic Confirmation Bias*, 33 YALE L. & POL'Y REV. 457, 459-60 (2015) (discussing confirmation bias in the context of forensic evidence). For example, the juror's prior beliefs may inform how much weight to afford other-acts character evidence. If the juror's prior beliefs are consistent with evidence of the defendant's prior sexual assault, then she may weight the character evidence heavily; whereas if her prior beliefs are inconsistent with the other-acts character evidence, she may afford the evidence little weight. Remember that contrary to many other forms of evidence, a juror can easily justify weighting character evidence heavily or lightly, regardless of whether she credits the evidence or not. This is partly because the probative value of character evidence depends on an inferential link for which the courts provide little or no guidance: how

In fact, there is reason to expect that a juror's prior beliefs and prejudices will be even more influential in a sexual-assault case than in other types of cases. This is because there is often an absence of direct evidence regarding the act in question. This absence of evidence forms one of the primary rationales underlying the exceptions to the ban on character evidence in sexual-assault cases. However, it also naturally exacerbates the effect of a juror's prior beliefs and prejudices on a juror's judgment.

This enhanced effect can be explained by the same Bayesian principles described in Part II. Simply stated, pursuant to the shrinkage principle, the absence of evidence in a case causes a juror to rely more heavily on her assessment of the defendant's character than she would in a case that involves an abundance of evidence. In turn, her judgment will rely more heavily on her prior beliefs and prejudices, which have heavily influenced her character assessment. The verdict will therefore be more heavily influenced by race, appearance, economic status, and other background characteristics of the defendant.¹⁹²

Of course, the impact of character evidence on inequality is not limited to trials. A defendant may well receive a better or worse plea bargain based on the race or appearance of the defendant in a sexual-assault case. This is due to the defendant's and the prosecutor's knowledge that prior beliefs and prejudices will substantially permeate the jury's verdict if the case goes to trial. Moreover, this unequal treatment surely has a substantial impact on society more broadly, and it may even deepen jurors' prior prejudices in future cases.

consistent is the defendant's behavior across varying circumstances? Thus, the juror's prior beliefs may influence the juror's view regarding the consistency of the defendant's behavior. If a juror is presented with strong evidence that the defendant has committed a prior sexual assault, and this evidence is consistent with the juror's prior beliefs, the juror may reason that this is extremely strong evidence that the defendant committed the sexual assault in question. If, on the other hand, the evidence is inconsistent with the juror's prior beliefs, she may reason that it is not very probative of whether the defendant committed the act in question.

¹⁹² The literature has identified various ways in which an individual's race and economic status impact the individual's likelihood of being falsely arrested for, charged with, and convicted of sexual assault, and in which Rule 413 may disproportionately increase the likelihood of false convictions for certain groups of people based on race and other background characteristics. See Baker, *supra* note 64, at 592-97.

In summary, the admissibility of character evidence in sexual-assault cases promotes prejudice and inequality in these cases and beyond. Moreover, the problem is not one that can be superficially remedied: the very idea of character evidence is directly tied to the defendant's background characteristics.¹⁹³ If this is correct, the status quo is unacceptable with respect to the courts' implicit reliance on jurors' prior beliefs and prejudices and the resulting inequality based on race, appearance, and other background characteristics.

Finally, in no way do I intend to minimize the importance of concerns underlying a sexual-assault exception to the rule against character evidence. As discussed above, they are critical to supporting sexual-assault claims, actions, and prosecutions. My aim in the current Section is only to highlight the implications of the reflective model for prejudice and inequality in sexual-assault cases and other cases that involve formal exceptions to the rule against character evidence. The aims of fairness and equality, as well as those underlying Rule 413, are critical.¹⁹⁴

¹⁹³ One possibility is for a court to instruct the jury, via Rule 105, not to consider factors such as race, sex, and appearance. See Capers, *supra* note 148, at 898-900. Such an instruction may have limited effectiveness in certain circumstances. However, as discussed previously, there is good reason to expect that a juror's reliance on her prior beliefs is *inherent* in any use of character reasoning to arrive at a judgment. See *supra* Sections II.C, III.A. As such, a limiting instruction can only go so far, and in some circumstances, it may even deepen a juror's reliance on prior beliefs by highlighting this impermissible mode of reasoning to the juror. In any event, the effectiveness of limiting instructions in this regard is an important topic for future empirical research. See *infra* Section III.D.

¹⁹⁴ The primary focus of this Section is Rule 413's exception to the ban on character evidence. However, the reflective model of character evidence carries similar implications for other formal exceptions to the rule against character evidence. Another context in which an exception to the ban on character evidence may give rise to verdicts that are heavily influenced by a juror's prior beliefs and prejudices regarding a defendant's race, appearance, and other background characteristics is Rule 609 of the FRE, which provides for a broad exception to the rule against character evidence for a witness's prior convictions. FED. R. EVID. 609. For the reasons discussed in this Section and in Part II, how a juror uses such evidence to make inferences regarding a witness — and worse, to make inferences regarding a witness-*defendant* — may be heavily dependent on the juror's prior beliefs and prejudices. See Anna Roberts, *Reclaiming the Importance of the Defendant's Testimony: Prior Conviction Impeachment and the Fight Against Implicit Stereotyping*, 83 U. CHI. L. REV. 835, 835-40, 860-83 (2016) (highlighting the significant role of racial stereotyping in criminal trials, and arguing that courts should consider the importance of a defendant's testimony for negating such stereotypes through

C. *The Trend Toward Broad Admissibility for Character Evidence*

As discussed in Part I, the courts and legislatures have been increasingly permissive of character evidence. As one author has stated:

The American rule barring character evidence in criminal cases is degrading in every sense of the word. The rule's vitality has degraded as courts and legislatures expand existing exceptions and add new ones. The rule's coherence has degraded so badly that the justification for the rule and the tools for applying it are anemic in all but the clearest cases.¹⁹⁵

In addition to a growing range of formal exceptions to the rule against character evidence, courts regularly admit character evidence under a misinterpretation of Rule 404(b)(2) as providing for an exception to the rule against character evidence — that is, as permitting other-acts evidence that, although relevant to a Rule 404(b)(2) purpose such as intent or motive, indeed requires propensity reasoning.¹⁹⁶ Additionally, it is common for courts to admit character evidence without explanation by simply referring to precedent that has developed in a particular context.¹⁹⁷

As discussed *supra* Section I.B, the Advisory Committee recently considered amendments to Rule 404(b) that would clarify the nature of Rule 404(b)(2) as providing for non-character uses of other-acts evidence rather than for exceptions to the rule against character evidence; however, the Advisory Committee declined to amend the rule, explaining that such an amendment would “add substantial complexity, while ignoring that in some cases, a bad act is legitimately offered for a proper purpose but is nonetheless bound up with a propensity inference

“individuating information” in deciding whether to admit or exclude prior-conviction impeachment evidence); Julia Simon-Kerr, *Credibility by Proxy*, 85 GEO. WASH. L. REV. 152, 153-59 (2017) (arguing “that the real function of American impeachment jurisprudence has been to embed notions of status in the law of evidence” and highlighting the impact on “perpetuating systemic biases in the justice system”). See generally Anna Offit, *The Character of Jury Exclusion*, 106 MINN. L. REV. 2173 (2022) (highlighting the role of racial stereotyping in the jury-selection process and proposing the application of Rule 404(a) to limit juror exclusions in this process based on propensity inferences).

¹⁹⁵ Milich, *supra* note 6, at 776.

¹⁹⁶ See *supra* Section I.B.

¹⁹⁷ See *supra* Section I.B.

...”¹⁹⁸ Arguably, however, leaving this room for propensity inferences risks altogether devouring the rule, or at least permitting the current trend toward increasingly broad and common carveouts to continue.

Meanwhile, the trend toward a more permissive rule against character evidence results in verdicts that are heavily influenced by jurors’ prior beliefs and prejudices. In particular, when Rule 404(b)(2) evidence is misinterpreted as an exception to the ban on character evidence, the role of prior beliefs in these situations is similar to that described in the previous Section involving formal exceptions to Rule 404.¹⁹⁹ Assume, for example, that a defendant is on trial for armed robbery. He is alleged to have used a knife in the robbery, and to show that the defendant committed the crime in question, the court admitted evidence that the defendant has committed a prior robbery involving a knife. Assume that the court admitted this evidence under a loose interpretation of Rule 404(b)(2) to prove the “identity” of the defendant as the person who committed the act in question.²⁰⁰

Applying the reflective model, although a juror develops prior beliefs (and informs her propensity hyperpriors) from the start of a trial, these prior beliefs remain relatively uninfluential in the absence of character evidence. However, when the court admits evidence of the defendant’s prior robbery, the juror is likely to rely on prior beliefs through two sets of inferences. First, she may make explicit and implicit character inferences — reasoning that the defendant has committed this act in the past and is therefore likely now acting in accordance with a character to commit such acts. As above, character inferences in this type of situation are often unavoidable, and to the extent that jurors make them, pursuant

¹⁹⁸ Advisory Committee Report, *supra* note 80, at 5; *see* United States v. Thorne, No. 18-389, 2020 WL 122985, at *5 n.4 (D.D.C. Jan. 10, 2020).

¹⁹⁹ As in the sexual-assault context, a limiting instruction under Rule 105 is unlikely to alleviate the concern that I address in the current Section, which involves the courts’ frequent misapplication of Rule 404(b)(2) to permit propensity reasoning — not situations in which courts apply Rule 404(b)(2) to admit evidence for a non-propensity purpose while providing a limiting instruction to prevent impermissible propensity inferences. Moreover, as previously discussed, a limiting instruction regarding juror prior beliefs in particular is unlikely to be effective in mitigating a juror’s reliance on them. *See* discussion *supra* Section III.A; *supra* note 193.

²⁰⁰ FED. R. EVID. 404(b)(2).

to the reflective model, they are likely heavily influenced by the juror's propensity hyperpriors — her implicit biases.

Second, even if the juror restricts her character reasoning to evaluate the *identity* of the defendant, it still involves propensity inferences and is therefore likely to be heavily influenced by the juror's prior beliefs. The juror is asked to make an inference that the defendant used a knife in a previous robbery and is therefore more likely to be the culprit in the current crime, since a knife was used. If a juror holds prior beliefs that indicate a high probability that the defendant would commit armed robbery, the juror may well infer that the defendant used a knife in a past robbery; that he is therefore the kind of person that uses knives in robberies; and is therefore more likely to be the person who committed the crime in question — since it involved the use of a knife. If the juror holds prior beliefs that indicate a low probability that the defendant would commit armed robbery, the juror may well infer that although there is evidence that the defendant used a knife in a past robbery, even assuming that he committed this prior act, his decision to do this may have been dominated by circumstances rather than character, or he may have changed; that he is, for example, a complicated person and not inherently the type of person that would commit this type of act; and that the defendant's use of a knife in a past robbery does not provide substantial evidence that he committed the crime in question, even if it did involve the use of a knife. Therefore, although the propensity inferences in this scenario may be for a more limited purpose than in the sexual-assault exception discussed in the previous Section, they still invite jurors to rely heavily on their prior beliefs, and these can be very influential on the case's outcome.

Note that if Rule 404(b)(2) is interpreted properly to provide for non-propensity uses of other-acts evidence, this problem does not arise. This is because jurors are not asked to arrive at a character assessment. For example, if a court admits evidence of a prior burglary to prove the defendant's knowledge of how to bypass an alarm system — a non-propensity use of other-acts evidence — jurors are simply asked to reason that the defendant knew how to bypass the alarm system, as evidenced by his previous act, and therefore is more likely to have committed the act in question, which involved bypassing a similar alarm system. This reasoning does not require propensity inferences. Therefore, it does not activate the juror's propensity hyperpriors as

described in Part II. The juror may of course employ prior beliefs and prejudices to the extent that she does for other evidence; however, as argued above, these beliefs are likely relatively subdued when the juror is not asked to assess the defendant's character.²⁰¹

Now contrast this burglary scenario with the robbery scenario in which jurors are in fact asked to make propensity inferences regarding the defendant's use of a knife. In the robbery scenario, a juror must assess — via propensity reasoning — to what degree the defendant's previous use of a knife implies that he used one in the incident in question. This is an inference that requires the juror to assess the defendant's character or propensity — even assuming that the juror does not make any inferences beyond those sanctioned by the court. It invites the juror to rely heavily on her prior beliefs and prejudices regarding the defendant to fill substantial informational gaps and to arrive at a character assessment and make inferences for the act in question.

Finally, common law carveouts from Rule 404 have a similar effect. As described in Part I, these carveouts have frequently developed to permit character evidence for a particular purpose — similar to (and often originating from) misinterpretations of Rule 404(b)(2).²⁰² Although the other-acts evidence is offered for a particular purpose, such as to prove knowledge or intent in a drug case,²⁰³ the relevance of the evidence relies on propensity reasoning. Therefore, it similarly invites jurors to rely heavily on their implicit biases in forming character assessments and arriving at a verdict.

In summary, in addition to formal exceptions to Rule 404, the trend toward a more permissive treatment of character evidence via misinterpretations of Rule 404(b)(2) and via common law carveouts to the rule against character evidence paves a substantial channel for prior beliefs and prejudices to enter trials and influence verdicts. This effect and its implications for unfairness and inequality provide a strong reason

²⁰¹ Note that to the extent that jurors make *impermissible* character inferences even when other-acts evidence is admitted for non-character purposes (it is widely recognized that they do), the influence of prior beliefs and prejudices should at least be considered by the court (as a form of prejudice) in reviewing the evidence's admissibility under Rule 403. See FED. R. EVID. 403.

²⁰² See *supra* Section I.B.

²⁰³ See discussion *supra* Section I.B; *supra* note 19 and accompanying text.

(among others) to curb the sharp trend toward a more permissive approach to character evidence.

D. *Developing Sound Empirical Evidence*

There is a marked gap in the literature regarding how character evidence interacts with variables such as race, sex, economic status, and other background characteristics in influencing a verdict. While there is an abundance of empirical scholarship regarding the impact of certain background characteristics on case outcomes, these studies generally provide no information regarding the interaction between character evidence and the background characteristics studied, and there is almost no empirical research directly relating to this interaction in particular.²⁰⁴

The model introduced in Part II facilitates empirical research in this regard by describing a testable theory, or hypothesis, regarding how exactly character evidence might interact with a juror's prior beliefs and prejudices — and particularly in relation to race, sex, economic status, and other background variables — in influencing a verdict. In particular, the reflective model explains why there is good reason to expect that the introduction of character evidence causes a substantial increase in the influence of a juror's implicit biases on a judgment or verdict.

Based on this model, it is straightforward to formulate a basic experimental design for examining the effects implied by the model. For example, to empirically study the interaction between character evidence and implicit bias regarding the race of a defendant, a study may define two factors, Character Evidence and Race, each having two “levels” — “Character Evidence” and “No Character Evidence,” and “Black” and “White,” respectively.²⁰⁵ It may define a Verdict (“Guilty” or “Not Guilty”) as the outcome of interest. Then, it could employ a causal framework known as the “potential outcomes framework” (also known as the Rubin Causal Model) to examine the following effect: the difference between the impact of Race on the Verdict when Character

²⁰⁴ For one notable study regarding the interaction between race and character evidence, see Maeder & Hunt, *supra* note 161, at 610-18.

²⁰⁵ See Tirthankar Dasgupta, Natesh S. Pillai & Donald B. Rubin, *Causal Inference from 2^k Factorial Designs by Using Potential Outcomes*, 77 J. ROYAL STAT. SOC'Y SERIES B (STAT. METHODOLOGY) 727, 727-43 (2015).

Evidence is introduced versus when Character Evidence is not introduced.²⁰⁶

Thus, in Figure 3, the study could examine the following difference: (Cell 1 – Cell 3) – (Cell 2 – Cell 4), where each cell would take either the value Guilty or Not Guilty (known as “potential outcomes” in the potential outcomes framework).²⁰⁷ That is, it would examine the difference in potential outcomes for Race = Black versus Race = White when there is Character Evidence versus No Character Evidence.²⁰⁸

Figure 3

	Character Evidence	No Character Evidence
Black	Cell 1	Cell 2
White	Cell 3	Cell 4

A study designed to examine the reflective model and the interaction between character evidence and race could employ a framework known as the “potential outcomes framework” to compare the difference between the impact of *Race* on a *Verdict* when *Character Evidence* is introduced (Cell 1 – Cell 3) versus when *Character Evidence* is not introduced (Cell 2 – Cell 4).

For example, over many replications, if it is observed that the introduction of Character Evidence is associated with a substantial difference in the likelihood of Guilty (versus Not Guilty) when comparing Race = Black and Race = White, while the introduction of No Character Evidence is associated with a far less substantial difference when comparing Race = Black and Race = White, then this would provide support for the reflective model and the argument that character evidence causes an increased influence of implicit bias.²⁰⁹ Such findings would have significant implications for the broad range of legislative and judicial exceptions to the rule against character evidence discussed herein.

²⁰⁶ See GUIDO W. IMBENS & DONALD B. RUBIN, CAUSAL INFERENCE FOR STATISTICS, SOCIAL, AND BIOMEDICAL SCIENCES: AN INTRODUCTION 3-30 (2015); Dasgupta et al., *supra* note 205, at 727-43; Donald B. Rubin, *Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies*, 66 J. EDUC. PSYCH. 688, 688-701 (1974).

²⁰⁷ See IMBENS & RUBIN, *supra* note 206, at 3-30.

²⁰⁸ See *id.*; see also Maeder & Hunt, *supra* note 161, at 613-18.

²⁰⁹ See IMBENS & RUBIN, *supra* note 206, at 3-30; Dasgupta et al., *supra* note 205, at 727-32.

CONCLUSION

Notwithstanding the rule against character evidence, courts regularly admit character evidence through an expanding set of exceptions.²¹⁰ These exceptions arise for various reasons, including evidentiary gaps, circumstances in which character evidence seems particularly probative, and non-evidentiary policies.²¹¹ In creating exceptions, legislatures and courts implicitly or explicitly rely on a common narrative regarding the risks associated with the admission of character evidence. In particular, they create exceptions for types of character evidence that seem particularly probative relative to the unfair prejudice described by this narrative, or for circumstances in which courts seem to be able to mitigate such unfair prejudice through limiting instructions and other counteracting measures.

However, this narrative fails to account for a highly pernicious effect of character evidence, and this effect is intrinsic to character-propensity reasoning and is unlikely to be mitigated through counteracting measures. According to the reflective model, regardless of the concerns that have led to an exception — some of which are of critical importance — character evidence inherently involves inferences that are substantially based on a juror's prior beliefs and prejudices regarding the background characteristics of a defendant. Consequently, verdicts that are based on character evidence are inherently biased against certain groups of people, and the admissibility of character evidence in many contexts may therefore be a significant driver of inequality and unfairness in the U.S. legal system. If this model is correct, it provides a strong reason to limit legislative and judicial exceptions to the rule against character evidence and to curb the current trend toward a more permissive rule.

²¹⁰ See *supra* Part I.

²¹¹ See *supra* Parts I, III.