BIAS, THE BRAIN, AND STUDENT EVALUATIONS OF TEACHING

DEBORAH J. MERRITT†

The complaints are never-ending, voluminous, and contradictory. I talk too loud or not loud enough. I walk too close to people and make them nervous. If I look at students, they are nervous. If I do not look at them they are angry. If I call on them, I am picking on them. If I do not call on them, I have a personal vendetta against them . . . .

When I talk to students in an attempt to ascertain what I do that is so different from the other professors teaching the same section of first-year students, they admit that I do no more in class than their white male professors—my class is no more rigorous, no more intimidating, no more work. In fact, they seem to like the class. . . . Most students appear to like the use of overheads, the introductory and periodic summaries, and question and answer periods . . . .

The only difference appears to be that I am a Black female . . . .[1]

INTRODUCTION

Professors of color have published poignant accounts of harshly negative student evaluations.[2] As the few empirical

† John Deaver Drinko/Baker & Hostetler Chair in Law, Moritz College of Law, The Ohio State University. Years of conversations with colleagues have helped shape the ideas reflected in this Article. For more immediate assistance, I am indebted to Ruth Colker, Andrew Merritt, Daniel Merritt, and Elaine Shoben. Kristin Harlow provided invaluable research assistance.

[1] Pamela J. Smith, *Teaching the Retrenchment Generation: When Sapphire Meets Socrates at the Intersection of Race, Gender, and Authority*, 6 WM. & MARY J. WOMEN & L. 53, 162–63 (1999).

[2] For descriptions in the legal literature, see, for example, Okianer Christian Dark, *Just My 'Magination*, 10 HARV. BLACKLETTER L.J. 21, 21–28 (1993); Richard Delgado & Derrick Bell, *Minority Law Professors' Lives: The Bell-Delgado Survey*, 24 HARV. C.R.-C.L. L. REV. 349, 349–54, 359–61 (1989), which reports results of a survey mailed to all minority law faculty; Trina Grillo, *Tenure and Minority Women Law Professors: Separating the Strands*, 31 U.S.F. L. REV. 747, 752–54 (1997); Joyce

studies examining instructor race and student ratings have confirmed, minority faculty receive significantly lower evaluations than their white colleagues.[3] Students' contradictory and often hostile comments on evaluations of minority faculty, as well as their occasional direct references to gender or race, raise troubling questions about the role of bias in these assessments.[4] White faculty members have also noted the possibility of bias in their student evaluations, particularly based on gender, appearance, or political ideology.[5] Throughout the academy,

---

Hughes, *Different Strokes: The Challenges Facing Black Women Law Professors in Selecting Teaching Methods*, 16 NAT'L BLACK L.J. 27, 29 (1998); Reginald Leamon Robinson, *Teaching from the Margins: Race as a Pedagogicial Sub-Text*, 19 W. NEW ENG. L. REV. 151, 151–52, 168–72 (1997); Smith, *supra* note 1, at 167–69, 175–91; Donna E. Young, *Two Steps Removed: The Paradox of Diversity Discourse for Women of Color in Law Teaching*, 11 BERKELEY WOMEN'S L.J. 270, 279–80 (1996); and Vincene Verdun, *The Ugly Truth: Was the Outburst Anything but Racism?*, 3 BUS. L. TODAY, May/June 1994, at 18, 18. *See also* Emma Coleman Jordan, *Images of Black Women in the Legal Academy: An Introduction*, 6 BERKELEY WOMEN'S L.J. 1, 4–5 (1990–91) (describing the controversy over Harvard Law School's lack of a single tenured minority female professor in the spring of 1990, and noting that students made "off-the-cuff" remarks to reporters regarding their visiting tort professor, Regina Austin, a woman of color); *cf.* Mari J. Matsuda, *Voices of America: Accent, Antidiscrimination Law, and a Jurisprudence for the Last Reconstruction*, 100 YALE L.J. 1329, 1332, 1352–54 (1991) (noting the tension that exists when discrimination claims are based on negative employment decisions related to foreign-born employees' accents and describing students' negative evaluation comments regarding their Asian American professors' accents).

[3] Until recently, no empirical work probed the relationship between race and student evaluations, a scholarly gap that is itself troubling. Published data on this subject from law schools remain unavailable. A recent analysis of almost 17,000 evaluations completed by undergraduate and graduate students at the University of Texas, however, reveals that minority faculty obtain significantly lower ratings than white professors, even after controlling for tenure status and course type. *See* Daniel S. Hamermesh & Amy Parker, *Beauty in the Classroom: Instructors' Pulchritude and Putative Pedagogical Productivity*, 24 ECON. EDUC. REV. 369, 373 (2005). Studies have not yet attempted to isolate the reasons for this difference.

[4] *See, e.g.*, Delgado & Bell, *supra* note 2, at 361 (stating that minority faculty reported student evaluations that "are sometimes both positive and negative for a single course"); Smith, *supra* note 1, at 167 ("The intensity of their anger and hatred was frigh-tening. Many of them attached notes to their evaluations, espous-ing crazy racial and/or sexist stereotypes."); Kathryn Pourmand Nordick, Essay, *A Critical Look at Student Resistance to Non-Traditional Law School Professors*, 27 W. NEW ENG. L. REV. 173, 188, 191 (2005) (observing that law school classmates' criticisms of black professors "were often harsh and condescending, while the criticisms of traditional professors were almost backward compliments" and that a recent review of randomly selected student evaluations yielded "statements that clearly indicated bias").

[5] *See, e.g.*, Richard L. Abel, *Evaluating Evaluations: How Should Law Schools Judge Teaching?*, 40 J. LEGAL EDUC. 407, 437–45 (1990); Kathleen S. Bean, *The*

faculty question whether student evaluations of teaching accurately reflect a professor's success in helping students learn. Many charge that evaluations actually undermine learning by encouraging lenient grading and superficial classroom presentations.[6] In an increasingly diverse and competitive workplace, can we rely upon conventional teaching evaluations to tell us what we want to know about a professor's classroom success? Or do these evaluations reflect—and perhaps reinforce—biases based on race, sex, and other unwelcome characteristics?

Despite the persistence of these questions, law schools and other academic departments continue to use traditional student evaluations of teaching. Indeed, many professors report growing reliance on these measures—particularly on isolated numerical averages—in tenure, promotion, salary, and other decisions.[7] The academy has been particularly silent in response to questions about racial bias in conventional teaching evaluations: Few articles engage the eloquent critiques that individual

Gender Gap in the Law School Classroom—Beyond Survival, 14 VT. L. REV. 23, 29 (1989); Martha Chamallas, *The Shadow of Professor Kingsfield: Contemporary Dilemmas Facing Women Law Professors*, 11 WM. & MARY J. WOMEN & L. 195, 195–208 (2005); Christine Haight Farley, *Confronting Expectations: Women in the Legal Academy*, 8 YALE J.L. & FEMINISM 333, 336–40 (1996); April Kelly-Woessner & Matthew C. Woessner, *My Professor Is a Partisan Hack: How Perceptions of a Professor's Political Views Affect Student Course Evaluations*, 39 PS: POL. SCI. & POL. 495, 495–500 (2006); Deborah Maranville, *Classroom Incivilities, Gender, Authenticity and Orthodoxy, and the Limits of Hard Work: Four Lenses for Interpreting a "Failed" Teaching Experience*, 12 WM. & MARY J. WOMEN & L. 699, 716–23 (2006).

[6] *See, e.g.*, VALEN E. JOHNSON, GRADE INFLATION: A CRISIS IN COLLEGE EDUCATION 235–37 (2003); Dennis E. Clayson & Mary Jane Sheffet, *Personality and the Student Evaluation of Teaching*, 28 J. MARKETING EDUC. 149, 149, 157–58 (2006); Charles R. Emery et al., *Return to Academic Standards: A Critique of Student Evaluations of Teaching Effectiveness*, 11 QUALITY ASSURANCE EDUC. 37, 37–45 (2003); Wendy M. Williams & Stephen J. Ceci, *"How'm I doing?,"* 29 CHANGE 13, 13–14 (1997).

[7] *See, e.g.*, Judith D. Fischer, *The Use and Effects of Student Ratings in Legal Writing Courses: A Plea for Holistic Evaluation of Teaching*, 10 J. LEGAL WRITING INST. 111, 111–12 (2004); Hugh Hinton, *Reliability and Validity of Student Evaluations: Testing Models Versus Survey Research Models*, 26 PS: POL. SCI. & POL. 562, 562–63 (1993); Richard S. Markovits, *The Professional Assessment of Legal Academics: On the Shift from Evaluator Judgment to Market Evaluations*, 48 J. LEGAL EDUC. 417, 417 (1998).

minority professors have raised,[8] and schools do not seem to have examined their practices in response to these concerns.[9]

It is time to take seriously the criticisms that scholars have voiced about student evaluations of teaching. Extensive research by psychologists and educators convincingly demonstrates that these evaluations *are* biased. The biases, however, are not simplistic ones based directly on race, gender, or other social categories. Professors do not suffer automatic, consistent penalties premised on race, gender, political ideology, or other commonly recognized categories. Indeed, some "nontraditional" professors obtain very positive teaching evaluations, accurately reflecting their teaching excellence, while some politically moderate, conventional white males receive ratings that seem unduly low.

Bias in student evaluations derives primarily from a relationship that most faculty overlook: the powerful link between student ratings and a small set of nonverbal behaviors.[10]

---

[8] *But see* Nordick, *supra* note 4, at 179–87 (reviewing minority law professors' narratives regarding their experiences with students' hostile criticism).

[9] *See, e.g.*, Therese A. Huston, *Race and Gender Bias in Higher Education: Could Faculty Course Evaluations Impede Further Progress Toward Parity?*, 4 SEATTLE J. SOC. JUST. 591, 591–92, 597–601 (2006) (discussing research showing that student evaluations reveal biases against women and minority faculty and contending that "most faculty and administrators are unaware of the bias in students' course evaluations of teaching"); Smith, *supra* note 1, at 93–96 (stating that although "racism is alive and well as we enter the twenty-first century, . . . [w]hite people and institutions deny the continued existence of racism and its effects of the ability of African-Americans to survive and excel in academia"); Kathryn L. Vaughns, *Women of Color in Law Teaching: Shared Identities, Different Experiences*, 53 J. LEGAL EDUC. 496, 500 (2003) ("One aspect of the law school environment that has especially bothered me is a reluctance to acknowledge that my experiences in the classroom, and those of other people of color, may well be different—sometimes vastly so—from those of my white peers.").

[10] Although this Article focuses on the substantial biases stemming from unconscious reactions to nonverbal behavior, other considerations can also distort law school evaluations. Some studies, for example, suggest that students award lower ratings to faculty if they receive grades before completing evaluations. *See, e.g.*, Dennis E. Clayson et al., *Grades and the Student Evaluation of Instruction: A Test of the Reciprocity Effect*, 5 ACAD. MGMT. LEARNING & EDUC. 52, 55–61 (2006). This dynamic can affect evaluations in courses like legal writing, in which students receive grades throughout the semester. *See, e.g.*, Judith D. Fischer, *How to Improve Student Ratings in Legal Writing Courses: Views from the Trenches*, 34 U. BALT. L. REV. 199, 199–202 (2004); Melissa Marlow-Shafer, *Student Evaluation of Teacher Performance and the "Legal Writing Pathology": Diagnosis Confirmed*, 5 N.Y. CITY L. REV. 115, 115–16 (2002) ("Many legal writing teachers claim that their evaluations contain degrading comments and are lower than doctrinal law professors."). Although they deserve serious attention, further discussion of these problems is

Conventional student evaluations are strongly influenced by a professor's smiles, gestures, and other mannerisms, rather than the professor's knowledge, clarity, organization, or other qualities more clearly associated with good teaching. The way in which a professor walks into the room or smiles at the class can affect student ratings much more substantially than what the professor says or writes on the blackboard.[11] In fact, students' ratings of professors show little, if any, correlation with objective measures of what students learn.[12] Evaluations collected from students after no more than five minutes exposure to a professor accurately predict assessments gathered at semester's end, leaving little doubt of the superficiality of student evaluations.[13]

The nonverbal mannerisms that drive teaching evaluations bear little relation to learning. Many of the nonverbal behaviors that influence teaching evaluations are related to race, gender, and other immutable characteristics; they stem from physiology, culture, and habit.[14] Social stereotypes filter perceptions of these behaviors so that even when faculty engage in identical classroom behaviors, students may perceive those behaviors differently depending on the professor's race, gender, and other characteristics.[15] Women and minority faculty, therefore, may experience bias on at least two levels. At the same time, white men can also suffer unfairly negative evaluations if their facial expressions or mannerisms trigger negative reactions.

These biases do not arise because students are incapable of evaluating teaching. Under the right circumstances, the most experienced decision makers will manifest the same biases.[16] The inaccuracies in our current system occur because of the *manner* in which we gather student feedback. Psychology research demonstrates that the human mind functions along two

beyond the scope of this Article.

[11] *See infra* Part I.

[12] *See infra* Part II.C.

[13] *See* Clayson & Sheffet, *supra* note 6, at 154; *infra* notes 69–75 and accompanying text (discussing in detail the psychological research finding a strong correlation between end-of-semester student evaluations and students' evaluations formulated after brief observations of the same professors).

[14] *See infra* Part II.A.

[15] *See infra* Part II.B.

[16] *See, e.g.*, Irene V. Blair et al., *The Influence of Afrocentric Facial Features in Criminal Sentencing*, 15 PSYCHOL. SCI. 674, 677–78 (2004) (discussing influences on sentencing by trial judges); *see also infra* notes 98–100 and accompanying text (discussing this research).

very different tracks, one that generates automatic, instinctive reactions and another that produces more reflective, deliberative decisions.[17] The way that we currently obtain teaching assessments from students taps their instinctive rather than reflective judgments. Law schools can mitigate the biases in student evaluations by designing evaluation systems that allow students to offer more thoughtful assessments of teaching. The most effective processes would give students additional information about a professor's pedagogic strategies and then engage them in facilitated small-group discussions of teaching.

The first part of this Article reviews the psychology research demonstrating the strong link between student evaluations and a professor's nonverbal behavior. The second part examines why this connection is so damaging: It allows race, gender, and other immutable characteristics to bias our assessment of good teaching, while doing little to identify faculty who genuinely enhance student learning. A third section probes the cognitive processes that produce student evaluations, explaining why the evaluation process we use—rather than the students who participate—generates these unreliable results. The final section proposes a system of gathering student feedback, one that makes greater use of facilitated small group discussions, to overcome many of the cognitive biases built into the current system.

Fairness to both students and faculty demands that we look critically at the student evaluation system we currently employ. Understanding the flaws in our teaching evaluation process, moreover, illuminates deeper truths about how the brain works. The legal system depends upon judgments that people make of others, from hiring decisions and due diligence reviews to negotiating strategies and jury deliberations. Appreciating the cognitive channels that distort those assessments will help lawyers improve decision making in all aspects of their practice. This Article begins that process by exploring cognitive paths in judgments that are familiar to all students, faculty, and practicing lawyers: student evaluations of teaching.

[17] Steven A. Sloman, *Two Systems of Reasoning*, *in* HEURISTICS AND BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT 379, 379 (Thomas Gilovich et al. eds., 2002); *see also infra* Part III.

I. NONVERBAL BEHAVIOR AND STUDENT EVALUATIONS

Nonverbal behaviors include a wide range of appearances and actions that influence communication apart from verbal content.[18] Smiles, frowns, raised eyebrows, and other facial expressions are nonverbal behaviors. So are shrugs, waves, and other gestures. The ways in which we move, position our bodies, and hold our arms further enrich the nonverbal repertoire. Dress, hairstyle, and other aspects of physical demeanor also contribute to nonverbal communication. So do many non-substantive aspects of speech, such as voice tone, accent, and cadence.

Humans respond instinctively and rapidly to these nonverbal cues.[19] Nonverbal behavior shapes employment interviews, coaching sessions, and other interactions in the workplace.[20] Expressions, gestures, appearances, and vocal tones influence every aspect of a judicial trial, from jury selection to sentencing.[21] Even a simple smile can communicate a wide variety of meanings, from genuine pleasure to discomfort or deception.[22] We rely constantly on nonverbal signals to detect the emotions, attitudes, and intentions of people around us.

Despite its cerebral connections, the law school classroom hums with nonverbal behavior. A quick scan of the room will tell the professor which students are paying attention and which ones are surfing the web or sending instant messages behind their laptop screens. Even in a large classroom, professors can often sense which students are engaged with the material and

---

[18] *See generally* JOSEPH A. DEVITO, THE INTERPERSONAL COMMUNICATION BOOK 162–92 (11th ed. 2007).

[19] *See* Y. Susan Choi et al., *The Glimpsed World: Unintended Communication and Unintended Perception*, *in* THE NEW UNCONSCIOUS 309, 309–10 (Ran R. Hassin et al. eds., 2005); Bella M. DePaulo, *Nonverbal Behavior and Self-Presentation*, 111 PSYCHOL. BULL. 203, 207 (1992).

[20] *See, e.g.*, *Business Applications of Nonverbal Communication*, *in* APPLICATIONS OF NONVERBAL COMMUNICATION 119, 119 (Ronald E. Riggio & Robert S. Feldman eds., 2003) (providing a summary of recent research).

[21] *See, e.g.*, Michael Searcy et al., *Communication in the Courtroom and the "Appearance" of Justice*, *in* APPLICATIONS OF NONVERBAL COMMUNICATION, *supra* note 20, at 41, 41.

[22] *See* Paul Ekman et al., *Smiles When Lying*, 54 J. PERSONALITY & SOC. PSYCHOL. 414, 418–20 (1988); Christine R. Harris & Nancy Alvarado, *Facial Expressions, Smile Types, and Self-Report During Humour, Tickle, and Pain*, 19 COGNITION & EMOTION 655, 665 (2005); Julie A. Woodzicka & Marianne LaFrance, *Working on a Smile: Responding to Sexual Provocation in the Workplace*, *in* APPLICATIONS OF NONVERBAL COMMUNICATION, *supra* note 20, at 139, 139.

which ones are bored, hostile, or confused. Just as much nonverbal communication flows in the opposite direction. From the moment a faculty member walks into the room, students perceive, process, and react to the professor's nonverbal signals.

Researchers have extensively documented the effect of these signals on student evaluations, often contrasting the dominance of classroom "style" over content. One early and well-known investigation into these classroom dynamics used a charismatic, distinguished-looking, and mellifluous actor to play the role of a scholar named "Dr. Fox."[23] The experimenters created a meaningless lecture on "Mathematical Game Theory as Applied to Physician Education," and coached Fox to deliver it "with an excessive use of double talk, neologisms, non sequiturs, and contradictory statements."[24] At the same time, the researchers encouraged Fox to adopt a lively demeanor, convey warmth toward his audience, and intersperse his nonsensical comments with humor. "In short," as one of the investigators summarized, Dr. Fox "gave a very enjoyable lecture in which he offered little or nothing of substance."[25]

Fox fooled not just one, but three separate audiences of professional and graduate students.[26] Despite the emptiness of his lecture, fifty-five psychiatrists, psychologists, educators, graduate students, and other professionals produced evaluations of Dr. Fox that were overwhelmingly positive.[27] In addition to awarding him strong numerical scores, audience members praised him for an "[e]xcellent presentation," "warm manner," "[g]ood flow," "[l]ively examples," "relaxed manner," and "[g]ood analysis of subject."[28]

Fox's use of warm, enthusiastic, and lively nonverbal behaviors would have been admirable if it had complemented a substantive presentation. Most faculty use stylistic elements to engage student interest and motivate learning. The disturbing feature of the Dr. Fox study, as the experimenters noted, is that

---

[23] *See generally* Donald H. Naftulin et al., *The Doctor Fox Lecture: A Paradigm of Educational Seduction*, 48 J. MED. EDUC. 630 (1973).

[24] *Id.* at 631.

[25] John E. Ware, Jr. & Reed G. Williams, *The Dr. Fox Effect: A Study of Lecturer Effectiveness and Ratings of Instruction*, 50 J. MED. EDUC. 149, 150 (1975).

[26] *See* Naftulin et al., *supra* note 23, at 631–33. The first group heard Fox's presentation live, while the other two saw a videotape. *Id.* at 632–33.

[27] *See id.*

[28] *Id.*

Fox's nonverbal behaviors so completely masked a meaningless, jargon-filled, and confused presentation. If style can trump substance so easily, even in the minds of a trained, professional audience, then what role do nonverbal behaviors play in more routine student evaluations?

Several researchers followed up on this question by using the Dr. Fox paradigm to conduct controlled classroom experiments.[29] These studies used videos that systematically varied a lecturer's content and nonverbal behaviors to examine their relative effect on student teaching evaluations.[30] A meta-analysis of this cluster of investigations concluded that nonverbal behaviors dramatically affected evaluations. For example, an entertaining style increased an instructor's ratings by about 1.2 points on a five point scale.[31] Lecturers who provided more content on the other hand received "inconsistent and generally much smaller" boosts in their evaluations.[32]

Other studies have isolated some of the specific nonverbal behaviors that generate positive student ratings. Based on a detailed analysis of university classes and student evaluations, Harry Murray determined that a professor's speech patterns, facial expressions, and humor had the greatest impact on student evaluations.[33] More learning-focused behaviors, such as giving

---

[29] *See, e.g.*, Herbert W. Marsh & John E. Ware, Jr., *Effects of Expressiveness, Content Coverage, and Incentive on Multidimensional Student Rating Scales: New Interpretations of the Dr. Fox Effect*, 74 J. EDUC. PSYCHOL. 126, 126–27 (1982) (reviewing earlier studies finding that students could be fooled into giving favorable evaluations of teachers when lectures are delivered in an enthusiastic and expressive manner); Ware, Jr. and Williams, *supra* note 25, at 151; Reed G. Williams & John E. Ware, Jr., *An Extended Visit with Dr. Fox: Validity of Student Satisfaction with Instruction Ratings After Repeated Exposures to a Lecturer*, 14 AM. EDUC. RES. J. 449, 449–50 (1977).

[30] *See, e.g.*, Marsh & Ware, Jr., *supra* note 29, at 126. One set of studies, for example, used the same actor who had portrayed Dr. Fox to create six video lectures on the biochemistry of memory. Ware, Jr. and Williams, *supra* note 25, at 151. Two of the lectures were designedly "low content" and contained only four substantive points, two were "medium content" and included fourteen points, and two "high content" lectures conveyed twenty-six points. *Id.* Within each pair, one lecture employed a highly charismatic style while the other avoided engaging mannerisms. *Id.* at 151–52. The researchers then randomly assigned undergraduates to view the videos, gathering teaching evaluations of the lecturer afterwards. *Id.* at 151.

[31] *See* Philip C. Abrami et al., *Educational Seduction*, 52 REV. EDUC. RES. 446, 455 (1982).

[32] *Id.* at 452.

[33] *See* Harry G. Murray, *Classroom Teaching Behaviors Related to College Teaching Effectiveness*, *in* USING RESEARCH TO IMPROVE TEACHING 21, 26 (Janet G.

"concrete examples of concepts," "point[ing] out practical applications," "repeat[ing] difficult ideas," or "providing sample exam questions" correlated less with student ratings.[34] While the Fox studies suggested that faculty could reap greater evaluation rewards by focusing on style rather than substance, Murray's investigation sounded a further disturbing note: Even when concentrating on the stylistic elements of their teaching, faculty can more effectively raise student evaluations by using certain facial expressions than by offering concrete examples or repeating difficult concepts.

A recent case study, centered on the eminent psychologist Stephen Ceci,[35] further illustrates the substantial connection between a professor's nonverbal behaviors and student evaluations of teaching. After participating in a short "teaching skills" workshop conducted by a media consultant, Ceci raised his evaluations in an introductory psychology course from an overall score of 3.08 out of five to 3.92.[36] Because he wanted to test the impact of the media training, Ceci carefully used the same syllabus, lecture content, audiovisual materials, assignments, and exams in sections of the course taught immediately before and after the training.[37] He altered only his vocal "pitch variability" and the extent of his hand gestures between the two versions of the course.[38] Notably, these small stylistic changes dramatically improved Ceci's score on *every* aspect of the college's

Donald & Arthur M. Sullivan eds., 1985) [hereinafter Murray, *Classroom Teaching*]; *see also* Harry G. Murray, *Effective Teaching Behaviors in the College Classroom*, *in* 7 HIGHER EDUCATION: HANDBOOK OF THEORY AND RESEARCH 135, 148–50 (John C. Smart ed., 1991) [hereinafter Murray, *Effective Teaching*].

[34] *See* Murray, *Classroom Teaching*, *supra* note 33, at 25.

[35] Ceci is the Helen L. Carr Professor of Developmental Psychology at Cornell University. He has published extensively on the accuracy of children's courtroom testimony, as well as other subjects, and recently won the American Psychological Society's James McKeen Cattell Award for lifetime contributions to an area of critical social importance. *See* Association for Psychological Science: 2004 James McKeen Cattell Fellow Award, http://www.psychologicalscience.org/awards/cattell/citations/ceci.cfm (last visited Oct. 19, 2007); Cornell University, College of Human Ecology: Bio Page for Stephen Ceci, http://www.human.cornell.edu/che/bio.cfm?netid=sjc9 (last visited Oct. 19, 2007).

[36] *See* Williams & Ceci, *supra* note 6, at 16, 20.

[37] *See id.* at 16–17. Ceci and his co-author took elaborate measures to assure similarities between the semesters. Independent raters, for example, viewed videotaped lectures from both semesters and confirmed that their content was identical. *Id.* at 18. Ceci's experience with the course, which he had taught for almost twenty years, made the controls feasible. *Id.* at 16–17.

[38] *Id.* at 15.

evaluation form, including items like instructor knowledge, organization, accessibility, textbook quality, fairness in grading, and other qualities unrelated to vocal pitch or gestures.[39]

Even a mere description of a professor's manner can affect students' evaluations. In one controlled experiment, a guest lecturer appeared before a large undergraduate class on a day when the regular professor was absent.[40] Students received written notes from their regular professor describing the guest as an experienced professor from another university who others considered "industrious, critical, practical, and determined."[41] Half of the notes further described the visitor as "a rather warm person," while the other half identified him as "a rather cold person."[42] Nonverbal behaviors often signal a speaker's apparent warmth or coldness to an audience. Here, the written descriptions primed the students to view the lecturer's manner through one of two contrasting lenses.[43]

After a forty minute informative lecture related to their course material, students evaluated the guest lecturer.[44] Those who had been told that he was "rather warm" rated him as significantly "more intelligent, more interesting, more considerate of the class, and more knowledgeable of his material" than did students who had read that he was "rather cold."[45] Changing just one word in the lecturer's written biography was enough to shift student perceptions of his personality, knowledge, and teaching effectiveness.[46]

---

[39] *See id.* at 19–20. Ceci's average rating in the category regarding instructor knowledge improved from 3.61 to 4.05: from 3.18 to 4.09 in regards to students' perceptions of his level of organization; from 2.99 to 4.06 for his accessibility; from 2.06 to 2.98 for textbook quality; and from 3.03 to 3.72 for fairness. *Id.* at 20. All of these shifts, as well as others on the evaluation form, were statistically significant at $p < .0001$. *Id.* at 20–21.

[40] *See* W. Neil Widmeyer & John W. Loy, *When You're Hot, You're Hot! Warm-Cold Effects in First Impressions of Persons and Teaching Effectiveness*, 80 J. EDUC. PSYCHOL. 118, 119 (1988).

[41] *Id.*

[42] *Id.*

[43] *Id.* at 118–19.

[44] *Id.* at 119.

[45] *Id.* Several other measures of personality and teaching effectiveness, including modesty, self assurance, and organization, did not differ significantly between the warm and cold conditions. *See id.* at 120. Widmeyer and Loy suggest that this confirms that the students perceived a relationship between warmth and teaching abilities, rather than a general "halo effect"—a frame of reference influencing others' total perception of an individual. *Id.*

[46] A large number of studies further explore the relationship among instructor

The most remarkable evidence of the link between a professor's nonverbal behavior and student evaluations of teaching, however, comes from recent research into "thin slice" judgments. Harvard psychologists Nalini Ambady and Robert Rosenthal coined the phrase "thin slices" in 1992 to describe brief observations of an individual that generate judgments about that individual's personality, intentions, and other characteristics.[47]

---

personality, nonverbal behaviors, and teaching evaluations. These investigations confirm that an instructor's personality exerts a substantial influence on evaluations, largely through his or her nonverbal behaviors. One pair of researchers, for example, recently concluded that the impact of personality is so large that evaluations "could most accurately be called a 'likeability' scale." Dennis E. Clayson & Debra A. Haley, *Student Evaluations in Marketing: What Is Actually Being Measured?*, 12 J. MARKETING EDUC. 9, 12–13 (1990); *see also* Stephen Erdle et al., *Personality, Classroom Behavior, and Student Ratings of College Teaching Effectiveness: A Path Analysis*, 77 J. EDUC. PSYCHOL. 394, 404–05 (1985) (finding in part that college instructors received higher ratings when they exhibited charismatic classroom behaviors such as the use of humor or encouraging student participation); Kenneth A. Feldman, *The Perceived Instructional Effectiveness of College Teachers as Related to Their Personality and Attitudinal Characteristics: A Review and Synthesis*, 24 RES. HIGHER EDUC. 139, 144 (1986) (finding statistically significant average correlations between students' and colleagues' perceptions of an instructor's personality traits and that instructor's ratings on student evaluations); James C. McCroskey et al., *Toward a General Model of Instructional Communication*, 52 COMM. Q. 197, 206–08 (2004) (finding that "teacher temperament is manifested in teacher communication behaviors which are observable by students" and that these behaviors were "associated with students' perceptions of their teachers' source credibility"); Harry Murray et al., *Teacher Personality Traits and Student Instructional Ratings in Six Types of University Courses*, 82 J. EDUC. PSYCHOL. 250, 259 (1990) ("[F]or any given type of course or for all types combined, student instructional ratings were strongly related to peer ratings of instructor personality traits."); Sally A. Radmacher & David J. Martin, *Identifying Significant Predictors of Student Evaluations of Faculty Through Hierarchical Regression Analysis*, 135 J. PSYCHOL. 259, 265–66 (2001) ("This robust relationship between instructor extraversion and students' perceptions of teaching effectiveness could be interpreted to support the fear of some faculty that student evaluations are just personality contests and may not be valid measures of teaching effectiveness."); Barbara R. Sherman & Robert T. Blackburn, *Personal Characteristics and Teaching Effectiveness of College Faculty*, 67 J. EDUC. PSYCHOL. 124, 130 (1975) ("[T]he evidence leans toward the importance of the personal characteristics as the cause of the perceived instructional effectiveness."); Marie Waters et al., *High and Low Faculty Evaluations: Descriptions by Students*, 15 TEACHING PSYCHOL. 203, 203–04 (1988) (discussing how students asked to describe teachers whom they had given high ratings on evaluations remembered those teachers for their positive personality traits "such as enthusiasm, personality, sense of humor, and enjoyment of teaching").

[47] *See* Nalini Ambady et al., *Toward a Histology of Social Behavior: Judgmental Accuracy from Thin Slices of the Behavioral Stream*, *in* 32 ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY 201, 203–04 (2000) (defining "thin slices" as "a brief excerpt of expressive behavior sampled from the behavioral stream," where a

Through a meta-analysis of forty-four studies[48] in which subjects observed no more than five minutes of a target's behavior, Ambady and Rosenthal concluded that these very quick observations "provide a great deal of information" and can trigger detailed predictions about another person's behavior.[49] Humans, in other words, make social judgments based on very short observations of other people.

Ambady and Rosenthal applied this insight to a detailed exploration of the assessments students offer on teaching evaluations.[50] They obtained videotapes of thirteen different instructors teaching undergraduate courses at Harvard University.[51] The courses spanned the humanities, social sciences, and natural sciences, and the instructors' ratings on end-of-semester evaluations were similarly diverse.[52] For each instructor, Ambady and Rosenthal abstracted thirty seconds of videotape from a single class session: ten seconds from the first ten minutes of class, ten seconds from the middle of the class, and ten seconds from the last ten minutes. They played these tapes, without sound, to groups of undergraduates who had never met the featured teacher.[53]

These students rated the instructors on fifteen different qualities, including competence, confidence, professionalism, enthusiasm, optimism, and warmth.[54] An initial group of raters viewed the entire thirty seconds of silent video for each instructor. Subsequent groups viewed redacted versions of just

brief excerpt is "any excerpt with dynamic information less than 5 min [sic] long"); Nalini Ambady & Robert Rosenthal, *Thin Slices of Expressive Behavior as Predictors of Interpersonal Consequences: A Meta-Analysis*, 111 PSYCHOL. BULL. 256, 256–57 (1992).

[48] *See* Ambady & Rosenthal, *supra* note 47, at 260 (describing their many studies regarding judgments based on observations of "thin-slices" of others' behavior in various contexts).

[49] *Id.* at 267.

[50] *See* Nalini Ambady & Robert Rosenthal, *Half a Minute: Predicting Teacher Evaluations from Thin Slices of Nonverbal Behavior and Physical Attractiveness*, 64 J. PERSONALITY & SOC. PSYCHOL. 431, 432 (1993).

[51] *Id.* at 431, 433.

[52] *Id.* at 433. The teachers in Ambady and Rosenthal's study were all graduate teaching assistants. *Id.* at 433. However, their work has been replicated with full-time faculty. *See infra* notes 63–68 and accompanying text.

[53] Ambady & Rosenthal, *supra* note 50, at 433. All clips focused on the teacher alone, without showing students. *Id.*

[54] *Id.* The full list of qualities was the following: "accepting, active, [not] anxious . . . , attentive, competent, confident, dominant, empathetic, enthusiastic, honest, likable, optimistic, professional, supportive, and warm." *Id.* at 433–34.

fifteen or six seconds of tape for each instructor.[55] Despite their very brief exposure to the instructors, these students produced highly consistent ratings of the instructors' qualities. After viewing no more than thirty seconds of an instructor's nonverbal behavior, the students substantially agreed with one another about which instructors were more competent, professional, and possessed other positive classroom qualities.[56]

Even more startling, these ratings showed highly significant correlations with end-of-semester evaluations the instructors had received from students enrolled in their courses. For students who viewed thirty seconds of silent videotape, their global rating (a sum of their ratings on the fifteen variables) of the instructor's personality produced a correlation coefficient of 0.76 ($p<.01$) with end-of-semester evaluations, explaining almost fifty-eight percent of the variance in those evaluations.[57] For students who

[55] The fifteen-second tapes used five seconds from each of the three portions of class time, while the six-second tapes used just two seconds from each of those portions. *See id.* at 437. Nine undergraduates rated the thirty-second segments, eight rated the fifteen-second clips, and eight judged the six-second versions. *See id.* at 433, 437. The researchers chose to use female undergraduate students for the study, attributing the basis for this decision to previous research suggesting that women are better than men at decoding nonverbal behaviors. *See id.* at 433.

[56] For the first group of students that viewed thirty-second video clips, reliability of the ratings ranged from a low of .60 (accepting, attentive, and honest) to a high of .89 (active and enthusiastic) with a mean of .72. *See id.* at 433. Reliability was comparable among students who viewed fifteen-second videos, and just slightly lower for those who saw only six seconds of video. *See id.* at 437–38.

[57] *See id.* at 434. A correlation coefficient expresses the strength and direction of a linear relationship between two variables. These coefficients range from -1.00, which signals a perfect negative relationship between the two variables, to 1.00, indicating a perfect positive association between the two. Most social scientists consider correlations between .1 and .3 (or -.1 and -.3) to be small; those between .3 and .5 to be moderate; and those over 5 to be large. *See* JACOB COHEN, STATISTICAL POWER ANALYSIS FOR THE BEHAVIORAL SCIENCES 79–81 (Lawrence Erlbaum Assocs. 2d ed. 1988) (1969); Will G. Hopkins, *A New View of Statistics: A Scale of Magnitudes for Effect Statistics*, SPORTSCIENCE (2002), http://www.sportsci.org/resource/stats/effectmag.html.

Although correlation coefficients are widely used, their relative sizes can mislead non-statisticians. For example, a correlation coefficient of .4 is not twice as large as one of .2; rather, it is four times as strong. To correct for this and offer a commonsense way of thinking about correlations, social scientists often refer to the "amount of variance" that one variable explains in another. This amount is the square of the correlation coefficient. A correlation coefficient of .2, for example, means that knowledge of one variable allows us to explain about four percent of the variation in the other variable. A coefficient of .4 means that one variable explains about sixteen percent of the variance in the other. *See generally* JACK LEVIN & JAMES ALAN FOX, ELEMENTARY STATISTICS IN SOCIAL RESEARCH 369 (10th ed. 2006).

viewed only *six* seconds of the same taped behavior, the correlation with end-of-semester evaluations was almost as high—a correlation coefficient of 0.71 ($p<.01$).[58] These correlations are strikingly high for social phenomena.[59]

Ambady and Rosenthal concluded that their findings were "quite remarkable: On the basis of observations of video clips just half a minute in length, complete strangers were able to predict quite accurately the ratings of teachers by students who had interacted with them over the course of a whole semester!"[60] Even participants who watched as little as six seconds of an instructor's silent classroom behavior "predicted with surprising accuracy" students' end-of-semester evaluations.[61] These findings confirmed "the considerable influence of very subtle affective nonverbal behaviors on the teaching process."[62]

Elisha Babad and Dinah Avni-Babad later collaborated with Rosenthal to build on the initial Ambady-Rosenthal study.[63] This second study used video clips of forty-seven different professors teaching sixty-seven different courses.[64] The instructors in this

---

The designation "$p < .01$" indicates the statistical significance of a correlation or other statistical relationship. The "p" level expresses the likelihood that a given relationship would have occurred purely by chance. When $p < .01$, the reported relationship would happen randomly—rather than because of an actual relationship—less than one time out of 100. By convention, social scientists accept p levels of less than .05 as "statistically significant." *See, e.g.*, *id.* at 230.

[58] *See* Ambady & Rosenthal, *supra* note 50, at 438. The students who rated the fifteen second clips achieved a smaller, statistically nonsignificant correlation with end-of-semester evaluations ($r = .44$, $p < .05$). *See id.*

[59] *See, e.g.*, COHEN, *supra* note 57, at 81 (stating that in social psychology, correlations of .50 are "about as high as they come"); Hopkins, *supra* note 57 (showing that correlation coefficients over .7 are "very large, very high, huge").

[60] Ambady & Rosenthal, *supra* note 50, at 435.

[61] *Id.* at 438.

[62] *Id.* at 440. The correlation between thin slices of behavior and teaching evaluations is so large, in fact, that these correlations are the highest researchers have obtained in thin-slice research. *See* Ambady et al., *supra* note 47, at 217–20. An earlier study, moreover, reached very similar results. In this study, conducted by Spallings and Spencer, ten participants viewed four-minute clips of teaching behavior by each of nine university accounting instructors. The participants agreed on how these nine instructors should rank on an overall measure of effectiveness. Their rankings, moreover, correlated significantly—a correlation coefficient of .70—with the instructors' rankings on end-of-semester evaluations. *See id.* at 209 (discussing the Stallings and Spencer study).

[63] *See* Elisha Babad, Dinah Avni-Babad & Robert Rosenthal, *Prediction of Students' Evaluations from Brief Instances of Professors' Nonverbal Behavior in Defined Instructional Situations*, 7 SOC. PSYCHOL. EDUC. 3, 3 (2004).

[64] *Id.* at 9.

study were experienced faculty, and their teaching spanned an even greater range of course types than in the Ambady and Rosenthal study.[65] Babad and colleagues also incorporated nonverbal aspects of speech in their study by using raters unfamiliar with the language spoken by the professors. Thirty-nine American undergraduates, graduate students, professors, or other professionals rated videotapes of professors teaching in Hebrew at the Hebrew University of Jerusalem.[66]

Each of these raters viewed nine seconds of tape drawn from each of the sixty-seven classes and evaluated each professor on three scales: "Likable, Warm, Friendly," "Competent, Effective, Professional," and "Boring, Passive" versus "Interesting, Active."[67] The researchers then correlated these ratings with end-of-semester evaluations offered by students enrolled in each course. Ratings of the professors' competence and ability to interest an audience, as manifested in the nine-second video clips, correlated strongly with the enrolled students' ratings of the professor's humor, enthusiasm, clarity, and overall classroom performance.[68]

Most recently, Dennis Clayson and Mary Jane Sheffet reproduced Ambady and Rosenthal's findings with more than 700 college students enrolled in business courses.[69] Clayson and Sheffet visited fourteen different sections of these courses just

---

[65] *See id.* at 7.

[66] *See id.* at 8–10. In this follow-up study, participants were exposed to professors' nonverbal aspects of speech such as pitch, pauses, speed, and other factors that do not depend on the content of speech.

[67] *Id.* at 10, 11. The judges showed a high degree of consistency in their ratings. *See id.* at 13.

[68] Babad and colleagues combined scores on these evaluation questions into an "instructional" composite score. *Id.* at 15. This score, in turn, correlated highly with ratings for competence and interest drawn from lecture video clips. *See id.* at 17. In this study, the ratings based on brief video clips did not correlate significantly with some end-of-semester ratings, such as those related to course workload and difficulty, instructor accessibility, readings, and course content overall. *See id.* at 15. The failure of these correlations to reach significance is understandable, given their more attenuated relation to characteristics discernible from brief video clips. The difference among ratings in this study, however, does not mean that educators can rely upon some portions of student evaluations to escape the "thin slice" effect. Multiple studies have shown that item-scores on teaching evaluations are heavily inter-correlated. No matter what specific questions a form asks, evaluations appear to tap a student's global evaluation of instruction. *See, e.g.*, Sylvia d'Apollonia & Philip C. Abrami, *Navigating Student Ratings of Instruction*, 52 AM. PSYCHOLOGIST 1198, 1199–1201 (1997).

[69] *See* Clayson & Sheffet, *supra* note 6, at 151–52.

after the first class had convened and the primary instructor had introduced him or herself to the students. The instructor had not yet distributed a syllabus or other materials to the class, and had spoken with the students for less than five minutes.[70] The instructor left the room after this brief exposure, while Clayson and Sheffet asked the students to complete questionnaires about the instructor's personality, rating the faculty member's agreeableness, creativity, conscientiousness, stability, and extroversion.[71] Consistent with Ambady and Rosenthal's research, the students' initial ratings of professors' personalities, based on less than five minutes of classroom contact, correlated significantly with conventional end-of-semester evaluations of the instructor's teaching.[72] Ratings of "agreeableness," "creativity," "conscientiousness," and "stability" each showed significant correlations with end-of-semester ratings, as did a global measure of personality combining all five dimensions.[73]

---

[70] *Id.* at 151. The fourteen sections included six in Organizational Management and eight in Principles of Marketing. *Id.*

[71] These five characteristics, as Clayson and Sheffet point out, are the "Big Five" personality dimensions. *Id.* Most psychologists agree that these five dimensions define a major portion of each individual's personality, are relatively fixed within each person, depend substantially on genetics, and have little cultural component. *See id.* Individuals high in agreeableness "tend to be friendly, trusting, and cooperative," while those who are conscientious are "methodical, well organized, and respectful of their duties." *Id.* Stability denotes people who are "relaxed, less emotional, and less prone to distress," and creativity characterizes those who are "open minded, creative, and interested in culture." *Id.* Extroverted individuals, finally, "will seek out the company of others and be energized by such interactions." *Id.*

[72] Students assessed the five personality factors by rating their instructor on five 7-point Likert scales. *Id.* at 152. For example, students indicated whether they found their instructor "disagreeable" or "agreeable," with one indicating most disagreeable and seven denoting most agreeable. *Id.* In addition to examining each of the five personality dimensions individually, Clayson and Sheffet averaged the five scores to obtain a "global" measure of personality. *See id.*

The end-of-semester evaluations included six fairly standard rating scales: (1) "[T]he instructor created an atmosphere conducive of learning;" (2) "[T]he instructor explains material appropriately;" (3) "[T]he instructor shows interest in student learning;" (4) "[T]he instructor sets high but reasonable standards;" (5) "[R]ate your satisfaction with your learning in this class;" and (6) "What grade would you give your instructor?" *Id.* Students answered each of these items with a letter grade (A–F). Clayson and Sheffet averaged the responses to obtain an overall assessment. *See id.*

[73] *See id.* at 154. Several other studies have reached similar results, showing very high correlations between evaluations gathered early in the semester and those collected at semester's end. *See* Richard G. Kohlan, *A Comparison of Faculty Evaluations Early and Late in the Course*, 44 J. HIGHER EDUC. 587, 589–91 (1973)

Students, in other words, rapidly form an impression of a professor's personality. An image based almost entirely on nonverbal behavior gels within the first few minutes of the semester. The students may refine their impressions as the semester progresses, but the initial image remains telling. The significant correlation between assessments completed after just five minutes of class and those offered at semester's end is daunting. Based on this correlation, Clayson and Sheffet concluded that traditional teaching evaluations "follow a seriously flawed paradigm."[74] They recommended the initiation of "research and discussions . . . to replace the current . . . system with some other form of evaluation."[75]

It is tempting to believe that law students are too sophisticated or well educated to react as strongly as other students to a professor's nonverbal behaviors. There is no evidence, however, to exempt any group of adults from this phenomenon. The subjects who applauded the meaningless lecture delivered in the initial Dr. Fox experiment were all graduate students or professionals, many with M.D.s and other advanced degrees.[76] Similarly, the students whose five-minute ratings of their professors accurately predicted end-of-semester evaluations were advanced business students with professional ambitions.[77] Even experienced trial judges react unconsciously to nonverbal behavior when sentencing defendants.[78] Responding to nonverbal conduct is not a sign of immaturity, low educational attainment, or carelessness; rather, these reactions are an essential element of how the brain functions.

(noting that evaluations collected after the first two hours of class correlated highly with those gathered during the last week of the semester); Matthew H. Sauber & R. Rodman Ludlow, *Student Evaluations Stability in Marketing*, 88 J. MIDWEST MARKETING 41, 43–46 (1988) (demonstrating a high correlation between evaluations collected during the second week of the semester and those gathered at semester's end).

[74] *See* Clayson & Sheffet, *supra* note 6, at 159.

[75] *Id.*

[76] *See supra* notes 23–28 and accompanying text. The authors of that study commented specifically on the "educational sophistication" of the audiences they had deceived. Naftulin et al., *supra* note 23, at 633.

[77] *See* Clayson & Sheffet, *supra* note 6, at 151–52; *supra* notes 69–73 and accompanying text.

[78] *See infra* notes 98–100 and accompanying text.

II. THE SHADOW SIDE OF NONVERBAL BEHAVIOR

A connection between nonverbal behaviors and teaching evaluations is not itself surprising. Teaching requires effective communication, and communication entails more than simply uttering words. An enthusiastic, expressive style maintains audience attention, emphasizes main points, and kindles deeper interest in the subject. Stylistic techniques can also enhance clarity and understanding. The fact that a professor's smiles, emphatic gestures, eye contact, changes in vocal pitch, and relaxed but confident movements correlate with more positive student evaluations of teaching is neither surprising nor threatening. Nonverbal behaviors surely play some role in good teaching, and many faculty members work to polish their classroom style, as well as their substantive knowledge.[79]

Yet, the research on student evaluations is troubling. It confirms not *some* connection between a professor's style and student evaluations, but an *overwhelming* link between those two factors. Nonverbal behaviors appear to matter much more than anything else in student ratings. Enthusiastic gestures and vocal tones can mask gobbledygook,[80] smiles count more than sample exam questions,[81] and impressions formed in thirty seconds accurately foretell end-of-semester evaluations.[82] The strong connection between mere nonverbal behaviors and student evaluations creates a very narrow definition of good teaching. By relying on the current student evaluation system, law schools implicitly endorse an inflexible, largely stylistic, and homogeneous description of good teaching. Rather than encouraging faculty to use nonverbal behaviors to complement excellent classroom content, organization, and explanations, the present evaluation system largely eliminates the "dog" of substance, leaving only the "tail" of style to designate good

[79] *Cf.* Nira Hativa, *Teaching Large Law Classes Well: An Outsider's View*, 50 J. LEGAL EDUC. 95, 104, 107–08 (2000) (offering advice to legal educators on ways to improve teaching, including tips on nonverbal behavior).

[80] *See supra* notes 23–28 and accompanying text.

[81] *See supra* notes 33–34 and accompanying text.

[82] *See supra* notes 50–62 and accompanying text.

teaching.[83] Neither law students nor faculty benefit from such a narrow definition of good teaching.

The psychology literature, moreover, identifies three further difficulties with the disproportionate role that nonverbal behaviors play in student evaluations. First, the behaviors that most influence these evaluations are rooted in physiology, culture, personality, and habit. Those behaviors are difficult for any faculty member to alter and they often reflect characteristics like race, gender, nationality, or socioeconomic class.[84] Second, the current evaluation process allows social stereotypes to filter students' perceptions of instructor behaviors. Students see the nonverbal behaviors of some faculty differently than they view identical behaviors in other professors, potentially placing women and minority faculty at a greater disadvantage.[85] Finally, the ratings that students award through the present evaluation system bear little relationship to objective measures of learning.[86] The current system of student evaluations, in other words, rewards and penalizes faculty according to relatively trivial indicia, rather than what they accomplish in the classroom.[87]

A. Nonverbal Behaviors and Mutability

Instructors are able to modify some of the nonverbal behaviors that affect student ratings: They can learn to move around the classroom with more ease, speak directly to students rather than lecture from notes, and gesture more emphatically.

---

[83] *Cf.* Clayson & Sheffet, *supra* note 6, at 158 (describing a finance professor who caused students' scores on a national exam to rise from the thirteenth to the ninety-seventh percentile, but whose scores on student evaluations "consistently placed in the lowest third of all faculty" and stating that "if good teaching is . . .what a student evaluation says it is, then this professor probably should be replaced"). "Should instructors be clones when it comes to behavioral manifestations? . . . If [these questions are] ignored, a hallmark of higher educational institutions, that is, diversity in its broadest sense, could become the victim." Audhesh K. Paswan & Joyce A. Young, *Student Evaluation of Instructor: A Nomological Investigation Using Structural Equation Modeling*, 24 J. MARKETING EDUC. 193, 200 (2002).

[84] *See infra* Part II.A.

[85] *See infra* Part II.B.

[86] *See infra* Part II.C.

[87] These flaws in the evaluation system, of course, do not mean that faculty members who receive high evaluations are bad teachers. A better assessment process, more focused on student learning, might identify many of the same teachers as effective. Eliminating arbitrariness and bias in the system, however, is essential for those who are unfairly penalized.

With training and practice, some faculty members can improve their evaluations—and their students' learning—by mastering these kinds of actions.[88]

Many of the behaviors that substantially shape student evaluations, however, are gestures, expressions, tones of voice, and other characteristics that stem from an instructor's physiology, culture, habit, and personality. These aspects of classroom behavior are "unintended" and "unconscious,"[89] and largely immutable. Professors who manifest these behaviors or appearances will never raise their evaluations beyond a settled ceiling, no matter how diligently they work at effective and engaging classroom presentations.[90]

Individuals with asymmetric facial features, for example, appear less agreeable, less conscientious, and more neurotic than individuals with symmetric features.[91] Professors with subtle facial asymmetries will seem more worried, nervous, careless, and disorganized to their students, as well as less helpful or sympathetic than their colleagues with more pleasing visages.[92] Overcoming these negative impressions is difficult, given the rapidity and lasting nature of "thin slice" judgments. Indeed, some research confirms that professors with attractive faces receive more positive student evaluations than those with less

[88] The psychologist Stephen Ceci, for example, improved his teaching evaluations dramatically through media training, and recounted his experience in the case study discussed above. *See supra* notes 35–39 and accompanying text.

[89] Ambady & Rosenthal, *supra* note 47, at 256.

[90] *See* Ambady et al., *supra* note 47, at 205; Clayson & Sheffet, *supra* note 6, at 158 ("[I]f . . . student perceptions are even marginally related to relatively long-lasting traits [in instructors], it may be true that some teachers never will receive consistently high evaluations in certain environments, irrespective of anything they do or possibly could do."); Murray, *Effective Teaching*, *supra* note 33, at 162.

[91] *See* Fahim Noor & David C. Evans, *The Effect of Facial Symmetry on Perceptions of Personality and Attractiveness*, 37 J. RES. PERSONALITY 339, 346 (2003); Todd K. Shackelford & Randy J. Larsen, *Facial Asymmetry as an Indicator of Psychological, Emotional, and Physiological Distress*, 72 J. PERSONALITY & SOC. PSYCHOL. 456, 464 (1997).

[92] *See* Noor & Evans, *supra* note 91, at 346.

desirable features.[93] Notably, the effect may be stronger for men than women, although it affects both genders.[94]

Similarly, some faces appear more competent than others and some facial structures look immature or unintelligent.[95] Although researchers have not tested this effect directly in the classroom, they have identified significant effects in several other contexts. Between 2000 and 2004, for example, the candidate with the more "competent" face won more than two-thirds of contested congressional elections.[96] Students are likely to incorporate the same biases, attributing more knowledge to faculty with "competent" faces and more warmth to those with babyish ones.[97]

Research likewise demonstrates that individuals with Afrocentric features appear more aggressive than people without those features.[98] Trial judges respond to that perception by

---

[93] *See, e.g.*, Hamermesh & Parker, *supra* note 3, at 369–71, 375 (detailing a regression analysis of 16,957 student evaluations completed at the University of Texas, which showed that attractive professors, as rated by undergraduates unfamiliar with the faculty member, received significantly higher evaluations than unattractive ones, with the rating difference between most and least attractive faculty constituting a full point on a five-point teaching evaluation scale); Todd C. Riniolo et al., *Hot or Not: Do Professors Perceived as Physically Attractive Receive Higher Student Evaluations?*, 133 J. GEN. PSYCHOL. 19, 30 (2006) (detailing results of naturalistic study based on www.ratemyprofessors.com, which confirms higher ratings for professors perceived as attractive).

[94] *See* Hamermesh & Parker, *supra* note 3, at 373.

[95] *See, e.g.*, Alexander Todorov et al., *Inferences of Competence from Faces Predict Election Outcomes*, 308 SCI. 1623, 1623 (2005); Leslie A. Zebrowitz et al., *Trait Impressions as Overgeneralized Responses to Adaptively Significant Facial Qualities: Evidence from Connectionist Modeling*, 7 PERSONALITY & SOC. PSYCHOL. REV. 194, 194 (2003) (noting individuals with child-like features are perceived as weak and naïve more often than those with more mature faces).

[96] *See* Todorov et al., *supra* note 95, at 1624. Competent appearance was judged by individuals from other states who did not recognize the candidates or realize that they were competing for an electoral position. *Id.* After briefly viewing a single black-and-white photo of each candidate, they indicated which individual they believed was more competent. *Id.* Competence ratings correlated with election results between 66.0% and 73.3% of the time. *Id.*

[97] The overall assessment of faculty with competent or babyish faces may depend on other attributes as well as the particular classroom context. Under some circumstances students may value highly a professor's intelligence, while under other circumstances, they may be more moved by a professor's warmth and support. *Cf. id.* at 1624 (noting that, although competent-appearing political candidates win elections significantly more often than those with less mature faces, the latter may secure an advantage in races in which integrity is a primary issue). The interplay of appearance and nonverbal behaviors is complex.

[98] *See* Irene V. Blair et al., *The Use of Afrocentric Features as Cues for Judgment in the Presence of Diagnostic Information*, 35 EUR. J. SOC. PSYCHOL. 59, 65 (2005); *cf.*

imposing longer sentences on criminal defendants with Afrocentric features than on those without those characteristics: A recent large-scale analysis identified a significant relationship between Afrocentric features and sentence length, even after carefully controlling for severity of the primary offense, concurrent offenses, and prior offenses.[99] The disparity, which effected both white and black defendants, was substantial: Individuals with heavily Afrocentric features received sentences seven to eight months longer than those with identical criminal histories but few Afrocentric features.[100] The same facial features that influence experienced trial judges very likely affect law students as well, prompting them to view professors with Afrocentric features as more hostile than their other professors.

Voice quality also affects interpersonal judgments. Individuals with attractive voices seem more competent, powerful, and warm than those with less desirable vocal qualities.[101] People who speak in babyish tones sound warmer than other people, but less expert or commanding.[102] Loud voices register as more authoritative and knowledgeable than soft

Joni Hersch, *Profiling the New Immigrant Worker: The Effects of Skin Color and Height* 1–7, 14–15 (Vanderbilt Law and Econ., Working Paper No. 07-02, 2007), *available at* http://ssrn.com/abstract=927038 (finding lighter skin color correlated with higher wages among recent legal immigrants to the United States).

[99] *See* Blair et al., *supra* note 16, at 676–77.

[100] *Id.* at 677–78. More precisely, a hypothetical white or black defendant with Afrocentric features one standard deviation above the mean for their race group, and with mean scores for each criminal history variable, would have received a sentence seven to eight months longer than a defendant with the same mean criminal history scores but Afrocentric features scoring one standard deviation below the mean for their group. *Id.*; *see also* William T. Pizzi et al., *Discrimination in Sentencing on the Basis of Afrocentric Features*, 10 MICH. J. RACE & L. 327, 333–36 (2005) (summarizing the work of Blair and her colleagues in this area).

[101] *See, e.g.*, Diane S. Berry, *Vocal Types and Stereotypes: Joint Effects of Vocal Attractiveness and Vocal Maturity on Person Perception*, 16 J. NONVERBAL BEHAV. 41, 51 (1992) [hereinafter Berry, *Vocal Types*]. "Attractive" voices differ by sex, but people show a high degree of consensus on which male and female voices are most attractive. *See* Diane S. Berry, *Vocal Attractiveness and Vocal Babyishness: Effects on Stranger, Self, and Friend Impressions*, 14 J. NONVERBAL BEHAV. 141, 141, 146–49 (1990).

[102] Berry, *Vocal Types*, *supra* note 101, at 51. Attractiveness and babyishness are separate dimensions of vocal quality, so these characteristics can interact to form a variety of distinct impressions. *See id.* at 43. Attractive and mature voices, for example, appear competent and powerful, but less warm than attractive, babyish voices. *Id.* at 51. Audiences perceive the latter as especially warm, but less powerful and competent. *Id.*

ones.[103] A professor's natural vocal tones, therefore, influence student perceptions of the professor's competence, warmth, knowledge, power, and other qualities.

In addition to these physiological features, a faculty member's learned mannerisms significantly affect student perceptions. By adulthood, these characteristics are as much part of us as our physical features. Speech patterns, for example, differ by culture and region. Americans tend to associate rapid speech with competence,[104] even though some Americans use more leisurely speech patterns. Students in U.S. law schools are therefore likely to prefer fast speaking faculty members, rating them as more intelligent and knowledgeable than slower speaking professors, even if the two groups of faculty deliver comparable content.

Similarly, white Americans engage in frequent eye contact, believing that it demonstrates honesty, integrity, and attention.[105] African Americans value eye contact less, and employ it less frequently while speaking.[106] When mulling the answer to a question, individuals in some cultures look up while members of other cultures look down.[107] Cultural differences like these can prompt a classroom of predominantly white American students to believe that faculty of color or foreign-born professors

---

[103] *See, e.g.*, Ying Peng et al., *The Impact of Cultural Background and Cross-Cultural Experience on Impressions of American and Korean Male Speakers*, 24 J. CROSS-CULTURAL PSYCHOL. 203, 214 (1993).

[104] *Id.* at 214–15; George B. Ray, *Vocally Cued Personality Prototypes: An Implicit Personality Theory Approach*, 53 COMM. MONOGRAPHS 266, 273 (1986). Speakers from other cultures, such as Korea, either disregard vocal speed in judging competence or rate slower speakers as more competent. *See* Peng et al., *supra* note 103, at 215–16 (finding that native Koreans did not associate vocal rate with competence and distinguishing an earlier study by Lee and Bolster finding that Koreans perceived slower speakers as more competent).

[105] *See, e.g.*, Elisha Babad, *Nonverbal Behavior in Education*, *in* THE NEW HANDBOOK OF METHODS IN NONVERBAL BEHAVIOR RESEARCH 283, 290 (Jinna A. Harrigan et al. eds., 2005).

[106] *See* Uwe Gielen et al., *Naturalistic Observation of Sex and Race Differences in Visual Interactions*, 9 INT'L J. GROUP TENSIONS 211, 213, 220 (1979); Marianne LaFrance & Clara Mayo, *Racial Differences in Gaze Behavior During Conversations: Two Systematic Observational Studies*, 33 J. PERSONALITY & SOC. PSYCHOL. 547, 549 (1976); Kyung Soon Lee & Angela Carrasquillo, *Korean College Students in United States: Perceptions of Professors and Students*, 40 C. STUDENT J. 442, 453 (2006) (discussing that white professors perceived that Korean students avoided eye contact during conversations).

[107] *See* Anjanie McCarthy et al., *Cultural Display Rules Drive Eye Gaze During Thinking*, 37 J. CROSS-CULTURAL PSYCHOL. 717, 721 (2006).

are less attentive, less candid, or otherwise less engaged with the material than white faculty members who more closely track white American cultural norms.[108]

Hand gestures and body movement also differ significantly by race and culture. African Americans, on average, use more intense body language than white Americans do.[109] Conversely, Chinese Americans, Japanese Americans, and Korean Americans use less expressive body language than whites, and display their emotions less visibly.[110] Japanese Americans are also less assertive than white Americans during verbal interactions.[111] Disparities like these can prompt white students, still the majority in most law school classrooms, to view African American professors as more hostile than white ones, while they view Asian American professors as cold, uncaring, or diffident.

Recent research, finally, suggests that sexual orientation also shapes nonverbal behavior. Using the "thin slice" method described above,[112] Nalini Ambady and two colleagues found that undergraduates correctly identified another person's sexual orientation about seventy percent of the time after viewing just

---

[108] Conversely, as the percentage of non-white and international students grows among law students, white faculty may find themselves rated negatively by students offended by their eye contact. *Cf.* Babad, *supra* note 105, at 290 ("[I]n some cultures, looking someone straight in the eye is not considered positive at all, but rather aggressive, daring, and impolite."); Judith A. Sanders & Richard L. Wiseman, *The Effects of Verbal and Nonverbal Teacher Immediacy on Perceived Cognitive, Affective, and Behavioral Learning in the Multicultural Classroom*, 39 COMM. EDUC. 342, 351 (1990) (explaining that for African American students, instructor eye contact was not related to perceived cognitive learning, but that eye contact was significant for Asian, white, and Hispanic students).

[109] *See* Sanders & Wiseman, *supra* note 108, at 351 (explaining that an instructor's tense body position correlated with white students' perceived learning, but not with that of black, Asian, or Hispanic students); Stella Ting-Toomey, *Conflict Communication Styles in Black and White Subjective Cultures*, *in* INTERETHNIC COMMUNICATION: CURRENT RESEARCH 75, 77 (Young Yun Kim ed., 1986) (explaining that blacks showed more confrontational conflict styles than whites did). *See generally* Thomas Kochman, *Force Fields in Black and White Communication*, *in* CULTURAL COMMUNICATION AND INTERCULTURAL CONTACT 193 (Donal A. Carbaugh ed., 1990).

[110] *See* Min-Sun Kim, *A Comparative Analysis of Nonverbal Expressions as Portrayed by Korean and American Print-Media Advertising*, *in* READINGS IN CULTURAL CONTEXTS 206, 213–14 (Judith N. Martin et al. eds., 1998).

[111] *See* William B. Gudykunst et al., *Uncertainty Reduction in Japanese-American/Caucasian Relationships in Hawaii*, 51 W.J. SPEECH COMM. 256, 269 (1987).

[112] *See supra* notes 47–62 and accompanying text.

ten seconds of silent videotape.[113] Students may make similar inferences about a faculty member's sexual orientation based on nonverbal cues. Those inferences can affect evaluations in a variety of ways. If heterosexual students are uncomfortable with gays and lesbians, and they perceive a professor's homosexual orientation through his or her nonverbal behavior, that professor may receive more negative evaluations. Alternatively, since judgments of sexual orientation are imperfect, some students may erroneously attribute a particular orientation to a professor and respond negatively to the gestures signifying that orientation. Research on the intersection of sexual orientation, nonverbal behaviors, and social judgments is just beginning, but may uncover significant concerns about teaching evaluations.

These traits constitute only some of the many nonverbal behaviors that influence judgments that students make about faculty.[114] The pervasive influence of these behaviors, combined with their immutability, explains the difficulty that many dedicated professors have encountered in trying to raise their scores on student evaluations. Although training has improved ratings for some faculty,[115] most professors realize modest gains at best. Even twenty weeks of professional instruction attended by one group of highly motivated faculty generated just a small increase in student evaluation scores.[116] Anecdotal reports express similar frustrations. Professors, for example, are puzzled to discover that expanding office hours and giving students

---

[113] *See* Nalini Ambady et al., *Accuracy of Judgments of Sexual Orientation from Thin Slices of Behavior*, 77 J. PERSONALITY & SOC. PSYCHOL. 538, 541 (1999).

[114] *See generally* Hillary Anger Elfenbein & Nalini Ambady, *On the Universality and Cultural Specificity of Emotion Recognition: A Meta-Analysis*, 128 PSYCHOL. BULL. 203, 203–04 (2002) (reviewing studies of cross-cultural emotion recognition); Donald L. Rubin, *Help! My Professor (or Doctor or Boss) Doesn't Talk English!*, *in* READINGS IN CULTURAL CONTEXTS, *supra* note 110, at 149, 149–50.

[115] *See, e.g.*, Williams & Ceci, *supra* note 6, at 23.

[116] *See* Harry Murray & Cheryl Lawrence, *Speech and Drama Training for Lecturers as a Means of Improving University Teaching*, 13 RES. HIGHER EDUC. 73, 86–87 (1980). *But see* Murray, *Classroom Teaching*, *supra* note 33, at 31. A professional actress taught the faculty members breathing and voice exercises, directed them in enacting short dramatic scenes, and gave them corrective feedback on their lectures. The group met for two hours each week. The faculty members who participated in the training averaged a 0.2 gain in student ratings, on a five-point scale, while a control group of faculty who were not participants in the training sessions realized no gains. An earlier study by one of Murray's graduate students identified even less pay-off from more modest attempts to provide feedback to professors on their classroom behaviors, although the lowest rated instructors realized some gains from that method. *Id.* at 29–32.

detailed contact information does not improve their ratings on "accessibility outside of class."[117] As two reviewers concluded, "accessibility" as measured on student evaluations "has more to do with personality than office hours."[118] And personality, especially as reflected through unconscious mannerisms, is notoriously hard to change.

B. Perceptual Filters: Stereotyping

A faculty member's gestures, voice tones, facial expressions, and other nonverbal behaviors profoundly shape what students believe about that professor's teaching effectiveness. The professor's actual behaviors, however, tell only half the story. Law students, like all humans, perceive other people's behavior through filters that are socially conditioned. None of us see the world through neutral, objective lenses. Instead, our minds classify individuals according to race, gender, age, and other socially salient categories with dizzying speed.[119] We then use those classifications to interpret a speaker's behavior so that the same gestures, expressions, and other components of nonverbal behavior look different depending on the speaker's race, gender, and other characteristics.

A long line of research, for example, demonstrates that Americans perceive smiles and friendliness more readily on white faces than on African American ones. Conversely, Americans more readily perceive anger and hostility in black individuals. Birt Duncan conducted one of the earliest studies in this area, showing a series of college students one of four videotapes in which a male student shoved another student after a heated disagreement.[120] The tapes systematically varied the

---

[117] *See* Clayson & Haley, *supra* note 46, at 13.

[118] *Id.*

[119] *See* Ambady et al., *supra* note 47, at 231 (citing various studies); *see also* Tiffany A. Ito et al., *The Social Neuroscience of Stereotyping and Prejudice: Related Brain Potentials to Study Social Perception*, *in* SOCIAL NEUROSCIENCE: PEOPLE THINKING ABOUT PEOPLE 189, 203 (John T. Cacioppo et al. eds., 2006). Kathleen Bean describes the power of these classifications in the context of a law professor: "[T]he gender gap . . . is born the moment I walk into the classroom. It has a life of its own before I open my mouth, before my body language speaks, and before my eyes make contact with anyone. My sex, and my sex alone, . . . opens the gender gap." Bean, *supra* note 5, at 29.

[120] *See* Birt L. Duncan, *Differential Social Perception and Attribution of Intergroup Violence: Testing the Lower Limits of Stereotyping Blacks*, 34 J. PERSONALITY & SOC. PSYCHOL. 590, 592–98 (1976) (discussing the phenomenon of

races of the disputing students, producing dramatically different responses among viewers. When a black student shoved a white peer, seventy-five percent of the viewers described the incident as "violent."[121] When a white student shoved a black, only seventeen percent of the viewers registered the same reaction. Instead, the students characterized this behavior as "playing around," "dramatizing," or "aggressive."[122] The race of the perpetrator, race of the victim, and interaction between the two significantly affected the viewers' perceptions.[123]

More recently, a research team led by Joshua Correll used a laboratory computer game to demonstrate dramatically different reactions when white and black actors displayed identical behaviors.[124] Correll and his colleagues constructed a game in which the researchers systematically varied the race of a series of men that appeared on the computer screen, the poses they adopted, and the objects they held. Game players were instructed to "shoot" men holding guns and "not shoot" those holding innocent objects like wallets.[125] The players proved significantly more likely to shoot blacks holding innocent objects

differential social perception of intergroup violence). The subjects believed they were viewing a live encounter between two students in a nearby room, using a closed circuit television. To standardize interactions, however, the interactions were taped. *See id.* at 592.

[121] *Id.* at 595.

[122] *Id.*

[123] The white viewers characterized a black student shoving another black as the most violent, a black shoving a white as next most violent, white shoving a black next, and a white shoving a white as most benign. *Id.* at 595. Race also affected the attributions that viewers attached to the students' behavior. Viewers attributed shoving by the black student to the student's personal attributes—for example, his violent nature—while they attributed shoving by a white to circumstantial factors such as that a stimulus caused the student to act. *Id.* at 597.

Sagar and Schofield obtained similar results in a study of sixth grade boys. *See* H. Andrew Sagar & Janet Ward Schofield, *Racial and Behavioral Cues in Black and White Children's Perceptions of Ambiguously Aggressive Acts*, 39 J. PERSONALITY & SOC. PSYCHOL. 590, 594 (1980) (exploring the way in which the interpretation of ambiguous social behavior can be influenced by racial stereotypes and cultural differences). When asked to judge ambiguous behavior in a hypothetical story, the boys rated the actions of black children as significantly more "mean and threatening" than identical actions by white children. *Id.*

[124] *See* Joshua Correll et al., *The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals*, 83 J. PERSONALITY & SOC. PSYCHOL. 1314, 1314 (2002) (examining the effect of ethnicity on participants' decisions to shoot or not to shoot while playing a simple video game).

[125] One computer key corresponded to "shoot," while another indicated "don't shoot," so subjects had to register a reaction one way or another. *Id.* at 1316.

than whites brandishing the same items.[126] They also showed a greater tendency to overlook whites holding guns than blacks.[127] Even when subjects accurately distinguished among targets, they took significantly longer to recognize whites wielding guns and blacks with innocent objects.[128]

These stereotypes affect the judgments of both black and white individuals. Correll and several other researchers have shown that both whites and blacks perceive black men as more dangerous than white men when they are engaged in identical actions.[129] Americans of both races unconsciously filter their perceptions of nonverbal behavior to see more aggression among blacks than among whites engaged in similar conduct.

Several scholars, moreover, have shown that these differences arise even in very subtle contexts. In a striking series of experiments, Kurt Hugenberg demonstrated that white students identify happy expressions on white faces significantly more quickly than they see sad or angry expressions on those faces.[130] The opposite pattern emerges when white students view

---

[126] *Id.* at 1318–19, 1322.

[127] *Id.* at 1319 (reporting results from the second study, finding that study participants were more likely to "not shoot" at video-game images of whites brandishing weapons than blacks). These results did not reach significance in the first and third studies, although they showed the same direction.

[128] *Id.* at 1317, 1322 (reporting results from the first and third studies finding that study participants were slower to respond when a video-game image of a white man holding a gun appeared on the screen than a black man, and were also slower to determine that objects held by images of black men were harmless objects than for similarly adorned white men). The window for response times in Study Two was so small that differences in response latencies were not detected. *See* Joshua Correll et al., *Event-Related Potentials and the Decision to Shoot: The Role of Threat Perception and Cognitive Control*, 42 J. EXPERIMENTAL SOC. PSYCHOL. 120, 120–28 (2006) (replicating the same results with yet another group of students); Justin D. Levinson, Forgotten Racial Equality: Implicit Bias, Decision-Making, and Misremembering 42 (Aug. 25, 2006) (unpublished manuscript), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=927547) (stating that subjects were more likely to remember aggressive acts attributed to African American actors than to white ones).

[129] *See* Correll et al., *supra* note 124, at 1325; Sagar & Schofield, *supra* note 123, at 594–95.

[130] *See* Kurt Hugenberg, *Social Categorization and the Perception of Facial Affect: Target Race Moderates the Response Latency Advantage for Happy Faces*, 5 EMOTION 267, 271–73 (2005). Students viewed the faces on computer monitors, using keystrokes to identify the expressed emotion. Hugenberg used character animation software to construct male faces that shared identical facial characteristics, varying only in skin color and expression. Pre-testing confirmed that students readily identified the expressions as happy, sad, or angry. In the main studies, each student responded to 160 trials using eight stimulus faces in random order. *Id.* at 270–71.

black faces. They identify sad or angry faces on blacks significantly more rapidly than they see happy expressions on those faces.[131]

Even neutral expressions elicit different responses depending on the race of the target. Pierre Philippot and Yanélia Yabar studied this phenomenon by showing white students a series of photographs that had been carefully selected for their neutral expressions.[132] The students were significantly more likely to associate neutral black faces, rather than neutral white faces, with emotions like "show[s] aggressiveness to others," "insult[s] others," and "boil[s] inwardly."[133]

These studies confirm that "how individuals perceive and categorize facial expressions can depend quite critically on who it is that is displaying the expression."[134] In the classroom, therefore, students are likely to detect warm, happy faces more quickly on white professors, while they perceive sad, angry, or hostile expressions more readily on blacks. A neutral expression on the face of a white professor may convey warmth to students,

---

[131] *See id.* at 271, 273. Hugenberg and a colleague determined that the effect is even greater among white students who show higher degrees of racial prejudice on the Implicit Attitude Test. *See* Kurt Hugenberg & Galen V. Bodenhausen, *Facing Prejudice: Implicit Prejudice and the Perception of Facial Threat*, 14 PSYCHOL. SCI. 640, 640–43 (2003) [hereinafter Hugenberg & Bodenhausen, *Facing Prejudice*]. The difference, however, occurs among even less prejudiced students. *See* Kurt Hugenberg & Galen V. Bodenhausen, *Ambiguity in Social Categorization: The Role of Prejudice and Facial Affect in Race Categorization*, 15 PSYCHOL. SCI. 342, 343–44 (2004).

[132] Pierre Philippot & Yanélia Yabar, *Stereotyping and Action Tendencies Attribution as a Function of Available Emotional Information*, 35 EUR. J. SOC. PSYCHOL. 517, 517–27 (2005). Philippot and Yabar conducted their study in Belgium, which has a significant North African immigrant population that has given rise to racial tensions and stereotypes. In particular, Philippot and Yabar state that North African immigrants in Belgium are "generally perceived as a threatening and aggressive group." *Id.* at 520.

[133] *See id.* at 523–24, 527. Philippot and Yabar tested two other stereotypes of North Africans: "Show exuberance" and "Show excitement." *See id.* at 522. It is less clear that these stereotypes mark black-white relationships in the United States. Unfortunately, Philippot and Yabar did not distinguish among reactions to these five different stereotypes.

The laboratory studies examining differential perceptions of hostility on black and white faces have been limited so far to male faces and actors. Hopefully researchers will expand these investigations to explore perceptions of black and white women. Meanwhile, accounts by black women faculty suggest that they, like black men, suffer from exaggerated perceptions of hostility. *See, e.g.*, Smith, *supra* note 1, at 114 (discussing the stereotype of black women as "Sapphire," a "tough, domineering, emasculating, strident and shrill" character).

[134] Hugenberg, *supra* note 130, at 275.

while the same expression on a black professor's face may connote hostility.[135] These differences can markedly affect student evaluations of faculty accessibility, caring, and other qualities.[136]

Perceptual filters also bias the ways in which people perceive identical behaviors of men and women. When a male employee establishes direct eye contact with another worker, he raises his credibility.[137] A female employee using the same behavior does not enhance her credibility; instead, she increases the perception that she will act coercively against the other worker.[138]

---

[135] Anecdotal evidence from campuses yields similar conclusions. A black student on one predominantly white campus noted: "Socially, blacks are pressed into being very passive and always grinning, otherwise they are immediately typecast as being hostile and aggressive." Carole Baroody Corcoran & Aisha Renée Thompson, *"What's Race Got to Do, Got to Do with It?" Denial of Racism on Predominantly White College Campuses*, *in* RACISM IN AMERICA 137, 142 (Jean Lau Chin ed., 2004).

[136] Much of the research on race biases, like this example, focuses on differences between blacks and whites. Analogous differences, however, mark comparisons of other race groups, and research is starting to illuminate these complex relations. *See, e.g.*, Glascock & Ruggiero, *supra* note 3, at 200–05 (analyzing evaluations of white and Hispanic professors by students of both ethnicities); Levinson, *supra* note 128, at 24–25, 36–48 (examining recall bias as applied to African Americans, whites, and Hawaiians); Dominic W. Massaro & John W. Ellison, *Perceptual Recognition of Facial Affect: Cross-Cultural Comparisons*, 24 MEMORY & COGNITION 812, 816–22 (1996) (comparing white American and native Japanese students).

[137] *See* Herman Aguinis et al., *Effects of Nonverbal Behavior on Perceptions of Power Bases*, 138 J. SOC. PSYCHOL. 455, 460–63 (1998). Aguinis and his colleagues asked 170 young adults with significant workplace experience to respond to a written vignette in which two employees, "John" and "Greg," discussed declining profits at their bank. The researchers systematically varied John's eye contact with Greg, his facial expression, and his body posture in the descriptions to capture a variety of nonverbal behaviors. *Id.* at 460. When the vignette described John as looking directly at Greg, readers rated John's credibility as significantly higher than when the vignette indicated that John looked around the room, glancing occasionally at Greg. *See id.* at 463. Credibility was assessed from reactions to a series of statements such as "John is a man who keeps his word" and "John tells the truth." *See id.* at 462.

[138] *See* Herman Aguinis & Christine A. Henle, *Effects of Nonverbal Behavior on Perceptions of a Female Employee's Power Bases*, 141 J. SOC. PSYCHOL. 537, 544–45 (2001). Aguinis and Henle used the same procedure adopted in Aguinis' earlier study, *see generally* Aguinas et al., *supra* note 137, but identified the two employees as "Mary" and "John." Aguinis & Henle, *supra*, at 541. When Mary was described as looking directly at John, readers did not believe that she was more credible than when she glanced at him only occasionally. Instead, they perceived her as possessing significantly more coercive power. *Id.* at 544–45. Coercive power is a supervisor's ability to punish a subordinate. Aguinis and Henle measured ability to coerce through reactions to statements like "Mary can give John undesirable job assignments" and "Mary can make things unpleasant on the job for John." *See id.* at 543.

Similarly, a relaxed facial expression in male employees connotes credibility, expert knowledge, legitimate power, and the ability to confer rewards on others.[139] A relaxed expression on the face of a male manager also suggests that the manager can make subordinates feel valued, approved, and important.[140] Calm expressions on the faces of female employees confer none of these positive benefits. Instead, a relaxed expression prompts observers to decrease their estimation of the woman's power.[141]

These biases, like those based on racial categories, can powerfully affect students' evaluations of male and female professors. Students are likely to perceive male instructors who establish eye contact with students and adopt a relaxed expression—attitudes typically effective in classroom teaching—as credible, knowledgeable, and in control of the classroom. The students will believe that the professor has legitimate authority to demand work from them and will focus on his power to reward them for good answers, rather than his authority to punish them

[139] *See* Aguinis et al., *supra* note 137, at 463. When the vignette described the facial expression of one worker, "John," as relaxed, readers rated him significantly higher on credibility, expert power, legitimate power, and reward power—all positive attributes in the workplace. *See id.* "Expert power" measures the special expertise that a superior employee can share with an inferior. Aguinis and his colleagues measured expert power through reactions to statements like "John can share with Greg his (John's) considerable experience and/or training." *Id.* at 462. Legitimate power refers to an employee's perceived authority to command others. It was measured in this study through reactions to statements like "John can make Greg recognize that he (Greg) has tasks to accomplish." *Id.* Reward power, finally, consists of the ability to confer benefits on another worker. This study established perceptions of reward power through reactions to statements like "John can increase Greg's pay level." *Id.*

[140] *See id.* at 463. In this study, a relaxed expression on a male manager's face significantly increased ratings of "referent power." That type of power includes the perceived ability to make others feel "valued," "approve[d]," "personally accepted," and "important." *See id.* at 462.

[141] *See* Aguinis & Henle, *supra* note 138, at 544–46. Aguinis and Henle identified just one nonverbal behavior that appeared to have a positive effect on perceptions of female managers. A relaxed body position, signified by leaning back in a chair with legs crossed, increased ratings of referent power. *See id.* at 544–45. Readers, in other words, were more likely to think that this female employee could confer feelings of acceptance, value, and importance on another worker. Aguinis and Henle termed this finding "surprising" and "did not consider this a finding of strong theoretical significance" because of its incongruity with prior theoretical work. *Id.* at 545. The finding, however, may hold a clue to how female professors *do* achieve positive ratings from students. When they appear relaxed in the classroom, that appearance may connote high status, consistent with a general correspondence between those variables, which for women translates into an ability to confer "soft" benefits like feelings of value, acceptance, and importance.

for bad ones. They will see the professor's ability to make them feel valued, approved, and important.

Female faculty who adopt the same classroom demeanors will evoke far different responses. They are less likely to gain credibility or the appearance of expert knowledge in students' eyes. Students will be less likely to acknowledge the female faculty member's authority to make assignments or demand work in the same way they will perceive that power in a male professor. They will also be less likely to focus on the female professor's ability to reward them for good work or to enhance their feelings of value and importance. Instead, students are more likely to focus on the female instructor's power to punish them for poor work.[142]

Fewer studies examine the impact of socioeconomic status on our perceptions of others, but those that exist suggest that it also shapes attitudes. One well-developed line of experiments uses vehicles—a common class marker in American life—to probe reactions to people of high and low socioeconomic class. Anthony Doob and Alan Gross performed the first of these experiments, demonstrating that drivers were significantly more likely to honk

[142] This type of power is a negative one leading to "resistance . . . , lower organizational commitment, and dissatisfaction." *Id.* at 544–45. Several studies comparing student comments on teaching evaluations reveal gender differences parallel to these differences. *See, e.g.*, Farley, *supra* note 5, at 339 (stating that evaluations of female law faculty criticize them disproportionately for "being too strict or for being 'task-masters' "); Michael A. Messner, *White Guy Habitus in the Classroom*, 2 MEN & MASCULINITIES 457, 458 (2000) (describing evaluations concluding that male co-teacher was "objective," "relaxed and comfortable," "flexible," "open-minded," "good humored," and "look[ed] at all sides of an issue," while a female colleague was perceived as "biased," having "an agenda," having a "chip on her shoulder," "rigid and dogmatic," "politically correct," "grumpy and angry").

Empirical surveys suggest that female faculty do, on average, receive lower evaluations than their male colleagues. A recent, multivariate analysis of almost 17,000 student evaluations at the University of Texas revealed significantly higher evaluations for male faculty, even after controlling for course type and instructor status. *See* Hamermesh & Parker, *supra* note 3, at 370, 373. Another recent study of evaluations gathered in 741 different courses taught at twenty-one different institutions showed that women faculty received significantly lower ratings from male students than females, while male and female students did not differ in their assessment of male faculty. *See* John A. Centra & Noreen B. Gaubatz, *Is There Gender Bias in Student Evaluations of Teaching?*, 71 J. HIGHER EDUC. 17, 26 (2000). Female instructors, however, received higher ratings from female students. *Id.* Earlier studies reached mixed results, but these most recent, and most comprehensive, studies suggest that male and female faculty receive somewhat different ratings.

at an old, rusty car that failed to move promptly through an intersection than at an expensive, new, and well-maintained car.[143] Andrew McGarva and Michelle Steiner observed similar driver responses when testing their reactions towards being honked at by a rusty Ford pickup, the low-status vehicle, and new Nissan Pathfinder SUV, a high-status vehicle.[144] Drivers accelerated away from the Ford pickup significantly more quickly than they drove away from the SUV after the other driver honked at them.[145]

Law students do not honk at professors who displease them, but the same attitudinal differences can affect relationships in the classroom. The "horn honking" studies expose a cultural tendency to vent frustration or hostility more readily against low-status individuals than high-status ones. Socratic classrooms, challenging material, and intense competition for grades are at least as frustrating to students as a stalled car at an intersection. Law students may express that irritation more readily on evaluations of professors with low-status mannerisms than in their assessments of faculty with more high-status appearances. Indeed, the horn-honking studies may explain the surprising degree of overt hostility that law students express on evaluations of some minority faculty.[146] Those evaluations are a type of classroom "honking."[147]

---

[143] *See* Anthony N. Doob & Alan E. Gross, *Status of Frustrator as an Inhibitor of Horn-Honking Responses*, 76 J. SOC. PSYCHOL. 213, 213–16 (1968). The drivers also honked more quickly at the low status cars and were more likely to honk repeatedly at those cars. *Id.* at 216. Andreas Diekmann and several colleagues complemented this research by finding that the status of the *blocked* car also affects responses in these situations. Drivers of higher status cars are more likely to honk their horns or flick their headlights at the waiting car than are drivers of lower status cars. Andreas Diekmann et al., *Social Status and Aggression: A Field Study Analyzed by Survival Analysis*, 136 J. SOC. PSYCHOL. 761, 761–64, 768 (1996).

[144] Andrew R. McGarva & Michelle Steiner, *Provoked Driver Aggression and Status: A Field Study*, 3 TRANSP. RES. PART F 167, 167–79 (2000).

[145] *Id.* at 173–74. The "provoking" driver also held "his mouth expressively agape" and "raised both arms impatiently." *Id.* at 172. Status influences on the road, however, may differ more among cultures than some other behaviors do. *See, e.g.*, Diekmann et al., *supra* note 143, at 768.

[146] *See supra* note 2 and accompanying text.

[147] One such comment was made in an anonymous student note given to an African American female professor:

> A black mammy like you is completely incompetent to judge anyone on anything. I do not care whether you are magna cum laude . . . . You do not belong in law school teaching. Black mammies should stay at home doing mammy things. Or they should stay in their place and it is not law school.

These examples comprise just a sample of the kind of stereotypes that routinely bias perceptions based on race, gender, class, and other categories. Researchers have documented a substantial number of other biases, with still others to be uncovered.[148] The bottom line is that, especially when the mind makes judgments based on "thin slices" of nonverbal behavior—the type of judgments students seem to be expressing on teaching evaluations—these stereotypes automatically and unconsciously alter perceptions. Students, like other humans, ascribe different meaning to the same behaviors depending on the race, gender, class, and other characteristics of the actor.

Familiarity does not necessarily reduce these perceptual distortions; instead, the differences may grow through self-reinforcing cycles. If students perceive hostility in the ambiguous expression of a black professor, coerciveness in the eye contact of a woman, or obstructive behavior in a professor from an underprivileged background, they will respond to those impressions with heightened hostility of their own. The faculty member, in turn, confronts negative reactions that have no apparent source. He or she has engaged in neutral (or even positive) behaviors that appear to have provoked an angry response. Faced with this seemingly irrational hostility, the black, female, or low-status faculty member may display subtle signs of their own discomfort and anxiety. Students will perceive those behaviors, reinforcing their initial negative images. False impressions generated by stereotypes, in other words, create a cycle of mutually reinforcing behavior.[149] By the semester's end, students may feel quite justified in rating the black, female, or

Smith, *supra* note 1, at 179.

[148] *See, e.g.*, Marc David Pell, *Evaluation of Nonverbal Emotion in Face and Voice: Some Preliminary Findings on a New Battery of Tests*, 12 TENNET 499, 502 (2002) (showing that subjects recognized neutral expressions more readily on male faces and expressions of disgust more readily on female ones); Ron Tamborini & Dolf Zillman, *College Students' Perception of Lecturers Using Humor*, 52 PERCEPTUAL & MOTOR SKILLS 427, 427–32 (1981) (demonstrating that students respond differently to male and female professors using similar humor).

[149] *See* Chamallas, *supra* note 5, at 202 (describing such a cycle when students respond critically to a female law professor); Hugenberg & Bodenhausen, *Facing Prejudice*, *supra* note 131, at 643; Lu-in Wang, *Race as Proxy: Situational Racism and Self-Fulfilling Stereotypes*, 53 DEPAUL L. REV. 1013, 1048–80 (2004) (discussing the phenomenon of self-fulfilling stereotypes in a variety of legal contexts); Carol O. Word et al., *The Nonverbal Mediation of Self-Fulfilling Prophecies in Interracial Interaction*, 10 J. EXPERIMENTAL SOC. PSYCHOL. 109, 119 (1974).

low-status professor as hostile and uncaring. Because humans are so sensitive to nonverbal signals of aggression, support, and other attitudes, even small differences in those perceptions have "powerful implications" for ongoing interactions.[150]

C. The Missing Link with Learning

The link between nonverbal behavior and teaching evaluations allows irrelevant characteristics like race and gender to distort student assessments of faculty. The same association hints that conventional evaluations bear little relationship to student learning. The professionals who lauded Dr. Fox's entertaining but meaningless discussion of mathematical theory and medical education could not have learned much: There was nothing in the lecture to learn.[151] Behaviors detectable in just thirty seconds of silent videotape seem unlikely to promote solid learning, yet these elements strongly influence student evaluations.[152] Do student evaluations of faculty correlate with student learning?

The cumulative research suggests that there is little, if any, positive association between the ratings students give faculty and the amount they learn. The most recent study, in fact, suggests a negative correlation between evaluations and learning. In a particularly well-designed investigation, two business professors gathered eight full years of data on students who completed two sequential accounting courses at a midwestern university.[153] After controlling for ACT scores, overall GPA, and grades in the first course, the researchers discovered that students who completed the first course with highly rated professors achieved significantly *lower* grades in the second course.[154] The professors with top evaluations, in other

---

[150] Hugenberg & Bodenhausen, *supra* note 131, at 643.

[151] *See supra* notes 23–25 and accompanying text.

[152] *See supra* notes 47–62 and accompanying text; *see also* Clayson & Sheffet, *supra* note 6, at 157–58 ("Finding an association between the final evaluation of the class at the end of the term and a personality evaluation made within 5 minutes of exposure . . . makes a validity argument for the relationship between personality and evaluation difficult to defend.").

[153] Penelope J. Yunker & James A. Yunker, *Are Student Evaluations of Teaching Valid? Evidence from an Analytical Business Core Course*, 78 J. EDUC. BUS. 313, 313–14 (2003).

[154] *Id.* at 315–16. Students' overall GPA, introductory course grade, and ACT scores, which are designed to measure pre-college aptitude, on the other hand, all showed significant positive correlations with their grades in the intermediate course.

words, did not prepare students for the more advanced course as well as lower rated faculty did.

This study is particularly noteworthy because it captured multiple facets of learning. By examining performance in a subsequent course, one that built directly on material learned in the first class, the investigation tapped students' long-term retention of material, as well as their motivation for learning. Some scholars suggest that even if highly rated professors do not enhance their students' immediate achievement, they inspire an enthusiasm for learning that has longer term pay-off.[155] In the business curriculum study, however, this relationship did not hold. Students who completed the initial course with highly rated professors performed less well than students who learned the introductory material from lower rated faculty. The results of this analysis lend support to speculation that a desire for better teaching evaluations may pressure faculty to "dumb down" the material they present to students.[156]

In several other studies, teaching evaluations failed to show any correlation, positive or negative, with student achievement. An examination of students in two sequential economics courses failed to find any significant relationship between faculty evaluations in the first course and student performance in the advanced offering.[157] Professors whom students rated poorly nonetheless prepared them as effectively as more highly-rated colleagues. Another more recent study directly probed the

Id. at 315. As one would expect, the students' general academic ability, as reflected by ACT score and overall GPA, their diligence, as reflected by overall GPA, and their relative performance in the initial course, all predicted achievement in the subsequent course. *See id.* The notable finding of the Yunkers' study is that, after controlling for these variables, students who learned the initial material from more highly rated professors performed less well than students who studied that material from professors with lower evaluations. *Id.* at 316.

[155] *See, e.g.*, Mike Allen et al., *The Role of Teacher Immediacy as a Motivational Factor in Student Learning: Using Meta-Analysis to Test a Causal Model*, 55 COMM. EDUC. 21, 28–29 (2006); Paul L. Witt et al., *A Meta-Analytical Review of the Relationship Between Teacher Immediacy and Student Learning*, 71 COMM. MONOGRAPHS 184, 200 (2004).

[156] *See, e.g.*, JOHNSON, *supra* note 6, at 49–50.

[157] Stephen Shmanske, *On the Measurement of Teacher Effectiveness*, 19 J. ECON. EDUC. 307, 308, 311, 313 (1988); *see also* Ganesh Mohanty et al., *Multi-Method Evaluation of Instruction in Engineering Classes*, 18 J. PERSONNEL EVALUATION EDUC. 139, 143, 146 (2005) (finding no significant relationship between student ratings of instruction and student learning as measured by comparing students' scores on tests administered both prior to and upon completion of engineering courses, although study included a small number of students).

relationship between the nonverbal behaviors that influence teaching evaluations and student performance in a course. Students enrolled in fifteen different sections of an introductory course scored their instructor's nonverbal behaviors shortly before completing a common midterm exam.[158] Scores on the midterm showed no significant correlation with the students' ratings of their instructors.[159]

Stephen Ceci, the psychology professor who documented his success in substantially raising student evaluations by adopting minor changes in his nonverbal behavior, offers further evidence that student evaluations bear little relationship to learning. After implementing a media consultant's advice, Ceci's evaluations rose almost a full point on a five point scale.[160] The students enrolled in the post-training course, moreover, believed that they learned more from a professor who gestured emphatically and varied his voice tones. On a five point scale, these students awarded Ceci an average of 4.05 for how much they had learned, compared to just 2.93 registered by students the previous semester.[161] The media training and changes in nonverbal behavior, however, did not affect students' actual performance on quizzes and exams; those scores were virtually identical across the two semesters.[162] Ceci's changed style, in other words, improved his evaluations but did not enhance the students' learning.

A few studies do identify a positive relationship between student evaluations and learning. Harry Murray, for example,

---

[158] Jon A. Hess et al., *Is Teacher Immediacy Actually Related to Student Cognitive Learning?*, 52 COMM. STUD. 197, 208–09 (2001).

[159] *Id.* at 210–11; *see also* Joseph L. Chesebro, *Effects of Teacher Clarity and Nonverbal Immediacy on Student Learning, Receiver Apprehension, and Affect*, 52 COMM. EDUC. 135 (2003); Debra Q. O'Connell & Donald J. Dickinson, *Student Ratings of Instruction as a Function of Testing Conditions and Perceptions of Amount Learned*, 27 J. RES. & DEV. EDUC. 18, 19, 22 (1993) (finding that student ratings of instructor in education course failed to correlate with student learning as measured by comparing students' scores on tests administered both prior to and upon completion of the course).

[160] *See supra* notes 35–39 and accompanying text.

[161] Williams & Ceci, *supra* note 6, at 16, 20. This finding is consistent with a large number of studies finding that student evaluations of teaching correlate highly with students' perceptions of how much they have learned. *See, e.g.*, O'Connell & Dickinson, *supra* note 159, at 22; Witt et al., *supra* note 155, at 201. Students, in other words, believe that they learn more from highly-rated professors—but usually, they do not. *See* O'Connell & Dickinson, *supra* note 159, at 18–19, 22.

[162] Williams & Ceci, *supra* note 6, at 21.

found a correlation of 0.30 between students' ratings of their instructor and final exam scores in a study of multiple sections of an introductory psychology course.[163] Differences in evaluation scores, in other words, explained about nine percent of the variance in student learning.[164] An earlier meta-analysis by Peter Cohen identified a somewhat higher correlation of 0.43 between teaching evaluations and student learning, suggesting that the former explain about 18.5% of the variance in the latter.[165] Researchers, however, have noted several flaws in Cohen's analysis: It included studies in which students rated their professors only after receiving course grades, as well as surveys based solely on students' subjective beliefs about the extent of their learning.[166] In light of these and other problems, Cohen joined two other evaluation experts in agreeing that additional research was needed.[167] Those more recent studies, as noted above, have shown no positive relationship between teaching evaluations and student learning.

In a meta-analysis of eighty-one studies published between 1979 and 2001, finally, Paul Witt and several colleagues identified a small positive relationship between "immediacy," the complex of personality and nonverbal behaviors that produce high student evaluations, and objective measures of student achievement like exam scores.[168] Given the relationship between immediacy and evaluations, this analysis could signal a parallel association between evaluations and learning. The correlation

[163] Murray, *Effective Teaching Behaviors*, *supra* note 33, at 151; *see also* Jeff Koon & Harry G. Murray, *Using Multiple Outcomes to Validate Student Ratings of Overall Teacher Effectiveness*, 66 J. HIGHER EDUC. 61, 68, 73–74 (1995) (analyzing the same data).

[164] *See supra* note 57 (explaining the concepts of correlation and variance).

[165] *See* Peter A. Cohen, *Student Ratings of Instruction and Student Achievement: A Meta-Analysis of Multisection Validity Studies*, 51 REV. EDUC. RES. 281, 295–96 (1981).

[166] *See, e.g.*, Philip C. Abrami et al., *Validity of Student Ratings of Instruction: What We Know and What We Do Not*, 82 J. EDUC. PSYCHOL. 219, 220, 222–24 (1990); Hinton, *supra* note 7, at 566. The former problem confounds attempts to correlate student achievement with faculty ratings because a high evaluation offered after receiving course grades may reflect student satisfaction with their grades rather than an independent judgment of faculty quality. *See supra* note 10 (discussing the relationship between student grades and teaching evaluations). The second flaw substantially diluted Cohen's conclusions: Most scholars consider student perceptions of learning a poor measure of their actual learning. *See, e.g.*, Witt et al., *supra* note 155, at 189.

[167] *See* Abrami et al., *supra* note 166, at 231.

[168] *See* Witt et al., *supra* note 155, at 184–85, 196.

calculated by Witt and his colleagues, however, was just 0.122,[169] meaning that variation in the measured nonverbal behaviors explained just 1.5% of the variance in student learning. Although statistically significant, Witt and his colleagues concluded that these "low associations" were "less than meaningful" in any practical sense.[170]

The nonverbal behaviors that students reward on teaching evaluations, in other words, produce little, if any, gain in student learning. Those behaviors seem to make learning more enjoyable for students, and that is a worthwhile goal for faculty to pursue. To the extent that nonverbal behaviors are mutable, professors can attempt to learn mannerisms that will increase students' enjoyment of their classes as well as the faculty member's teaching evaluations. Dedicating substantial time to polishing these nonverbal behaviors, however, is unlikely to improve student understanding. Faculty can do more to enhance actual learning by skipping the media training sessions and devoting that time directly to their students or class materials.

The price of our current system of student evaluations, moreover, is much higher than the opportunity cost of time spent trying to change nonverbal behaviors. As the previous sections explain, student evaluations impose serious risks of bias. Some of those distortions specifically burden white women, faculty of color, and other traditionally disadvantaged groups. Others penalize any faculty member with the wrong type of face, voice, gestures, or other nonverbal behaviors. The correlation between conventional teaching evaluations and student learning would

[169] *Id.* at 196. This figure represents the correlation of student learning with combined measures of verbal and nonverbal immediacy. *Id.* Attempts to separate verbal and nonverbal immediacy yielded even lower correlations. *Id.* "Cognitive learning" embraces all types of learning that law faculty attempt to convey. In other words, it includes thinking and analysis skills, as well as mastery of basic facts and principles.

[170] *Id.* at 200. Social science conventions agree that correlations of this magnitude are weak, with little practical significance. *See supra* note 57.

have to be very high to justify this kind of unfairness.[171] Instead, the relationship to learning outcomes is minimal at best.

III. THE COGNITIVE FOUNDATIONS OF JUDGMENT

The psychology literature paints a disturbingly negative picture of student evaluations: These assessments respond primarily to minor aspects of a professor's classroom style; many of those behaviors reflect characteristics like race, gender, and class; and evaluation scores bear little, if any, relationship to student learning.[172] Why are student evaluations so biased?

The problem does not lie with students. It inheres in the process we use to gather their input. Despite some claims to the contrary, students have essential feedback to offer faculty on teaching. They can tell professors what they learned from a course and how that compared to what they expected to learn. They can describe the educational techniques that worked for them and those that did not. They can provide suggestions for how a faculty member might teach differently. Law students can assess the quality of their educational experience in myriad ways.

Obtaining useful feedback from any evaluator, however, requires collecting information under circumstances that allow for meaningful, deliberative reflection. As the Nobel prize-winning psychologist Daniel Kahneman explains, there are two types of human thought processes: System One and System Two.[173] System One processes are "fast, automatic, effortless,

[171] Evaluations are most problematic when they are used to determine a particular faculty member's fitness for tenure, promotion, salary increase, or other reward. Even a relatively large correlation between two variables in a population tells us very little about any one population member. A hypothetical correlation of .50 between student learning and evaluations, for example, would be higher than any ever detected but still would explain only twenty-five percent of the variance in student learning. *See supra* note 57. Fully three-quarters of the variation in student learning, under these circumstances, would stem from other factors. And despite the overall correlation, some faculty with low evaluations would generate more learning in their students than professors with higher evaluations. For this reason, overall correlations provide a poor basis for judging population members individually. One observer argues that a correlation between two variables should reach .90, a very high threshold, before it is used to rate individuals. Hopkins, *supra* note 57.

[172] The extensive literature on teaching evaluations raises other questions on the validity of these measures, particularly their relationship to students' expected and actual grades. *See supra* note 10. Some of these issues are important to legal education but are beyond the scope of this Article.

[173] Daniel Kahneman, *A Perspective on Judgment and Choice: Mapping*

associative, implicit (not available to introspection), and often emotionally charged; they are also governed by habit and are therefore difficult to control or modify."[174] System Two processes are "slower, serial, effortful, more likely to be consciously monitored and deliberately controlled; they are also relatively flexible and potentially rule governed."[175]

As rational adults, we assume that most of our judgments derive from System Two. We perceive ourselves as reasoned creatures who "think things through." In fact, however, almost all judgments begin as System One intuitions; we have gut reactions to just about everything.[176] System Two deliberations challenge and override our initial System One decisions in a surprisingly limited number of cases.[177] Instead, we are more likely to use System Two cognitive processes to justify the intuitive judgments we generate using System One's more emotional leaps.[178]

Students' evaluations of faculty begin, like most judgments, as System One processes. As such, they are heavily influenced

---

Bounded Rationality, 58 AM. PSYCHOLOGIST 697, 698 (2003). Kahneman credits Stanowich and West with proposing the terms "System One" and "System Two," although the framework rests on work done by Kahneman and his collaborator Amos Tversky. *Id.* Other scholars have used terms like "experiential" and "analytic," or "experiential" and "rational." *See, e.g.*, Seymour Epstein, *Integration of the Cognitive and the Psychodynamic Unconscious*, 49 AM. PSYCHOLOGIST 709, 710 (1994); Paul Slovic et al., *Rational Actors or Rational Fools: Implications of the Affect Heuristic for Behavioral Economics*, 31 J. SOCIO-ECON. 329, 330 (2002).

[174] Kahneman, *supra* note 173, at 698.

[175] *Id.*

[176] *See id.* at 716–17.

[177] *See id.* at 717. Although it is impossible to know precisely how often System Two overrides System One in daily life, Kahneman estimates that the "most common" paths of decision making are either those that involve only System One—when System Two offers no corrections at all—or those in which System Two only partially and inadequately corrects System One. *Id.* All judgments, in the common meaning of that word, do activate System Two to some extent because they require explicit choices. *Id.* at 699. As Kahneman explains, however, many judgments occur without System Two overriding the instinctive direction proposed by System One—they are "intuitive" judgments. *Id.*

[178] R.B. Zajonc, a pioneer in this area of psychology, observed:

> We sometimes delude ourselves that we proceed in a rational manner and weigh all the pros and cons of the various alternatives. But this is probably seldom the actual case . . . . Most of the time, information collected about alternatives serves us less for making a decision than for justifying it afterward.

R.B. Zajonc, *Feeling and Thinking: Preferences Need No Inferences*, 35 AM. PSYCHOLOGIST 151, 155 (1980).

by first impressions—even though the students have known the faculty member for at least a semester. When the brain makes a System One decision, it does not assess all of the information gathered on a subject over time as with System Two processes. Instead, the mind uses its quick-fix System One toolbox: It seizes its first impressions of the subject. In a sense, the students' minds reach back to the first impressions they formed of the instructor, using those as the basis for their evaluation. Those impressions, of course, incorporate all of the biases described above.

The dominance of System One thinking offers a possible explanation for the particularly high correlation between end-of-semester teaching evaluations and "thin slice" judgments of the same teachers. Students' assessments at a course's summation may not just correspond to those first impressions—they may *be* those impressions. Because the process we use to collect end-of-semester evaluations does not encourage System Two—deliberative—reflection, the students' brains most likely retrieve their first impressions and report those as their current evaluation.[179]

Obtaining more meaningful evaluations from students, therefore, requires finding ways to engage System Two thinking. There are at least three reasons why the current evaluation system fails to encourage that deeper thought. First, System One works with special efficiency in deciding whether things are good or bad.[180] Research, in fact, suggests that the brain has specialized neural circuitry that renders these good/bad distinctions promptly and confidently.[181] When the brain is asked to make any good/bad judgment, it is less likely to invoke System Two's more reflective thinking because System One handles these decisions so efficiently. This does not mean that

[179] Other mechanisms may also contribute to the high correlation between evaluations of teaching based on brief observations of a professor and end-of-semester evaluations of the same professor. Different theorists have suggested that first impressions may be accurate in this context, that those impressions may be self-reinforcing, or that students may be incapable of forming deeper, more considered assessments of teaching. *See* Ambady & Rosenthal, *supra* note 50; Clayson & Sheffet, *supra* note 6. The simplest explanation, however, may be the best: The mind draws upon first impressions as part of its System One thinking.

[180] *See* Kahneman, *supra* note 173, at 701; Zajonc, *supra* note 178, at 154–56.

[181] Kahneman, *supra* note 173, at 701 (citing Joseph E. LeDoux, *Emotion Circuits in the Brain*, 23 ANN. REV. NEUROSCIENCE 155 (2000)).

the brain invariably makes correct good/bad decisions using System One, but it does tend to rely upon that process for these decisions.

Rating a professor's teaching requires the brain to make good/bad distinctions: We ask students to indicate how "good" or "bad" various aspects of that teaching were. This is especially true of the numerical judgments that most evaluation processes demand from students. The very nature of these forms, which ask students to rate value without requiring other types of reflection, discourages the brain from moving beyond System One instinctive reactions.[182]

Second, System Two thinking requires time. Our brains excel at intuitions, first impressions, and quick judgments, but struggle to make more complex decisions. As one pair of psychologists recently observed, "human cognitive capacities are limited" so "[p]roblem solving is hard work."[183] Humans have restricted working memory, making it difficult for us to process multiple pieces of data.[184] To make thoughtful evaluations, the brain needs time to recall diverse bits of data, compare them, group pieces of information into larger chunks of partial judgments, and ultimately yield a reasoned response. Researchers have repeatedly established that time pressure induces reflexive System One thinking rather than more reflective System Two processes.[185]

---

[182] Many law school forms, paralleling those used in other university departments, ask students both to make numeric ratings and to offer comments. The forms, however, almost invariably seek numeric ratings before comments, channeling the brain into fairly simplistic "good/bad" thinking. The number of students who offer comments, moreover, almost always falls far short of those providing numerical scores.

[183] Klaus Oberauer & Reinhold Kliegl, *A Formal Model of Capacity Limits in Working Memory*, 55 J. MEMORY & LANGUAGE 601, 601 (2006).

[184] Some psychologists have concluded that the human brain can hold only four "chunks" of data in working memory at one time. *See, e.g.*, Nelson Cowan, *The Magical Number 4 in Short-Term Memory: A Reconsideration of Mental Storage Capacity*, 24 BEHAV. & BRAIN SCI. 87, 88 (2000). Other psychologists conceptualize the limits on working memory in other ways. *See id.* (noting seven alternative views); Oberauer & Kliegl, *supra* note 183, at 608 (proposing a particularly sophisticated model). All, however, agree that working memory is constrained.

[185] *See, e.g.*, Anton J. Dijker & Willem Koomen, *Stereotyping and Attitudinal Effects Under Time Pressure*, 26 EUR. J. SOC. PSYCHOL. 61, 72 (1996) (discussing how stereotypes exerted a greater role on decisions made under time pressure); Melissa L. Finucane et al., *The Affect Heuristic in Judgments of Risks and Benefits*, 13 J. BEHAV. DECISION MAKING 1, 1–8 (2000) (explaining that subjects perceived greater inverse relationship between risks and benefits of activities when

For students, evaluating a professor's teaching requires more complex deliberation than most faculty members realize. As instructors, our beliefs about the teaching process are omnipresent. They shape our class preparation as well as our actions in the classroom on a regular basis. Even if we do not focus consciously on how we teach, the process is salient to us. Students, however, concentrate on the product of our teaching rather than the process. They attend daily to what they are learning and what they still need to know; the underlying process is secondary. For students to evaluate a semester's worth of teaching, they must recall the assignments, lectures, explanations, and other characteristics of a professor's work, examine those memories from a new perspective, and judge their efficacy in achieving a variety of goals. Comprehensive, accurate, and reflective evaluations take more than five, ten, or even fifteen minutes.[186]

Finally, since the human brain has limited capacity, any cognitive distraction impairs System Two thinking. Laboratory research demonstrates that divided attention reduces reasoned thought and thus promotes reliance on intuitive, stereotyped, and other automatic System One processes.[187] People have trouble thinking about two issues at once or switching quickly from one mental task to another.

Students complete traditional teaching evaluations under tremendous cognitive load. They have just finished a challenging law school class. They may be filling gaps in their notes or digesting the professor's final comments.[188] With the

considering them under time pressure than when assessing them without time pressure); Kahneman, *supra* note 173, at 711; David M. Sanbonmatsu & Russell H. Fazio, *The Role of Attitudes in Memory-Based Decision Making*, 59 J. PERSONALITY & SOC. PSYCHOL. 614, 614–22 (1990) (describing how impressions affect consumer decisions under time pressure).

[186] Although many professors allot ten to fifteen minutes of class time for students to complete evaluations, most faculty observe that students rarely use all of this time. Law students are anxious to move on to other tasks such as lunch, another class, reviewing their notes, or a meeting with a friend, and tend to complete these evaluation forms quickly.

[187] *See, e.g.*, Kahneman, *supra* note 173, at 711; C. Neil Macrae et al., *Creating Memory Illusions: Expectancy-Based Processing and the Generation of False Memories*, 10 MEMORY 63, 71–80 (2002); Sabine Sczesny & Ulrich Kühnen, *Meta-Cognition About Biological Sex and Gender-Stereotypic Physical Appearance: Consequences for the Assessment of Leadership Competence*, 30 PERSONALITY SOC. PSYCHOL. BULL. 13, 13, 17, 20 (2004).

[188] Conversely, if a professor distributes evaluations at the beginning of a class

end of the semester near, they probably are worried about their performance on final examinations. They may be thinking about an upcoming job interview, a meeting with a friend, or a reunion with family members; the mind offers a large number of distractions. Laboratory research suggests that even very small distractions, such as memorizing a nine digit number between tasks, can significantly reduce reflective thought and enhance reliance on stereotypes.[189] The cognitive demands that law students face are much more substantial than these distractions, virtually forcing them to rely on intuitive System One channels to complete teaching evaluations.

The key to more meaningful evaluations of teaching, in sum, lies in creating the circumstances that allow students to engage more reflective System Two thought processes before providing those insights. Students need time and mental space to transcend intuitive System One judgments. They need mental prompts to help them review the semester's work. Like all decision makers, they need assistance in moving beyond reflexive "good/bad" judgments. The final section of this Article explores how law schools might design such a system.

IV. DEEPER THINKING, BETTER EVALUATION

The legal system, like all social structures, rests on judgments that individuals make about others. Lawyers decide which law graduates to hire and promote; clients choose which firms to retain. Negotiators appraise their opponents to calculate a successful offer, and deal makers judge how far they can press an advantage. Trial lawyers size up jurors while jurors assess lawyers, witnesses, and parties. Appellate advocates tailor their arguments to the bench while the judges register the advocate's sincerity. Student evaluations of teaching are just one type of interpersonal judgment occurring within a constant stream of judgments that individuals make about others.

All of these social evaluations risk the biases described above. Humans rest their judgments of others on intuitive System One thinking. Nonverbal behaviors, filtered by social stereotypes, powerfully affect those assessments. Research that

period, students may initially be focused on their previous class or the material they have come to discuss. They must put all of these thoughts aside to focus on evaluation.

[189] Sczesny & Kühnen, *supra* note 187, at 17, 20.

has identified aids to deliberative decision making in other contexts, therefore, can be used to design a process that promotes more reflective student evaluations of teaching. Moreover, reducing bias in student evaluations of teaching may illuminate paths to better decision making in other aspects of the legal profession.

At a minimum, thoughtful evaluation of teaching requires time and attention; otherwise, System Two thought processes fail to engage. Psychologists have identified a number of other conditions that promote more reflective, deliberative judgments. These include (1) encouraging decision makers to be as accurate as possible,[190] (2) focusing evaluators on the individuality of the person they are assessing,[191] (3) reminding decision makers to consider relevant data,[192] (4) facilitating group discussion of judgments,[193] and (5) establishing accountability to a third party.[194]

Although it is possible to identify workable processes that embrace the features that promote reflective, deliberative judgments, our current system of gathering student evaluations incorporates few, if any. Gregory Munro, for example, recommends using Small-Group Instructional Diagnosis ("SGID") to assess law school teaching.[195] In this process, a facilitator meets with small groups of students to gather their impressions of a course and instructor. The students discuss their

---

[190] Mary E. Wheeler & Susan T. Fiske, *Controlling Racial Prejudice: Social-Cognitive Goals Affect Amygdala and Stereotype Activation*, 16 PSYCHOL. SCI. 56, 57 (2005).

[191] Wheeler & Fiske, *supra* note 190, at 57.

[192] Kahneman, *supra* note 173, at 711–12.

[193] Richard F. Martell & Keith N. Leavitt, *Reducing the Performance-Cue Bias in Work Behavior Ratings: Can Groups Help?*, 87 J. APPLIED PSYCHOL. 1032, 1033–34 (2002).

[194] Thomas E. Ford et al., *The Role of Accountability in Suppressing Managers' Preinterview Bias Against African-American Sales Job Applicants*, 24 J. PERS. SELLING & SALES MGMT. 113, 113–24 (2004); Wheeler & Fiske, *supra* note 190, at 57.

[195] GREGORY S. MUNRO, OUTCOMES ASSESSMENT FOR LAW SCHOOLS 136 (2000). Professors Gerald Hess and Eric Orts have each adopted similar techniques to gather ongoing feedback from law students. *See generally* Gerald F. Hess, *Student Involvement in Improving Law Teaching and Learning*, 67 UMKC L. REV. 343 (1998) (referring to the groups as "Student Advisory Teams"); Eric W. Orts, *Quality Circles in Law Teaching*, 47 J. LEGAL EDUC. 425, 425–26 (1997) (describing his classroom use of "quality circles"). The technique derives from management tools developed by Edward Deming and implemented by Japanese industries and other businesses worldwide. *See* Hess, *supra*, at 347–48; Orts, *supra*, at 425–26.

perspectives as a group, expanding the information available to each student, checking individual biases, establishing accountability, and implicitly noting the seriousness of the process and need for accuracy.[196] These group discussions reduce cognitive overload by focusing attention and providing adequate time for thoughtful assessment.

To achieve the best results, the facilitator meets with the faculty member before these group discussions, obtaining information about the course goals, the subject matter, and the professor's pedagogic strategies.[197] The faculty member may provide a brief written statement of his or her teaching objectives, offering students further focus for their discussion. A professor may also note issues that troubled him or her during the semester, inviting student feedback on those matters.

Students who participate in these small-group discussions applaud them enthusiastically.[198] Indeed, research suggests that students do not like traditional, end-of-semester written evaluations, but prefer group discussions that promote dialogue with the professor.[199] Group discussions more thoughtfully involve students in assessment and underscore the school's interest in their input. Students have the opportunity to view the purposes of legal education from a perspective different from their own, often enhancing their own educational commitment and learning strategies.

Faculty members also learn more from these discussions than they do from standard teaching evaluations. Thoughtful discussion with students who possess awareness of the professor's pedagogic objectives produces more detailed and informative feedback. With modest training, law faculty can successfully facilitate these sessions for one another. As they do so, they broaden their personal knowledge of classroom

---

[196] *See* MUNRO, *supra* note 195, at 137.

[197] *See id.* at 136.

[198] *See, e.g.*, Robert D. Abbott et al., *Satisfaction with Processes of Collecting Student Opinions About Instruction: The Student Perspective*, 82 J. EDUC. PSYCHOL. 201, 203–206 (1990); Hess, *supra* note 195, at 351–52, 355–61; Orts, *supra* note 195, at 425.

[199] *See, e.g.*, Abbott et al., *supra* note 198, at 201–06. That dialogue may not occur directly. When evaluating faculty for tenure or promotion, especially, it is better to use another faculty member to facilitate the student discussion. Faculty, however, can establish an ongoing discussion with their own students about their teaching, and can teach in a manner that is responsive to students' concerns.

techniques, student responses, and pedagogic successes or failures. The process of facilitation and evaluation may create an ongoing dialogue among faculty, underscoring their joint commitment to teaching.

Small group discussion of teaching can also inform broader curricular goals within the law school. Faculty members have few opportunities to consider the curriculum as a whole, or even the significance of their courses within a student's three-year law school career. Discussions of a particular course and instructor can include questions about why the students took the course, how it fit within their broader educational goals, how prepared they feel for advanced offerings in the area, and whether they would have preferred a different type of offering. These individual discussions can fuel broader curricular reflections within the school.

This type of faculty review, of course, takes more time than traditional student evaluations. Most institutions would find it difficult to evaluate every course every semester using small group discussions. It is unlikely, however, that schools genuinely need to assess every course and faculty member that frequently.[200] Schools might aim to evaluate each course once every three years, meaning that about one-sixth of the curriculum would undergo full evaluation each semester. Faculty eligible for tenure or promotion could be evaluated somewhat more frequently. The current evaluation system produces a large amount of data every semester, but those data have limited value. Generating smaller amounts of high-quality information would better serve institutional needs.[201]

Schools can require faculty members to continue gathering regular feedback for formative purposes in every course they teach.[202] Some faculty may do this with conventional written

[200] Research suggests that even students tire of the number and repetitiveness of the evaluation forms they complete. *See, e.g.*, Abbott et al., *supra* note 198, at 201.

[201] Students currently devote about fifteen minutes per course—a total of seventy-five minutes per semester for students enrolled in five courses—to completing faculty evaluations. Thus, requiring each student instead to participate in one small group evaluation session would consume no more of the students' time while giving students a more satisfying experience and generating more productive evaluations.

[202] Evaluators distinguish between "formative" and "summative" evaluations. Formative evaluations are used to inform an ongoing process, such as a class, and help adapt the process to the participants' needs. Summative evaluations are used to evaluate an experience once it has concluded.

evaluations, although schools should explain the drawbacks of these forms. Other professors may use a modified version of the Small-Group Instructional Diagnosis, conducted by the faculty member himself or herself during the semester.[203] Other techniques for assessing one's own performance include "minute paper[s]" and informal evaluation forms designed by the professor.[204] Students particularly like these techniques when they are used mid-semester, with the faculty member responding to student suggestions.[205]

Informal evaluations of this nature would allow students to comment regularly on all courses while giving faculty members useful feedback in time to implement changes. Indeed, students would gain more from these formative evaluations—because they allow a faculty member to respond and adapt—than they do from the current overload of summative assessments that are used primarily to rate faculty after a course has finished. This combination, then, of informal evaluations administered by the professor, used to inform his or her teaching, with small group assessments in selected classes each semester, would give students the broadest opportunity to provide feedback on instruction and benefit from that process.[206]

Student-centered methods of instructional evaluation can also fit within a broader framework that includes teaching assessments offered by colleagues, alumni, education experts, and professors themselves.[207] Evaluation scholars repeatedly

---

[203] *See* Hess, *supra* note 195, at 343 (describing a Student Advisory Team as "a group of students who meet periodically with the teacher to help the teacher improve the course"); Orts, *supra* note 195, at 425–27 (discussing the structure of "quality circles," groups that meet with the professor regularly throughout the semester and are comprised of student representatives, either voluntary or elected, of the entire class).

[204] *See* Hess, *supra* note 195, at 346.

[205] *See* Abbott et al., *supra* note 198, at 205.

[206] It is essential, however, that formative evaluations remain informal, confidential exercises between the professor and students enrolled in the course. Including these assessments in evaluations of faculty for promotion, salary increases, or other purposes would undercut the goals of the small group discussion process.

[207] *See* Filippa Marullo Anzalone, *It All Begins with You: Improving Law School Learning Through Professional Self-Awareness and Critical Reflection*, 24 HAMLINE L. REV. 324, 371 (2001) ("The responsibility of the legal academy is to provide the forum and incentives for faculty to become better teachers . . . ."); Laurie A. Babin et al., *Teaching Portfolios: Uses and Development*, 24 J. MARKETING EDUC. 35, 40 (2002) (explaining how statements from a teacher's colleagues, alumni of the institution, or clients of students in client-based classes "can provide evidence of

stress the need to complement student evaluations with other forms of assessment. Yet traditional student evaluations, so easily implemented and appearing to generate "hard" data, predictably overwhelm these other techniques. Shifting the nature of student evaluations will allow faculties to match those assessments more effectively with those offered by other evaluators.

Given the multiple benefits of replacing conventional student evaluations with more meaningful processes, law school faculties should commit any time required to work out the details of small group discussions or other new methods of evaluating teaching.[208] As scholars, we criticize others for adopting "quick fixes" rather than expending the time and resources needed for meaningful evaluations and fair processes. We should be willing to apply the same standards to ourselves. Failing to do so is particularly unjust to our minority colleagues, who appear to suffer disproportionately from current evaluation systems.[209] It is also unfair to students, who deserve both to be heard more

effective teaching"); Gerald F. Hess, *Improving Teaching and Learning in Law School: Faculty Development Research, Principles, and Programs*, 12 WIDENER L. REV. 443, 458–61 (2006) (describing different ways that one's colleagues can contribute to faculty development); Melissa J. Marlow, *Blessed Are They Who Teach an Upper-Level Course, for They Shall Earn Higher Student Ratings*, 7 FLA. COASTAL L. REV. 553, 574–75 (2006) (suggesting that experts be used to evaluate teachers). *See generally* Daniel Gordon, *Does Law Teaching Have Meaning? Teaching Effectiveness, Gauging Alumni Competence, and the MacCrate Report*, 25 FORDHAM URB. L.J. 43 (1997) (comparing the usefulness of student evaluations and alumni surveys); Markovits, *supra* note 7 (discussing the use of market indicators to determine an academic's teaching skill).

[208] Law faculty and other scholars who have already devoted considerable attention to these details have cleared the way. *See, e.g.*, Hess, *supra* note 195, at 354 (describing the steps taken by one professional throughout the Student Advisory Team process); Orts, *supra* note 195, at 427.

[209] *See supra* notes 1–3 and accompanying text. Especially disheartening is the dearth, despite the extensive research and scholarly commentary on teaching evaluations, of studies focusing on the influence of race on these evaluations. Katherine Grace Hendrix, who conducted one of the few exploratory studies in this area, commented: "[R]esearchers have overlooked the classroom experiences of teachers and professors of color. In particular, the experience of being a member of a subordinate minority functioning as a professional within a predominantly White educational environment has escaped the interest of the White social scientist." Katherine Grace Hendrix, *Student Perceptions of the Influence of Race on Professor Credibility*, 28 J. BLACK STUD. 738, 739 (1998) (citation omitted); *see also* Huston, *supra* note 9, at 600–01 (commenting on the scarcity of research related to racial biases in student evaluations). Law schools have an opportunity to lead the rest of the academy in identifying and remedying these biases.

effectively on teaching quality and to learn to thrive under instructors representing diverse races, cultures, and backgrounds.

Improving our method of evaluating teaching is not a panacea for eliminating bias and the role of intuitive judgments in that process. Stereotypes affect how we remember information, as well as how we perceive it.[210] Reasoned decisions remain anchored in first judgments, giving those initial impressions lasting power.[211] Stereotypes based on race, gender, and class are persistent and resist efforts to overcome them. A professor's reputation, shaped in part by stereotypes, may influence even reasoned discussions.[212] And some influences on teaching evaluations, such as students' tendency to downgrade faculty who have graded them negatively,[213] fall outside the scope of the biases discussed in this Article. Progress in any field, however, requires taking small steps. If we do not attempt to improve the quality of our decision making, we remain trapped forever at its lowest levels.

CONCLUSION

Law and legal education assume reflective, rational decision making. Yet psychologists have shown that most of our judgments originate with intuitive preferences. The human brain reacts automatically to nonverbal behaviors and other subtle cues in the environment. Social stereotypes further shape our perceptions on an unconscious level. Creating conditions that support deliberative, reflective thinking is much harder than we believe.

As educators, we can take an important step toward understanding the interplay of intuitive and analytic thinking by examining those processes in the context of routine teaching evaluations. Those assessments draw heavily on the brain's automatic processes, allowing minor stylistic mannerisms and

---

[210] *See* Levinson, *supra* note 128, at 11, 22–28.

[211] *See* Kahneman, *supra* note 173, at 712.

[212] The few studies on the influence of instructor reputation on current evaluation systems find a very strong effect. *See generally* Bryan W. Griffin, *Instructor Reputation and Student Ratings of Instruction*, 26 CONTEMP. EDUC. PSYCHOL. 534 (2001) (discussing how instructor reputation correlated significantly with end-of-semester ratings, even after controlling for numerous other factors). It is possible that this effect would shape other modes of evaluation as well.

[213] *See supra* note 10.

stereotypes to color ratings. Designing evaluation systems that prompt more reflective, rational input would accord students enhanced respect, improve instruction, and treat faculty colleagues more fairly. Exploring such systems will also increase our own understanding of the intricate processes that drive decision making, knowledge we can apply to almost every field of law.