

Neutral Competence?
Polygraphy and Technology-Mediated Administrative Decisions

Vera Katelyn Wilde
Charlottesville, Virginia

M.A., Foreign Affairs, University of Virginia, 2010
B.A., Political Theory, University of Virginia, 2005

A Dissertation presented to the Graduate Faculty of the University of Virginia in
Candidacy for the Degree of Doctor of Philosophy

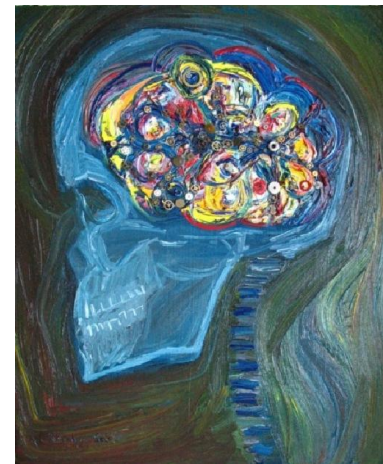
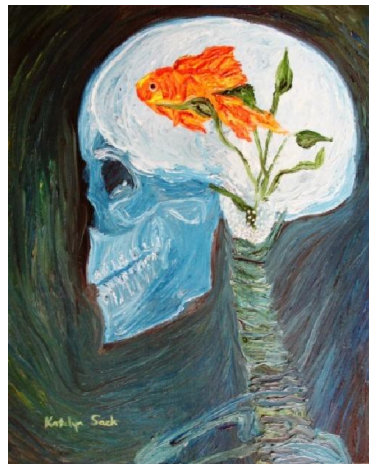
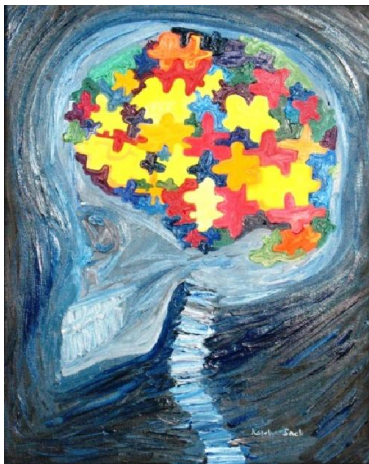
Department of Politics

University of Virginia
May, 2014

Haiku Summary:

Technology seems
neutral and data-driven.
Its use is human.

Chapter 1: Neutral Competence? Polygraphy and Technology-Mediated Administrative Decisions	p. 1
Chapter 2: Truth, Lies, and Polygraph Tape: Earlier Development of the Contemporary Surveillance State	p. 13
Chapter 3: Truth, Lies, and Polygraph Tape: Later Development of the Contemporary Surveillance State	p. 53
Chapter 4: g Factors for G Men? How Law Enforcement Hiring Tools Affect Police Department Diversity and Brutality	p. 87
Chapter 5: Neutralizing Prejudice or Smuggling Bias in a Trojan Horse? Race, Fairness, and Technology-Mediated Administrative Decisions	p. 123
Chapter 6: Does Procedural Fairness Make Inequality Just? Attitudinal Effects of the Ideology of Administrative Neutrality	p. 161
Chapter 7: Technology, Neutrality, and the Future of Politics: Bias-Free Does Not Mean Value-Neutral	p. 174
Bibliography	p. 182
Appendices as supplemental documents.	



“Psychic X-Ray: Jigsaw,” “... Fish Bowl,” and “...Clockwork,” oils on 16” x 20” stretched canvasses.

Neutral Competence?
Polygraphy and Technology-Mediated Administrative Decisions

How neutral are decision-making technologies such as lie detection in security contexts, medical diagnosis tools in healthcare, and welfare benefits administration tools in social work? The growth of bureaucracy is a core characteristic of 20th century American politics. Administrative decision-making technologies seem to constrain bureaucratic discretion: they seem relatively rule-bound, objective, and thus bias-free. Yet existing evidence does not actually establish the neutrality of those technologies. This matters, because these technologies affect most Americans' life chances in terms of life, liberty, and the means to pursue happiness. As denizens of the information age, we already know that increases in the number of such technologies and their use are steady and large: technology pervades modern life. Specifically, as a series of recent whistleblower leaks have established, we are living in a post-9/11 era in which means for mass data collection (the Big Data revolution) have changed the parameters of technology-mediated government. But mainstream public discourse is just beginning to incorporate this knowledge, and scholars do not know how neutral administrative technologies in general really are.

Decision-making technologies might be biased in three sorts of ways. First, they might systematically institutionalize cognitive biases, perhaps outside the conscious awareness of the people using them. For example, street-level bureaucrats such as police, doctors, and social workers might manifest racial bias, confirmation bias, and the combination of the two (intersectional bias), in interpreting information used to make technology-mediated decisions. Specifically, due to associations between blacks and

crime, police polygraphers might be more likely to interpret a polygraph chart as indicating deception if the subject’s background investigation indicates that he is a black felon than if the background investigation indicates the subject is an upstanding white guy. Some evidence suggests that this type of racial, confirmation, and intersectional bias adversely affects hiring decisions for blacks, felons, and especially black felons (Pager 2007; Bertrand and Mullainathan 2004). But scholars have not yet assessed whether these biases systematically affect technology-mediated decisions in the criminal justice realm.

In the medical realm, associations between blacks and mental illness might lead doctors to enter cognitive/emotional symptoms before physical ones for an indigent black patient than for a white patient not on welfare or disability when using a medical diagnosis tool to generate a probability-ranked differential diagnosis. Some research suggests that this type of racial substereotype shapes diagnosis and treatment of mental illness (Harris-Perry 2011; Metzl 2009; Strakowski et al. 1996; Trierweiler et al. 2000). But scholars have not yet established whether these substereotypes systematically affect technology-mediated decisions in the medical diagnostic realm.

In the welfare realm, stereotypes about single black mothers might lead people using a simple food stamp calculator to make different amount eligibility estimates for single mothers, black mothers, and especially single black mothers, as compared to their married, white counterparts. Public opinion research tends to show strong racial substereotype effects in relation to welfare in particular (Winter 2008; Gilens 2000). But scholars have not yet extended this research to address the question of whether these biases can systematically affect technology-mediated welfare benefits calculations.

Moreover, cognitive biases that street-level bureaucrats might institutionalize through apparently neutral decision-making technologies could combine with potential attitudinal-level effects of these technologies to weaken egalitarian norms. What if people *think* administrative decisions are completely fair and neutral – regardless of whether they actually are? This belief might affect a range of social and political attitudes, such as beliefs that the world is fair, or “just world beliefs” (Lerner 1980; Lerner and Miller 1978; Lerner and Simmons 1966). In this way, unequal outcomes, such as racial disparities in criminal justice, health, and social welfare, could become more normatively acceptable at the attitudinal level even as they become less factually equal at the procedural level.

In addition to these potentially inequalitarian behavioral decision-making and attitudinal effects, decision-making technologies could cause unintended consequences or simply not work at the institutional level. For example, technology-mediated employee screening tools such as polygraphs and voice stress analysis are increasingly used in nationally representative local and state law enforcement agencies to select recruits. But the efficacy of these tools in decreasing police misconduct, such as excessive use of force against civilians, or in increasing departmental diversity in terms of race and gender, has not been established.

Finally, decision-making technologies might be value-laden in deeper ways. Quite apart from potentially granting ascriptive biases a patina of neutrality by institutionalizing (perhaps unconscious) prejudices at the behavioral decision-making level, or simply not working as intended at the organizational level, technology might construct state power and social control. For example, decisions about what can be standardized, how, and in

what contexts can be decisions about how to prioritize values that may be in tension.

Where these decisions appear administrative and apolitical, as opposed to being constitutive conflicts of the political realm, one might argue that they are being shaped by power dynamics at discursive and epistemic as well as practical levels. In this way, technology-mediated decisions might be constrained by, and illustrate lack of democratically responsive constraint on, the changing boundaries of what is considered legitimate government.

Overall, then, there are three sorts of ways in which technology-mediated decisions may limit equality and liberty while increasing the appearance of fairness and limited government. Researchers have not yet connected the literatures on public administration and bias in the context of technology. So the very question of the neutrality of technology-mediated decisions constitutes a novel synthesis of diverse literatures. This dissertation addresses that question in a systematic fashion across a range of contexts, using qualitative, quasi-experimental, and experimental methods.

At the same time that this motivating question is itself novel, a broad array of scholarship suggests bias in technology is worth exploring. Indeed, leading theories of implicit and explicit racial attitudes would predict that bias systematically hurts minorities (Mendelberg 2008; Sears, Sidanius, and Bobo 2000). More broadly, social psychology theories predict that mental shortcuts influence decisions that may appear neutral (Bargh 1994; Pettigrew 1979). Public policy scholarship evaluating the efficacy of many widespread social interventions is not much more encouraging about whether racial and confirmation bias affect technology-mediated public administrative decisions. It tends to show common programs, such as Drug Abuse Resistance Education

(D.A.R.E.) in schools and Critical-Incident Stress Debriefing (CISD) for first responders and others, either do not work or cause harm by increasing rates of Post-Traumatic Stress Disorder development (T. D. Wilson 2011). And research on the sociology of science and technology tends to trace how conceptions of expertise, due process, and privacy – concepts at the heart of how technology is used in the exercise of legitimate political power – are fluid and bound by cultural context (Frickel and Moore 2006; Jasanoff 1995). For instance, DNA evidence – often considered the gold standard of objective forensic science – is actually fraught with factual questions about proper statistical interpretation (frequentist or Bayesian?), database use (do database searches disadvantage blacks by inflating posterior odds for disproportionately represented subgroups, or not?), and privacy rights or privileges (what are they, when do they apply, and to whom?). Meanwhile, the FBI denies independent researchers access to the National DNA Index System in defiance of its founding statute – underscoring the political nature of expertise in relation to technologies the state partly controls (Krane et al. 2009).

In this politically constrained information environment, my research demonstrates that apparently scientific and neutral technologies often have imperfect evidentiary bases. Consequently, they can wind up institutionalizing cognitive bias in some contexts, and otherwise value-laden political constructions like the appropriate limits of government or the bounds of the discourse about those limits in other contexts. Yet I also generate novel evidence supporting the possible realization of the ideal of the neutral administrative state.

My answer to the question of whether decision-making technologies really are neutral is based on diverse forms of evidence that I evaluate using multiple methods.

These methods include experimental and quasi-experimental quantitative methods, both of which rely heavily on diverse sources survey data, and qualitative historical analysis including synthesis of interviews and previously unreleased government documents from multiple federal agencies that I obtained under the Freedom of Information Act. By triangulating distinct sources of data and methods of analysis including original survey, survey experimental, and psychophysiology study evidence, I generate reliable and valid estimates of some forms of bias – and often of surprisingly robust neutrality – in technology-mediated administrative decisions. By combining the insights these analyses generate with qualitative historical insights, I build a fresh account of the political development of polygraph programs and the surveillance state.

This dissertation follows an inverted pyramid structure, starting with the highest level of analysis and ending with the lowest before returning to the original theoretical moorings to conclude. Accordingly, Chapters 2 and 3 focus on the construction of state power and social control through technology at the highest level of analysis (i.e., the federal legal regime of the surveillance state), Chapter 4 quantifies the political institutional-level effects of technology at the next-lowest level of analysis (i.e., the effects of police polygraph and other screening tools on departmental diversity and brutality), Chapter 5 tests for potential racial and confirmation bias at the behavioral decision-making level with street-level bureaucrats as the subjects of interest (across multiple public administrative technologies including polygraphy), and Chapter 6 assesses potential attitudinal effects of the perception of technology-mediated administrative neutrality among citizenry at the lowest level of analysis.

As Chapters 2 and 3 explore in greater detail, lie detection programs are common, costly, and chronic. Millions of polygraph exams are administered annually in the U.S., the federal government exports polygraph equipment and expertise to allies as part of anti-corruption programs, and millions more people have been subjected to lie detection screenings using next-generation polygraphs (i.e., wireless versions of psychophysiological and behavioral deception detection protocols) at busy airports, border crossings, and public spaces. Including the latter travel security screening lie detection programs, lie detection is about a \$4 billion/year industry within the U.S. alone. Excluding them, polygraphy is about a \$3 billion/year industry by my conservative estimate. Either way, the vast majority of associated expenses are paid by taxpayers. The recent growth of public sector polygraph programs has been tremendous and sustained – despite insufficient scientific evidence supporting the efficacy of these programs to detect or deter crime or misconduct (Iacono and Lykken 1997; Iacono 2001; National Research Council 2003; Lykken 1998).

However, asking why interventions without scientific basis persist and often grow in widespread use is the wrong question to ask about any public policy intervention. Social and political interventions are not evidence-based as a general rule (Campbell 1998; T. D. Wilson 2011). So the fact that polygraphs are not evidence-based does not by itself make polygraph programs, despite their widespread use and remarkable growth pattern, exceptional. What makes them exceptional is the convergence of watershed moments in the development and growth of polygraph programs with major turning points in the political development of the surveillance state. Investigating the conceptual shifts that characterize these junctures sheds novel light on the evolution of concepts and

practices surrounding core political concerns: changes in the understanding of how we construct evidence and proof for the public validation of truth, in the culture of political institutions and how they define coercion, and in the limits of legitimate government – changes in truth, lies, and polygraph tape.

Chapter 4 addresses the follow-up questions: Do more or less technology-mediated administrative decisions lead to more or less diversity at the political institutional level? And do they lead to better or worse hiring decisions that generate better or worse institutional outcomes? I address these questions by examining effects of supposedly neutral screening techniques including polygraph screenings on diversity and brutality in nationally representative state and local law enforcement agencies. Local and state police departments nation-wide have recently increased their use of all widely available pre-employment screening tools as part of efforts to create more highly skilled, professional workforces. But scholars have not yet established how these tools affect departmental diversity and brutality. Police diversity matters, because representativeness of political institutions is a measure of equal opportunity. Police brutality matters both as an extreme abuse of public trust, and as a proxy measure of broader police misconduct. Thus these measures, which matter normatively on their own, also matter as indicators of how effective (or in some cases, counter-productive) attempts to professionalize law enforcement agencies through apparently objective screening tools are.

This chapter is based primarily on data from the Law Enforcement Management and Administrative Statistics Survey (LEMAS), which collected nationally-representative data on state and local law enforcement agencies from 1997 to 2007. First, I document the steady increase in law enforcement agency use of many pre-employment screening

tools over the study period. Then I demonstrate how these tools affect department diversity and brutality. Using matching in combination with difference-in-differences regression analysis (DID), I compare similar police departments that enacted and did not enact these pre-employment screening tools to estimate the effects of these tools on departmental diversity and brutality. The combination of this matching technique with DID produces more accurate and less model-dependent estimates than traditional, commonly used modeling approaches, although I adopt an agnostic attitude to the power of matching to enable causal inferences from observational data. The analysis proceeds in two sections.

Specifically, I estimate the causal effects of screening tools on police diversity in terms of race and gender. I find that my previously estimated null racial bias effects from polygraphy generalize to the use of polygraphy to screen police recruits in the field. This suggests that results from survey experiments presented in Chapter 5 have good ecological validity. More broadly, my findings in this analysis suggest that some tools harm departmental diversity and others decrease citizen complaints of excessive officer use of force. However, no tool affects both total complaints and complaints that are found after investigation to be sustained. This might lead one to question the utility of these screening tools. These analyses control for agency type, size of population served, region, whether the department authorizes collective bargaining, and whether the department has an internal affairs unit, and are robust to a range of matching specifications. This chapter contributes to our understanding of how widely implemented but empirically unproven public policies – policies that look like they simply apply neutral tools and technologies

to solve problems – can influence the composition and character of political institutions and actors, and their citizen interactions.

Chapter 5 present results from a series of six survey experiments testing for possible racial, confirmation, and intersectional (racial plus confirmation) bias in technology-mediated decisions. Chapter 6 follows up on these results by presenting results from an additional survey experiment testing for possible attitudinal effects of these results with respect to egalitarianism. At the behavioral level, my findings suggest negative background information can be systematically used as a decision-making heuristic in the interpretation of ambiguous stimuli in the context of technology-mediated decisions. This means that technologies that seem to produce independent evidence can instead sometimes merely dress up pre-existing beliefs as science. This confirmation bias result holds in polygraph chart interpretation across four survey experiments under varying experimental conditions, including conditions of time pressure, emotion, and threat. The fourth survey experiment on polygraph chart interpretation clarifies this result by establishing the following scope condition: given a more rational way to apply the negative background information heuristic (i.e., the ability to opt to set the polygraph to “suspicious mode”), people tend to take that rational choice – and then make unbiased interpretations themselves.

Across all four polygraph experiments, race is not used as a heuristic in interpreting the ambiguous chart, even when it is combined with negative or positive background information that cues racial stereotypes. The fifth and sixth survey experiments replicate these null racial bias results, extending the tests of racial, confirmation, and intersectional bias in technology-mediated decisions to different

technologies. Those technologies are a medical diagnosis tool that uses symptoms entered along with patient demographic information to generate a probability-ranked differential diagnosis, and a food stamp calculator that uses client information to generate a benefits eligibility estimate. Results do not support racial, confirmation, or intersectional bias in these technologies. At the behavioral level, then, my results show confirmation bias is possible but not necessarily endemic to technology-mediated decisions, contrary to what one might theorize on the basis of the vast majority of published research on attribution bias, prejudice, and automatic processing. In fact, evidence does not support a racial bias effect in these decisions at all.

How does this overarching pattern of technology-mediated administrative neutrality – a form of procedural fairness – affect people’s attitudes about the justness of the world, race, and distributional fairness? At the attitudinal level, my findings suggest that micro-level neutrality does not legitimate or otherwise appear to affect egalitarianism or racial attitudes. Overall, then, “thinking technologies” can help bureaucrats make fair, unprejudiced decisions with respect to race. And they can do this without causing an increase in inegalitarian attitudes by promoting an ideology of neutrality that says procedural fairness legitimates inequality. Chapters 5 and 6 contribute, then, to our understanding of administrative neutrality, race, and fairness from the perspectives of both street-level bureaucrats and citizens. Findings from these seven survey experiments suggest that far from invisibly institutionalizing bias, technology-mediated decisions tend to be neutral with respect to race, sometimes with respect to background information, and generally with respect to the intersection of race and background information.

In a concluding essay (Chapter 7), I summarize this dissertation’s contributions and significance, and suggest directions for future research. The novel evidence presented in Chapter 5 shows that decision-making technologies can truly help street-level bureaucrats refrain from institutionalizing what many argue are pervasive biases against racial minorities and particularly vulnerable subsets of these minority groups. Chapter 6 establishes that this neutrality does not subsequently have inegalitarian attitudinal effects by rationalizing distributive injustice as the outcome of neutral processes. Against post-structuralist criticisms of the Enlightenment project, this evidence supports – within limits – the positivist conception of a functionally objective state apparatus. But the additional evidence explored in Chapter 4 shows these technologies are also often employed on a large scale without sound evaluation of their efficacy or effects – and in this way, they may harm political institutional diversity and even have serious, counter-intuitive effects. Thus rational-legal administration in accord with egalitarian norms may be an ideal-typical and practical norm, but ideology and error can still shape administration. Indeed, as Chapters 2 and 3 suggest, at the meta level, administrative decision-making technologies embody a reliance on neutral means – in service of deeply political ends – that can make value-laden processes seem apolitical. This appearance of neutrality is thus an ideological one that reflects the construction of state power and social control through information technology. Future research might evaluate how policing policies such as use of lie detection screenings affect people’s trust, well-being, and use of public services.

Truth, Lies, and Polygraph Tape:
Earlier Development of the Contemporary Surveillance State

“It is important to create a picture of objectivity, capability, being in charge, and above all, having the ability to ascertain if the examinee is responding truthfully or deceptively”
– Stan Abrams, *The Complete Polygraph Handbook*, 1989.

Introduction

This dissertation presents a sweeping analysis of administrative technology as an evolving way of understanding expertise and the Enlightenment fact-value distinction that has tremendous implications for politics from the construction of state power to the psychology of street-level bureaucrats. These first two empirical chapters set up that puzzle and address it at the broadest possible level of analysis by examining the development of the federal legal regime surrounding polygraph programs and the surveillance state. Subsequent empirical chapters descend down levels of analysis one step at a time, covering next the institutional-level effects of screening tools including polygraph tests (Chapter 4), then how some common cognitive biases might affect the behavioral decision-making of street-level bureaucrats making administrative decisions with the help of technologies including polygraphs (Chapter 5), and finally addressing the question of whether perceived neutrality of administrative decision-making technologies influences attitudes about racial and socio-economic inequalities in particular and fairness in general (Chapter 6).

The subset of the over-arching puzzle this chapter addresses is how the realm of what we tend to see as primarily subjective versus objective evolved from the Progressive Era to the present. This dissertation argues that evolution is expressed in terms of shifting constructions of state power and social control. This chapter identifies and explores four

political moments of conceptual shift that move where we as a society draw the line between facts and values. This shifting fact-value distinction has fundamentally changed political institutions and society by changing how we think about truth.

The fact-value distinction is a boundary between facts and opinions. It traces its lineage through ancient Greek, Enlightenment, and Progressive eras of moral-political thought that are central to contemporary Western politics. As a duality, the fact-value distinction maps onto other classical dualities, including male-female and public-private. In this network of associations, politics (gendered male) should be the realm of reason, not emotion. Yet, whether we think of the boundary as a line, a continuum, overlapping spheres, or something else, the very idea that the objective and subjective realms are separable is a contested one.

Research increasingly suggests emotion and reason are not truly separate. For instance, emotional appeals primarily account for the efficacy of political campaigns, especially among politically informed citizens (Brader 2006). Individuals with brain injuries that damage emotional signal processing but not most cognitive functions have impaired decision-making, suggesting emotional signals underpin what we tend to think of as rational decision-making abilities (Bechara 2004). Some research suggests emotion operates at the neurobiological level to often optimize, rather than denigrating the rationality of, economic choice (Seymour and Dolan 2008). It seems emotion and reason are twin cognitive powers that work together to make human beings the distinctive social and political animals we are. This revelation echoes and in some ways substantiates post-structuralist criticisms that the fact-value distinction is socially constructed in ways that

privilege the powerful and reproduce inequalities by denigrating the value of more popular or personal truth.

In this two-part qualitative historical chapter, I argue that landmark efforts to protect individual rights and limit government power – although they succeeded in important ways – also created a moral-political landscape that eroded the parallel fact-value and public-private boundaries gradually farther and farther until the external shock of 9/11 turned that erosion into a landslide. Four landmarks mark the path of the developing federal legal regime surrounding lie detection. These landmarks among others delineate important conceptual shifts that illustrate and contribute to this key boundary erosion. As a result, they are also major turning points in the development of the contemporary surveillance state.

Chapter 2 explores the first pair of these four turning points, and then Chapter 3 explores the second pair. The first in this first pair is the Progressive Era precedent *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923). *Frye* established the general inadmissibility of polygraph test results as evidence (O’Donnell 2007). This precedent was part of a broader response to Progressive Era reform pressures and backlash against widespread police use of torture. In *Frye*, the courts at once enacted the Progressive professionalization impulse by setting a legal standard for expert testimony – and checked public reform movements by institutionalizing a conservative conception of expertise. The *Frye* standard privileges generally accepted scientific knowledge as expertise over other types of expertise: methodologically sound but new knowledge, and ordinary people’s experiences.

This standard of expertise reflects a shift in how we construct truth. On one hand, it limited the discretionary power of fact-finding agents including police and judges to effectively decide on a case-to-case basis what constituted fact. On the other hand, it limited the ways in which scientists with new ideas and ordinary people with first-hand knowledge could present their opinions as expertise in court. The *Frye* standard for admissibility of scientific evidence and expert testimony held for at least seventy years, when *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993), sometimes called “Frye in drag” because of its similarity to the Frye standard, established a new standard that supersedes or qualifies Frye in many jurisdictions (Rice 2009). As a landmark of the Progressive Era legal development of the fact-value distinction, *Frye* exemplified greater efforts to professionalize, standardize, and quantify policing and other aspects of administration more broadly. But it did so without actually establishing effective oversight or evidence-based social policy or practice, so that this appearance was often just that.

Second, *Miranda v. Arizona*, 384 U.S. 436 (1966) established a new interpretation of Fifth Amendment rights as part of the Warren Court’s due process revolution (O’Brien 2008). In the due process revolution, the courts nationalized (some) substantive due process standards. *Miranda* nationalized the standard for judging what counts as a coerced confession and why it has to be excluded at trial. Like Progressive Era backlash against police torture, this standardizing response to what were perceived as coercive policing practices had some unintentional consequences. Namely, in addition to adding standardized protections, *Miranda* also created a chain of backlash and retrenchment that generated mounting structural incentives for police to lie to and otherwise

psychologically coerce suspects. The distinction between so-called “real” and “testimonial” is illustrative.

Polygraphs sit at the nexus of this constructed dichotomy. Polygraph charts, along with blood, voice analysis, handwriting analysis, medical and other records, are considered “real” evidence, because they record facts in the form of physiological responses. Ironically, “real” evidence in the case of polygraphs is inadmissible because it lacks sufficient evidence base for the scientific community to consider polygraph charts *valid* as indicators of deception. Lie detectors do not “really” detect lies. Statements made during polygraph interrogation, by contrast, are generally admissible as “testimonial” evidence.

The Second Reconstruction Era of which the Warren Court was part had myriad phenomenally positive social impacts, such as making formal equal opportunity on the basis of race, religion, and nationality into federal law. But as a set of social pressures, it also presupposed the fact-value distinction in general and the evidentiary basis of pseudoscientific practices building on that distinction in particular. This was a significant strategic error on the part of civil rights and liberties advocates. *Miranda* exemplifies this strategic error in its institutionalization of the real versus testimonial evidentiary distinction. The Fifth Amendment literally protects citizens from being compelled in any criminal case to be witnesses against themselves. The real-testimonial evidence distinction constrained the bounds of what that protection against self-incrimination means. In this way, it expanded governmental power under the auspices of science.

Chapter 3 explores the next second pair of these four landmarks on the path of the developing federal legal regime surrounding lie detection, that also delineate important

conceptual shifts in the moving boundary of the fact-value distinction. Overall, one of the counter-intuitive features of the development of this regime is the disconnect between the growth of polygraph programs and repeated institutionalization of powerful checks on these programs. As I explore in greater detail later in Chapter 3, polygraph results are generally inadmissible in court, forbidden in private sector employment screenings, and regarded by the contemporary scientific community as having insufficient evidence to justify their use in precisely the security-focused contexts in which they have experienced steady and remarkable growth for decades. One plausible explanation for this curious disconnect is that legal constraints on lie detection, while advancing substantive due process in some important ways, have also created a federal legal regime that shields public sector polygraph programs from public scrutiny, market pressures, and other forms of external oversight. The checks have therefore proven both constraint and shelter.

Progressivism, *Frye*, and the Professionalization of (Some) Expert Testimony

The Progressive reform movement (1890s-1930s) brought a renaissance of popular and political belief and action predicated on the intertwined Enlightenment ideals of scientific progress, efficiency, linear social and economic development, and the systematic amelioration of corruption. This Progressive pendulum swing gave rise first to Prohibition, and then to backlash against it. Some of that backlash was progressive in the sense of pushing Progressive ideals further, and some was conservative in the sense of counter-balancing them. The progressive backlash came from the executive branch in the form of a police torture scandal and its reverberations, and the conservative counter-balancing came from the judiciary. This matters in the context of the political

development of polygraphy because scientific lie detection techniques both partly replaced police use of torture in response to that progressive backlash against Prohibition’s unintended, corrupt consequences – and failed to set a truly scientific standard for the admissibility of expert testimony in criminal courts, in the polygraph-constraining precedent *Frye v. United States* (1923).

Briefly, the Eighteenth Amendment’s Constitutional prohibition of alcohol became effective in 1920 and resulted in a thriving black market for booze, increasing both crime (de jure) and corrupt law enforcement (de facto), as police sought to make money from the trade in some cases and enforce the law – albeit often lawlessly – in others.

Lawless law enforcement in the U.S. was hardly new. For instance, during World War I, for example, American military prisons routinely “high-cuffed” deserters, insubordinates, and conscientious objectors, forcing prisoners to stand all day with their hands cuffed high above their heads (Rejali 2007). Conscientious objectors were also subjected to ice-water baths until losing consciousness. Although the public was generally unsympathetic to these groups, there was public outcry over these practices, and the War Department subsequently stopped authorized use of high-cuffing.

When lawless enforcement of the law crescendoed again in police response to the booming black market that developed in response to Prohibition, the public outcry was again resounding. In response, in 1929 President Hoover established the National Commission on Law Observance and Enforcement (a.k.a. the Wickersham Commission – headed by former Attorney General George W. Wickersham), to assess causes of crime and how to reduce it. The commission produced several reports, most notably the 1931

Report on Lawlessness in Law Enforcement, which spurred police reforms by systematically investigating misconduct in American justice for the first time. This report documents and advances the early history of the struggle to curb police misconduct that eventually generated the Warren Court due process revolution, bringing the issue from the local to the national level of political discourse for the first time in U.S. history.

Most of this misconduct had to do with Prohibition, its widespread violation, and aggressive and extensive police efforts nation-wide police address the apparently resulting crime wave by coercing confessions. Remarkably, Wickersham Commission consultants observed police interrogations directly to make these observations. One might well wonder what interrogation tactics they did not observe, since torture – no matter how widespread – is usually neither found documented as a written state policy nor observed first-hand by administrative investigators. Their evidence also involved detailed interviews with participants and observers of “the third degree,” such as current and former police officials, in addition to judges, prosecutors and defense attorneys, law professions, and journalists from fifteen cities nation-wide. The third degree included not only traditional methods of torture that can leave marks, such as beating and burning, but also lengthy detention, threats, and other techniques of psychological coercion including pseudoscientific lie detection practices. The poor, racial minorities, and political agitators such as labor union activists disproportionately bore the burden of this widespread police misconduct.

Amid the scandal of widespread police use of torture to extract confessions, combined with the failure of law enforcement to address violent crime by catching perpetrators, American police feared that judges would take over their fact-finding

mission. This model of investigative justice is common in other countries, and it would leverage the strengths of the existing local and state justice infrastructure in the U.S. to increase accountability in general rather than introducing new structures designed to enhance police accountability specifically. Thus Progressive reform pressure to decrease police brutality following the Wickersham Commission generated a process of accelerated police professionalization. As a result of the report and ensuing public outcry, police departments began to create formal internal affairs units addressing alleged police misconduct (Walker, Boehm, and Hall 1997). They also began favoring the use of “clean torture” (coercive interrogation, interview, and detention practices that do not leave marks), including “lie detection” interrogation techniques, over previously widespread techniques such as beatings (Leo 2008). But it turns out the police did not have to worry, because the judiciary had other plans.

Progressive Era courts and the U.S. Supreme Court (SCOTUS) in particular did take on new police powers in the sense of upholding regulations intended to protect the health, safety, or morals of the community. But SCOTUS also tended to see their role as checking and balancing Progressive reform pressures in the economic realm, pushing the pendulum hard in the other direction. The most famous example of this trend is *Lochner v. New York*, 198 U.S. 45 (1905), in which SCOTUS struck down a maximum hour law as interfering with freedom of contract – instantiating a relatively new, substantive rather than procedural conception of what due process means that would change in 1937 and take on new meaning again under the Warren Court. In 1937, the court’s conception of due process changed with *West Coast Hotel Co. v. Parrish*, 300 U.S. 379 (1937) upholding the constitutionality of an economic regulation limiting freedom of contract –

the so-called “switch in time that saved nine” and averted FDR’s court-packing plan. Under the Warren Court, substantive due process completed its pendulum swing from a conservative concept protecting freedom of contract for individuals to work as many hours as they and their employers mutually agreed, to a progressive concept protecting individual rights to talk with police. Thus *Lochner* and the judicial social Darwinism it exemplifies actually expanded the power of the courts in a lasting way by reading substance into the due process clause, rather than effectively limiting the power of the courts to advance communal health, safety, and welfare when it interfered with the right of the individual to strike out on his own, or temporarily expanding judicial power relative to legislatures before 1937.

At the same time, the judiciary’s Progressive Era backlash was not purely economic. This era was a time of social challenges to an inegalitarian status quo in many other ways, and the courts responded to all of them with some degree of conservative backlash. In 1922, racial prejudice at Columbia University drove young Langston Hughes to drop out, embracing his calling as a poet in the thick of the Harlem Renaissance (Rampersad 1986, p. 56). Alderman Peter J. McGuinness of the Greenpoint section of Brooklyn sought an ordinance forbidding women from smoking (“ ‘sucking cigarettes’ ”) in public, because “ ‘The young fellows lose all respect for women and... vampired by these smoking women, desert their homes, their wives and children, rob their employers and even commit murder so that they can get money to lavish on these smoking women’ ” (Dismore 2011). And a newly minted Harvard Ph.D., supposedly inspired by one of his two simultaneous wives and professional collaborators’ observation that “ ‘When she got mad or excited, her blood pressure seemed to climb,’ ” attempted to use a

novel “lie detector” test to exonerate a black man who had been convicted of murder (Lamb 2001; O’Donnell 2007). The court denied the admissibility of scholar’s testimony in a precedent that has since shaped legal and political conceptions of scientific expertise and its relevance to public life. That scholar was William Moulton Marston, and he was one of the inventors of the modern-day polygraph.

That *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923), was a racialized case may or may not have driven its outcome. Regardless, it set a new standard for the admissibility of expert testimony, institutionalizing a new conception of expertise that is neither inegalitarian nor particularly scientific. It is a quintessentially conservative conception in the Burkean sense of relying on social mores in the form of ingroup acceptance of the thing as the key metric of value. The *Frye* standard for admissibility of expert scientific testimony established the inadmissibility of polygraph test results as scientific evidence because they were not sufficiently established to have gained “general acceptance” in their field. This standard still applies in many jurisdictions, although in others it has been replaced by a similar standard (Daubert or hybrid Frye-Daubert).¹

Frye thus in some sense bolstered Progressive Era reform efforts by professionalizing and standardizing bureaucratic decision-making about whether or not to allow expert scientific testimony. But it simultaneously checked the Progressive reform pendulum swing by institutionalizing a conservative conception of expertise. The general

¹ The *Daubert* standard applies to all expert witness testimony in federal courts, and allows for exclusion of unqualified evidence. Its language stems primarily from *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993), which said the general acceptance test *Frye* institutionalized was not incorporated into Rule 702 of the Federal Rules of Evidence. It replaced general acceptance with a flexible reliability standard, under which judges can exclude expert testimony that does not adequately relate the evidence and the expert’s conclusion. Building on *Daubert*, *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999), applied the flexible reliability standard to all expert testimony.

acceptance test privileges generally accepted scientific knowledge over both methodologically sound but new knowledge, and more pedestrian forms of expertise.

A few examples of what would be formally admissible as expert testimony on scientific evidence if we relied on this standard in its pure form serve to show the standard’s most obvious flaw: it does not define who the relevant group of experts is. A recent survey of head and neck surgeons (N=647, response rate \approx 6%) found that 27 percent believe Obamacare includes a provision for a “ ‘government panel [that] makes end-of-life care decisions for Medicare (Rocke et al. 2013)’ ” In other words, a sizeable minority of doctors appear to believe in death panels. Under *Frye*, those doctors might be considered the relevant experts on whether testimony relating to the existence of death panels should be admissible in court. Similarly, police investigators and others with apparently relevant training and expertise tend to perform worse than college students and other uninitiates at identifying liars versus truth-tellers in contexts including confessions (Meissner and Kassin 2002; Kassin and Fong 1999; Kassin, Meissner, and Norwick 2005). The general acceptance test does not specify whether the relevant group of experts in the context of lie detection are the police interrogators or polygraphers who practice it as professionals, or the psychologists who almost universally agree deception detection lacks sufficient scientific basis. Thus while *Frye* laid out a standard for what counts as expertise in how evidence and proof are constructed in courts of law, it also institutionalized a wide degree of discretionary power for fact-finding agents to decide what this standard means in practice.

Another problem in this conception of expertise that the lie detection example underscores is how it privileges experts over pedestrians, when sometimes regular people

make better decisions. In the examples of Kassir’s lie detection studies, we know that sometimes experts make systematic errors that civilians do not make. Far from being limited to the realm of lie detection, this is a much broader pattern: expertise has significant limits to both diagnose and predict. For instance, most political and economic forecasts over a few decades were neither much better than chance nor much better than projections of casually informed pedestrians (Tetlock 2006). Privileging expert testimony over regular people’s absent an established scientific basis thus not necessarily buy courts much in terms of accuracy.

In addition to producing questionable accuracy gains, this conception of expertise is subject to criticisms that it is anti-democratic. Scholars working in the tradition of the new political sociology of science tend to suggest that expertise must conform to norms of transparency and deliberative adequacy, and that consensus in the form of *Frye*’s general acceptance test lacks the procedural specificity to conform to this requirement (Guston 2006; Jasanoff 1995; Jasanoff 2003). This overarching concern contains several potential criticisms with deep echoes throughout American political thought, the most important of which in the context of the political development of the modern surveillance state is the Madisonian worry that the courts are supposed to constrain the public when it persecutes minorities – not when special interests claim to know better. In this way, standards of admissibility of expert testimony that do not rely on procedural fairness (or scientific basis itself) might make courts and legislatures more vulnerable to special interest capture, including simple deference to executive branch power and secrecy.

This weakly democratic evidentiary standard thus foreshadows executive capture of the formal judicial branch check on surveillance in the form of the U.S. Foreign

Intelligence Surveillance (FISA) Court. Congress created this parallel judiciary during the backlash against warrantless domestic surveillance stemming from the U.S. Senate’s Church Committee (D. B. Cohen and Wells 2004). Ironically in light of its origins, under the Foreign Intelligence Surveillance Act of 1978, the FISA court considers warrants for surveillance of suspected foreign agents in the U.S. by hearing arguments from one side – the government’s – and seldom making its findings public (Lichtblau 2013). It is a FISA court that, in a top-secret order Edward Snowden leaked to the press, issued an order requiring a Verizon subsidiary to give the N.S.A. an ongoing daily feed of call detail records including those for domestic calls to and from millions of regular Americans (Greenwald 2013). The court order forbids Verizon from disclosing the government’s request or the order granting it.

The procedural and substantive weakness of the evidentiary standard set forth in *Frye* also foreshadows executive and special interest groups’ capture of legislative checks on surveillance in the form of the Employee Polygraph Protection Act (EPPA). As discussed in greater detail in the section after next, EPPA significantly constrained lie detection programs in the private sector. But it also carved out several exemptions, including the entire public sector and several private sector groups, creating large classes of second-class citizens whose due process rights are now considered secondary to state and community interests.

Overall, then, *Frye* was a deeply conservative and unscientific turning point in policing and the development of national evidentiary standards. It circumscribed the legal power of lie detection, but left untouched police power to psychologically coerce confessions without having them considered a priori coercive. It institutionalized a

nebulous and lasting gray area between what courts consider scientific and unscientific expert testimony. This gray area reproduces power by privileging expert over non-expert narratives in often nontransparent, unaccountable ways.

Frye also foreshadowed the trend of replacing what the Wickersham Commission identified as lawless police practices (i.e., torture) with what the National Academic of Sciences reports on polygraphy and forensics have identified as pseudoscience – in spite of the exclusion of polygraph test evidence in *Frye*. This trend is politically significant because it involved the development of several closely related police interrogation techniques. These techniques are nontransparent processes of supposedly scientific but actually insufficiently evidence-based truth verification, intended to maximize convictions while improving the public image of police. They include behavioral lie detection, statement analysis, voice stress tests, polygraphy, and the Reid interview and interrogation technique. The Reid technique is used across a wide variety of federal, local, state, and international law enforcement agencies, both as part of polygraph test protocols and as part of normal interrogations. Polygrapher Fred Inbau, director of the Chicago Police Scientific Crime Detection Laboratory, taught polygraph interrogation technique to John A. Reid, who went on to establish his own polygraph practice and co-publish on interrogation with Inbau (Zulawski 2002). Although extremely widely used – John E. Reid & Associates Inc. trains more interrogators than any company globally, most North American police use the technique, and an accidentally leaked FBI interrogation manual confirms federal law enforcement continue to use it at the highest levels – the Reid technique is associated with false confessions, and leading psychologists consider it inherently coercive (Starr 2013; Kassin, Appleby, and Perillo

2010; Baumann 2013). The Reid technique incorporates behavioral lie detection and judgments of deception, exemplifying the widespread use of deception detection in American justice in spite of its apparent constraint in *Frye*.

Broadly, this professionalization of police interrogation that was not really professional, and standardization of admissibility of scientific evidence that was not really standard or scientific at all, set the stage for the logical next progression of the development of the modern surveillance state in the form of the Warren Court due process revolution as exemplified by *Miranda*. *Miranda* changes the conception of what process is due, further nationalizing the substantive due process standard created during Progressive Era judicial backlash in *Lochner*. It further professionalizes proof as something that must not be coerced to be valid and admissible, and coercion as something that can be psychological as well as physical.

Frye is thus significant in the development of the modern surveillance state because it institutionalized an apparently powerful check on polygraphy that does not actually function as much of a constraint in the bigger picture. Its general acceptance test marks one of four watershed moments in the development and growth of polygraph programs, all of which share this peculiar quality of dysfunctional constraint. In different ways, each of these moments also shares another key feature of *Frye*: the institutionalization of a two-tiered approach to due process, in which the collection of evidence (for various purposes) is artificially separated from its admissibility in court. *Frye* thus signaled that instead of judges overtaking the police role in fact-finding, as police feared they would during Progressive Era outrage over widespread police use of torture, fact-finding in the legal process would continue to take place within autonomous

bureaucracies – raising continuing concerns about democratic accountability and questions about what circumstances trigger due process protections for what classes of people.

Interrogating Rights: The Warren Court Due Process Revolution, *Miranda*, and the Nationalization of (Some) Due Process Standards

The Warren Court (1953-1969) enshrined as Constitutional rights many of the civil rights and liberties we take for granted today. In particular, during the Warren Court due process revolution (1963-1966), a liberal majority SCOTUS headed by Chief Justice Earl Warren interpreted the Fourteenth Amendment as incorporating the first eight amendments in the Bill of Rights so that they applied not only to the federal government but also directly to the states. In cases a series of decisions about procedural fairness in criminal justice, the Warren Court nationalized protections for accused criminals. These protections prohibited prosecutors’ use of evidence seized in illegal searches (*Mapp v. Ohio*, 367 U.S. 643 (1961)), guaranteed indigent defendants’ access to counsel (*Gideon v. Wainwright*, 372 U.S. 335 (1963)), and mandated that police clearly explain to suspects in their custody that they have certain rights, such as the right to an attorney (*Miranda v. Arizona*, 384 U.S. 436 (1966)). Critics charge the Warren Court with judicial activism and “handcuffing the cops” (Cassell and Fowles 1998). At the same time, the Court generally read due process narrowly, notably limiting rather than expanding the substance of key due process rights. *Miranda* exemplifies this constraint.

But like Progressive Era reform movements, the Warren pendulum swing wrought both unintended consequences and significant backlash that drove further growth of lie detection programs in particular and the surveillance state in general. Its

unintended consequences include an expansion of federal law enforcement powers. Its backlash includes a relatively immediate Congressional act (the federal Omnibus Crime Control and Safe Streets Act of 1968, statute 18 U.S.C. § 3501, establishing admissibility of criminal defendant’s voluntary statements in federal courts regardless of Miranda warnings) and the piecemeal development of a wide variety of exceptions enshrined in ensuing court cases, generating structural incentives for police to lie to and otherwise psychologically coerce criminal suspects, and raising questions about the continuing practical relevance of the Miranda warning that has grown so ubiquitous in popular culture the Rehnquist Court refused to expressly overturn it despite its reduced impact on law enforcement practices in light of subsequent rulings (*Dickerson v. United States*, 530 U.S. 428 (2000)).

In this way, the legislative and judicial response to *Miranda* was overwhelmingly counterbalancing – despite the fact that the due process revolution as a whole actually constrained the same Constitutional rights it incorporated. Moreover, contention surrounding these due process cases has proved more widespread and lasting than public outcry over desegregation or reapportionment. Chief Justice Warren considered these cases protections of Constitutional rights, while opponents (including current Justice Scalia) charge SCOTUS with reading rights into the Constitution.

Miranda, and its socio-historical moment and backlash, are politically significant in the context of the development of the surveillance state because they reflect a focus on what counts as interrogation, how we define coercion, and what process is due under the law in relation to both. Laying out a standard addressing these questions limited the (again still enormous) discretionary power of law enforcement to decide on a case-by-

case basis what due process was due to criminal suspects. But the subsequent chain of exceptions has arguably made the standard pro forma in two ways. First, it does not apply to some contexts or classes of people. Second, the chain of exceptions it generated structurally incentivized the use of deception and psychological coercion by police.

Appropriately enough, *Miranda* is a story about lie detection from start to finish. In 1963, an eighteen-year-old intellectually disabled woman was abducted while walking home from work, bound in the back of a stranger’s car, and raped at a secluded location (Sonneborn 2004; Van Meter 2007). Police doubted the woman’s story and submitted her to a lie detector test, with inconclusive results. Because her attacker had not injured her severely, police continued to doubt her story and did not proceed with an investigation.

But the woman had kept working, taking the same bus route and walking the same path as before. She was too afraid now to walk home alone, so her brother-in-law was walking her home. He saw a car that resembled the one “Lois” (the victim’s pseudonym) had reported being abducted in, driving slowly near her bus stop, and wrote down part of the tag. Police, again, did not obtain this easily available information themselves, because they were not actively investigating the rape complaint since they doubted the veracity of the report.

Even with this new lead, had Lois failed her polygraph, police might not have investigated her case any further – in fact, they might have charged her with the crime of making a false report. This counter-factual underscores a latent conservatism in the apparently liberal Warren Court due process revolution: it institutionalized protections for criminal suspects in narrow circumstances under which they are already criminal suspects. It did not extend these protections to ordinary people, private employees under

possible employer surveillance or questioning, or victims and witnesses – even though all these groups of people may swiftly become criminal suspects in the course of a conversation with police, an administrative investigation at work, or an attempt to report a crime.

But police forgave Lois her inconclusive polygraph, and followed up on the lead her brother-in-law had produced for them. The car belonged to Ernesto Miranda’s live-in girlfriend. Miranda had a history of attempted rape and resembled the alleged attacker, and his paint-smelling old, green car with brown upholstery and a rope attached to the inside was also consistent with the victim’s description of the car in which she was allegedly abducted. Police took Miranda into custody and interrogated him without first warning him that he had due process protections under the Bill of Rights, especially the right against compelled self-incrimination, the right to silence, and the right to counsel including a court-appointed lawyer if he could not afford an attorney. Miranda was tried and convicted of kidnapping, rape, and (in a separate complaint) armed robbery. On appeal, SCOTUS ruled police had violated Miranda’s Fifth and Sixth Amendment due process rights, and that his subsequent confessions could not be admitted at criminal trial. Miranda was subsequently retried and reconvicted with his confessions excluded from evidence.

Although it is often considered the keystone case in the Warren Court due process revolution – and, by critics, an excessively liberal constraint on law enforcement – *Miranda* actually constrained due process in three ways that contributed to increased use of lie detection and that would eventually underpin the modern surveillance state. First, *Miranda* incorporates Constitutional rights to due process, but it does so through a

narrow reading of the Fifth Amendment relative to what a literal interpretation would imply. That is, it guarantees a privilege against self-accusation, rather than a right against self-incrimination. This privilege to not confess, which is much narrower than a right to not incriminate oneself in other ways (e.g., through one’s records, phone conversations, blood, fingerprints, or other “real” as opposed to “testimonial” evidence) has obvious import with respect to the surveillance state, particularly in combination with the second way in which *Miranda* constrained due process rights. In this way, even as *Miranda* nationalizes a new conception of illegal coercion as being potentially psychological as well as physical, it also introduces a distinction between real and testimonial evidence that constrains what it means under the Constitution to incriminate oneself.

Second, *Miranda* institutionalizes a national standard for criminal defendants’ protections, but it simultaneously constrains the classes of people and conditions under which these protections apply. The Fifth Amendment contains some broad language about people and their due process rights in relation to life, liberty, and property. It is not limited entirely to criminal contexts, explicitly encompassing also, for example, the just compensation clause. Yet in the Warren Court’s avowedly radical reading, the scope of these due process protections constricts from all people to criminal defendants. This excludes regular people, foreshadowing massive warrantless wiretapping under the surveillance state post 9/11. It also excludes employees in administrative investigations (who may currently be asked to take polygraphs), and job candidates (who were routinely asked to pass lie detector tests before the Employee Polygraph Protection Act of 1988). In addition, it also excludes victims and witnesses in criminal cases. This means, as in *Lois*’s case, alleged victims and witnesses who report crimes can be interrogated without

being Mirandized, and anything they say can be used against them if police decide – lie detector or no – that they are lying. This fluidity of the category of criminal defendant underscores one reason why due process rights might otherwise be applied equally to alleged victims/witnesses and defendants in an adversarial criminal justice system. In addition to being limited to a narrower class of people than due process protections might have otherwise been applied, then, *Miranda*’s applicability is also limited in terms of context. For example, the criminal defendant must be in custody, being interrogated by a state agent.

Third, *Miranda* also constrained due process by generating significant and sustained conservative backlash across all branches of government. Congress immediately worked to pass the federal Omnibus Crime Control and Safe Streets Act of 1968, purporting to overrule *Miranda* in federal criminal court, restoring the previous “totality of the circumstances” test. The courts have since constructed an intricate latticework of exceptions, discussed in further detail below, that have significantly eroded the protections *Miranda* recognized. And executive branch powers in the form of law enforcement agencies at the federal, state, and local levels have further eroded these protections by using these exceptions and improvising their own. In this way, *Miranda* is significant in political developmental terms. Its continuing chain of backlash and retrenchment of police power that over time created a vast array of structural incentives for law enforcement to deceive and coerce suspects has shaped the growth and trajectory of polygraph programs in particular and the surveillance state in general.

This backlash matters because it reflects a political cultural sea change in how political institutions conceptualize coercion – as a narrative frame of normal operations to

be tactically skirted (lawful) rather than a substantively abhorrent moral and political crime to be prosecuted (lawless). Viewing psychological coercion as a legitimate component of the state monopoly on legitimate use of force is a necessary but not sufficient condition of the development of the modern surveillance state. In this way, *Miranda* supported continuing police use and widespread growth of not just lie detection and the closely related Reid interview and interrogation technique, but also the ironically diminished conception of privacy.

A state-recognized right to privacy in the context of criminal proceedings originates in the common law that developed out of liberal backlash against the inquisitorial Star Chamber in England. The Star Chamber almost always found in favor of the state (i.e., the monarchy), using various methods of coercive interrogation to force or construe confessions, often for apparent disloyalty. The reformist response to this official lawlessness was one of the key political historical motivating forces behind the Fifth Amendment protections against self-incrimination. One of the rationales for these protections in this historical context that has support in British political thought and American legal history is spiritual in a broad sense. Theoretically, the Lockean classical liberal idea of limited government is based on the premise that God gave man reason to understand natural law including natural rights to life, liberty, and property, and the purpose of legitimate government (as distinct from religious authority) is purely to protect those rights (Locke 1980). Historically, before the Warren Court’s incorporation of the Bill of Rights’ due process protections, in *Adamson v. California*, 332 U.S. 46 (1947), a defense attorney argued with partial success on appeal that his client’s conviction violated the Fifth Amendment privilege against self-incrimination due to a

prosecutorial comment on the defendant’s refusal to testify. The privilege was held to mean the state cannot comment on one’s refusal to testify, because that violates the inner conscience of the defendant. (Justice Black’s famous dissent in this case center on the fact that – consistent with the political theoretical basis of the right – SCOTUS relied on natural law and the idea of “fundamental fairness” rather than incorporating the Bill of Rights in this decision.)

This spiritual rationale for a right to privacy suggests that individuals must be protected from being coerced into confessing their sins to the state, because the soul does not answer to a secular power – rather, you owe information about your internal state and moral truth only in the confessional to your God, and the state does not obtain the same status as God (O’Brien 1978). This is one rationale for limited government itself, and thus for a key element of the classical liberal conception of what legitimate political power looks like. In some sense it is therefore part of the original basis of American government. But already in *Adamson* and earlier, it was blended with another, more limited and conceptually distinct rationale for the same protections, centering on the idea that criminal justice is like a fox hunt in which the criminal (the fox) must get a fair chance to escape. This justification for privacy protections as a matter of procedural fairness rather than internal sanctity and limited government also has a basis in British political thought and American legal history. Theoretically, it is a manifestation of utilitarian philosopher and legal reformer Jeremy Bentham’s instrumentalism (Bentham 1838; O’Brien 1978). In this context, the “fox hunter’s reason” for a right against self-incrimination is the instrumental achievement of a fair balance between individual and - state power, as opposed to a natural right. Historically, it reflects a Weberian focus on

rational-legal rule-bound government as being more modern and legitimate than its relatively venal alternatives (Weber 2004a).

The post-*Miranda* backlash then, far from being purely reactionary, has some theoretical and historical basis. It reflects both a longstanding theoretical tension between competing justifications for privacy in general and protections against self-incrimination in particular, and a historical legal battle among competing views about how much coercive power (and of what sorts) the state should have to enforce the law in the course of using its monopoly on legitimate use of force to promote lawfulness. *Miranda* suggests that this monopoly itself makes custodial police interrogations inherently coercive, triggering Constitutional protections.

Conservative backlash generates exceptions to the application of these protections, whittling away the due process rights the Warren Court nationalized. This latticework of exceptions effectively incentivized police use of deception and coercion to obtain confessions. This matters politically because it has systematically unequal effects, thus impoverishing what due process means in practice for the most vulnerable criminal suspects. In this way, *Miranda* and its limits contributed to the development of the modern, two-tiered criminal justice system that tends to disproportionately punish minorities, the poor, immigrants, juveniles, the mentally handicapped, political radicals, and other outgroups (Bach 2009; Healy 2004; Roberts and Stratton 2008). According to some legal commentators, the average American commits three federal felonies a day – primarily because vague federal laws forbid everyday activities that would not normally be considered criminal – but discretionary law enforcement and prosecutorial powers tend to target minorities, the poor, and those who threaten state power (Silverplate 2009).

Others consider Watergate and the pattern of elite immunity it initiated the origin of the sprawling modern carceral state that tends to harshly punish the weak while refraining from prosecuting large-scale political and financial crimes (Greenwald 2011). These explanations for the contemporary degradation of due process that characterizes the surveillance state are not conflicting, and indeed it would be surprising if such a massive socio-political development had one rather than many contributing factors.

Miranda contributed to this development by creating a framework within which those who are more knowledgeable about the rules of the game (law enforcement) have substantial power over those who are less knowledgeable (e.g., people with lower levels of human and social capital – disproportionately still minorities, the poor, immigrants, juveniles, the mentally handicapped, and other traditionally disadvantaged groups). In the remainder of this section, I examine select cases that clarified (or constructed) *Miranda*’s limits. My treatment is far from exhaustive – such a project would go beyond the scope of what is necessary to show how what seemed to be a nationalization of a due process standard – a resounding civil rights victory – swiftly and consistently turned into the codification of intricate rules applied inconsistently on the ground, systematically disadvantaging the vulnerable on a procedural level rather than leveling the playing field of justice. Then, I summarize *Miranda*’s primary contributions to the political development of polygraph programs and the contemporary surveillance state as discussed in this section, before evaluating the next watershed moment in this developmental trajectory in the following section.

Three periods of rulings illustrate the development of an extensive latticework of limits of *Miranda*. These limits are important to the political development of

contemporary polygraph programs and the surveillance state, because they establish diminishing due process protections and correspondingly increasing police powers to interrogate suspects, collect evidence, and otherwise induce self-incrimination without triggering what is according to a literal reading of the Fifth Amendment a right against just that. This latticework of limits, precisely because it preserves *Miranda* in name while weakening it in practice, incentivizes various forms of police deception including the use of the Reid interview and interrogation technique, which is known to produce false confessions, and closely related polygraph interrogation protocols.

First, during the Burger Court (1969-1986), rulings illustrate the beginning of conservative retrenchment of the due process revolution. Second, during an intermediate period spanning most of the Rehnquist Court (1986-2001), the Court continued its conservative trend, further trimming due process protections. Finally, after 9/11, SCOTUS rulings on due process limited these protections even more, winnowing the meaning of interrogation, custody, and asserting of due process rights still farther. The continuities across these eras are in many ways more striking than their differences, but I examine a few cases in each era precisely to underscore that striking continuity in spite of vast differences in political cultural context. Contrary to popular belief, there was enormous and fairly early judicial momentum supporting the erosion of due process and thus enabling the development of the surveillance state.

Four Burger era SCOTUS rulings exemplify early post-*Miranda* retrenchment. These cases are *Harris v. New York*, 401 U.S. 222 (1971), *Brewer v. Williams*, 430 U.S. 387 (1977), *Rhode Island v. Innis*, 446 U.S. 291 (1980), and *New York v. Quarles*, 467 U.S. 649 (1984). Together, these cases illustrate a pattern of early retrenchment that is

important in establishing the pattern of exceptions to the due process standard *Miranda* nationalized in a clear bright line rule. Rather than overturning the Warren Court’s bright line rule and *Miranda*, future courts beginning with the relatively conservative Burger Court opted instead to carve out an extensive series of broad and deep exceptions. This pattern continues through the present era, underscoring judicial branch contributions to the development of the surveillance state. Together, these early retrenchment rulings establish significant limits on the meaning, scope, and overall relevance of due process protections for criminal defendants.

In *Harris v. New York*, the Burger Court held that a coerced or otherwise illegal confession may be used to impeach a defendant’s trial testimony. In *Brewer v. Williams*, the Court sanctioned indirect interrogation of criminal suspects by adding the “inevitable discovery” rule as an exception to the exclusionary rule. This means that evidence that is discovered through blatant *Miranda* violations may sometimes be admissible, even if statements made in response to indirect interrogation were not at the time themselves considered admissible. In *Rhode Island v. Innis*, the Court made a partial return to the pre-*Miranda*, “totality of circumstances” standard of what is considered an interrogation and thus triggers the need to Mirandize a suspect. The ruling permitted admittance of statements made in response to *indirect* police questioning of a criminal suspect without a *Miranda* warning. Finally, *New York v. Quarles* created a “public safety” exception to *Miranda* warning requirements. So the “ticking time bomb” sort of example commonly bandied about to justify lower due process standards for terrorist and other criminal suspects has actually had a basis in post-*Miranda* Constitutional law since 1984. The Obama Justice Department famously invoked this exception to justify not Mirandizing

surviving 2013 Boston Marathon bombing suspect Dzhokhar Tsarnaev before his initial joint FBI-CIA interrogations (Gerstein 2013). (Tsarnaev was subsequently Mirandized and ceased cooperating with the investigation.)

This first post-Warren Court period of *Miranda* retrenchment reflects the court’s gradual conservative drift that continued and escalated in the second, intermediated period of retrenchment. Four Rehnquist era SCOTUS rulings exemplify this continuing pattern of post-*Miranda* retrenchment. These cases are *Colorado v. Connelly*, 479 U.S. 157 (1986), *Duckworth v. Eagan*, 492 U.S. 195 (1989), *Sacramento v. Lewis*, 523 U.S. 833 (1998), and *Dickerson v. United States*, 530 U.S. 428 (2000). These cases as a group illustrate an escalating pattern of substantive due process protection retrenchment, establishing the continuity of *de facto* and *de jure* disintegration of a nationalized bright line due process standard procedure. This disintegration further incentivized police use of deception and coercion in interrogation procedures and evidence collection. It undermined *Miranda* without overturning it.

In *Colorado v. Connelly*, the Rehnquist Court ruled that a mentally incompetent suspect could waive Miranda rights as long as he reasonably appears to understand his actions. In *Duckworth v. Eagan*, the Court held that actual *Miranda* warnings in the form of a particular (or even a factually correct) script are not necessarily required, as long as a suspect is fully informed of his rights on the whole. Meanwhile, it held that “actual physical compulsion” is required to establish that an interrogation violates *Miranda*. The dicta “actual physical compulsion,” like the right against self-incrimination, is rooted in the Star Chamber history of the use of physical torture to extract confessions from criminal suspects. Its use in this case, like the distinction between real and testimony

evidence in *Miranda* itself, prefigures more recent use in SCOTUS rulings rejecting a right to privacy and against self-incrimination with respect to attorney, banking, and medical records. This was a significant step back in terms of recognizing a national standard for informing criminal suspects of their due process rights, recognizing privacy rights, and accounting for psychological as well as physical coercion. In *Sacramento v. Lewis*, the Court revived the pre-*Miranda* “shocks the conscience” test, ruling that police who hit and killed a motorcycle passenger, after the motorcycle driver fled an attempted police stop that was initiated during another police car chase, did not behave with “deliberate indifference to life” or in a way that shocked the conscience. Of note is the fact that the Court did *not* hold that there exists a due process right, stemming from *Miranda* or the Fifth Amendment, to not be assaulted or otherwise coerced by the police. Finally, in *Dickerson v. United States*, the Court explicitly addressed the Congressional response to *Miranda*, the federal Omnibus Crime Control and Safe Streets Act of 1968, striking down a section of it to affirm *Miranda*’s Constitutionality (and thus its applicability to the federal government and the states). In its reasoning, the Court also addressed the contrasting realities of the long chain of judicial exceptions to *Miranda* versus its omnipresence in popular culture as a hallmark of the procedural fairness of the U.S. justice system. It is worth noting that in addressing this tension, the Court did not reassert the precedence of *Miranda* warnings or other due process protections over apparently conflicting standards such as the “totality of circumstances” standard, or other constraining holdings.

This second post-Warren Court period of post-*Miranda* retrenchment reflects the continuation and escalation of retrenchment of due process rights. This pattern further

escalated in the final, continuing period of this retrenchment. Several post-9/11 SCOTUS rulings stand out as establishing new limits on what process is due under *Miranda* in police interrogation: *Berghuis v. Thompkins*, 560 U.S. 370 (2010), *J.D.B. v. North Carolina*, No. 09-11121 (2011), *Howes v. Fields*, 565 U.S. No. 10-680 (2012), and *Salinas v. Texas*, 570 U.S. No. 12-246 (2013). These rulings have enough temporal distance from 9/11 to illustrate that the trend of prioritizing security over liberty was not a temporary, reactive one, but rather a long-term one with strong pre-9/11 roots and continuing post-9/11 momentum. Together, these cases underscore continually growing limits on what counts as a custodial interrogation, when *Miranda* warnings are due as a matter of nationalized procedure, and how due process rights must be asserted in order to apply.

Berghuis v. Thompkins established that police may interrogate suspects for lengthy periods of time after *Mirandizing* them, even if the suspects are silent and otherwise behaving as though they are asserting their right against self-incrimination. The suspect has to say the magic words (something unambiguous along the lines of, “I am invoking my *Miranda* rights to remain silent, and I would like to talk to a lawyer”) in order to stop the interrogation. But police are under no obligation to clarify that this is how to stop the interrogation. The right does not, then, have to be explicitly waived in order for police to conduct an interrogation from which they may use a confession in court. Not coincidentally, proponents of coercive interrogation practices celebrated this ruling as essential in the post-9/11 era (Yoo 2010).

In *J.D.B. v. North Carolina*, police and school administrators interrogated a 13-year-old special education student about two robberies without *Mirandizing* or giving

him an opportunity to call his guardian. SCOTUS ruled the lower court should have considered his age as part of the totality of circumstances in deciding whether the minor was in custody and thus had to be Mirandized. It is worth nothing that the Court did not rule that Miranda warnings had to be issued any time police questioned a juvenile criminal suspect, or that police interrogations in schools are custodial interrogations, or speak more specifically to how taking age into account should affect the delivery and explanation of due process rights (e.g., making the implications of the rights more explicit so young people are more likely to understand their options). This case constrained due process rights by failing to explicitly extend Miranda rights to juveniles questioned by police in school – constraining what counts as a custodial police interrogation.

In *Howes v. Fields*, the Court redefined what it means to be in police custody and under interrogation, reapplying the old “totality of circumstances” rule to hold *Miranda* did not apply in the given circumstances. In this case, the criminal suspect was already an inmate in prison, and police wanted to question him about criminal activities unrelated to his imprisonment. They invited him to leave his cell, and police interrogated him for several hours without Mirandizing him. He subsequently confessed to a new criminal allegation and was indicted for that crime. His confession was deemed admissible because, given the totality of circumstances, he was not considered to be in custody. Even though he said he did not want to talk, he never asked to return to his cell, even though police had said that he was free to do so.

Salinas v. Texas, 570 U.S. No. 12-246 (2013) established a criminal suspect’s *selective silence* during police interrogation can be used against him at trial. Salinas

answered some questions and was silent in response to others, including a question about whether bullet casings at a murder scene would match his gun. SCOTUS held that this interrogation, because Salinas was not under arrest, was a non-custodial police interview. Salinas did not say the magic words (e.g., “I invoke my Fifth Amendment right to silence in response to this specific question”). Therefore his silence could be used against him. This case thus further constrained due process rights by again constraining what counts as a custodial police interrogation.

In *Salinas*, SCOTUS also directly addressed the “magic words” criticism of the contemporary due process rights dance. Denying the existence of a “ritualistic formula” for claiming the right against self-incrimination, the ruling nevertheless insisted that silence does not trigger the protected right to remain silent. The right must be invoked to be covering. This is akin to asserting you have to tell someone with a gun who is demanding your wallet that you would really prefer to assert your property rights as protected by the law – or you will have not been robbed when you hand over the goods. Hobbes in fact argues this is how natural law works, and this is no coincidence (Hobbes 2009). As I will discuss in greater detail in the penultimate section, on the post-9/11 moment in development of polygraph programs and the surveillance state, a Hobbesian as opposed to a Lockean construal of natural rights characterizes this new Constitutional regime surrounding due process. Defenders of wartime constraints on due process, such as Bush II administration former assistant attorney general for the Justice Department's Office of Legal Counsel John Yoo, specifically cite Hobbes's philosophy that individual rights are always legitimately limited by their necessary balance with security, whether the limitation comes from your own choices in private interactions with other individuals

(e.g., to give a highway bandit all your money rather than your life), or from your interactions with state actors (Yoo 2006; Yoo 2005).

On the whole, these three sets of four cases spanning three distinct post-*Miranda* eras of increasing judicial conservatism and due process rights retrenchment illustrate both significant continuity and escalation in the limits of *Miranda* and the Warren Court due process revolution. They both respond to and institutionalize mounting police incentives for psychological coercion, such as interrogation practices including lies and lie detection that are based on insufficient scientific evidence and known to generate false confessions. Moreover, they foreshadow and contribute to a growing disconnect between the meaning of due process for the privileged and the vulnerable.

As this long pattern of exemptions and exceptions illustrates, *Miranda* and its legacy cemented the separations between legal (especially criminal) and extra-legal proceedings, circumscribing the physical right to privacy while – at least at first – recognizing the psychological or spiritual one in the privilege against self-incrimination. This material limitation and its momentum foreshadowed and laid the foundations for the post-9/11 development of the modern surveillance state. It did this in part by constraining what due process meant under the official interpretation of the Fifth Amendment, and also by creating conservative backlash that generated a cascading series of exemptions, classes and contexts to which due process (in this newly substantive and nationalized, but also newly limited conception) does not apply.

Conceptually, *Miranda*’s redefinition of coercion to include psychological coercion created more than tremendous backlash and a series of exceptions. It replaced, at least in part and for a time, the old standard (“totality of circumstances”) for determining whether

coercion had poisoned information police obtained during interrogation. Instead of replacing it with a new definition of coercion, it assumed police interrogations to be coercive interactions in which at least some classes of people (criminal defendants) in some contexts (custodial interrogations) had some Constitutional protections (the right to remain silent and to access counsel, and the right to end interrogation by asserting those protections).

Miranda is also hugely politically significant in the context of the development of polygraph programs and the surveillance state because it expands federal governmental power and influence in law enforcement. Previously, the Bill of Rights was unincorporated and no national standard existed for what constituted a coerced confession (which then had to be excluded at trial). The Warren Court attempted to equalize what due process meant in American by nationalizing this standard. This cut into the heart of federalism, effectively making the courts the managers of local and state law enforcement practices. While federal management of local and state police practices has taken less liberal forms more recently, the Warren Court was highly conscious of the power dynamics – and especially the racialized nature – of criminal justice. This was a time when federal power was widely seen as a liberalizing force in race relations, and with just cause (Tushnet 2006). The Warren Court ruled against school segregation, anti-miscegenation laws, and other racist practices. And just as *Frye* involved a black criminal defendant, so did many of the Warren Court’s seminal due process cases involve minority defendants (e.g., *Miranda* of the eponymous case was a Mexican-American defendant).

But by nationalizing a due process standard, the Warren Court in *Miranda* also laid the foundation of the contemporary surveillance state in two primary ways. First, it

initiated a long period of broadening and deepening federal governmental influence on local and state law enforcement at a managerial level. This trend is in tension with federalism, as exemplified by Article X of the U.S. Constitution, and its continued American political development in the form of the Posse Comitatus Act of 1878 (18 U.S.C. § 1385) in particular. Article X reserved powers not delegated to the federal government for the states or the people. Posse Comitatus limited the ability of federal military forces to enforce state laws, in reaction against Reconstruction.

Yet local and state law enforcement agencies have increasingly followed federal leads in adopting operating procedures, as exemplified by growing use of polygraph programs. Second, local and state law enforcement responses to particular issues or groups are also increasingly influenced by the federal government. This trend is exemplified by DHS and FBI-coordinated local police responses to the Occupy Wall Street movement in 2012 (The Partnership for Civil Justice Fund 2012; Wolf 2012). In this light, increasing local and state law enforcement use of federally subsidized and otherwise supported equipment, training, methods of operation, and even ways of dealing with specific political movements – contemporary trends of increasing federal management of ostensibly autonomous, state and local law enforcement agencies – are part of the counter-intuitive legacy of *Miranda*. This legacy ironically disadvantages some of the same sorts of minority groups, such as the Occupy Wall Street protesters – a contemporary reincarnation of pro-labor reformers, who were historically disproportionately subject to police coercion and abuse, and whom the Warren Court sought in its due process revolution to protect. As in Skowronek’s political developmental analyses of American state-building in the Progressive Era, we see here a

deepening of federal power in general – more federal state building – in response to external and internal threats (Skowronek 1982).

Conclusion

Current interrogation practices are the continuing legacy of Progressive reforms, centering on various forms of lie detection, such as polygraphs, voice stress analysis, behavioral analysis, and the Reid interrogation technique that sprang from and substantially overlaps with lie detection research and practice. Yet almost immediately, these techniques came to be used disproportionately against the marginalized – racial, religious, and sexual minorities, whistleblowers, and criminal suspects without the psychosocial resources to resist psychologically coercive, fraudulent interrogation practices (Alder 2007). Ironically, twentieth- and early twenty-first civil rights reformers failed to interrogate these methods, instead pressuring for their effective expansion as the framing of scientific neutrality was accepted and embraced as an attainable ideal by victim rights’ groups and others. Post-*Miranda* retrenchment presented opportunity after opportunity to contest this framing. These opportunities were missed. Progressive legalism was on the defensive against a rising conservative tide that effectively gutted the due process revolution without ostensibly overturning its keystone cases such as *Miranda*.

By keeping the reform discourse focused on procedural fairness at the level of means-neutrality, this failure contributed to the racialization of the punitive trends that have altered the nature of citizen-state interactions (Alexander 2012). This racialization is especially marked in major cities, where increasing evidence suggests punitive criminal

justice policies fail to prevent crime (Sunshine and Tyler 2003). Moreover, people who experience unjust encounters with law enforcement are subsequently less likely to engage in public activities like voting (Lerman and Weaver 2014; Lerman and Weaver 2012; Weaver and Lerman 2010). This civic disengagement might be part of a broader withdrawal response with a long tail of individual and community costs – extending the socio-political effects of the two-tiered justice system from the criminal justice realm, to the health of the social contract and the community in much broader ways (Travis 2013).

If *Miranda*’s primary contributions to the political development of psychologically coercive, deceptive police interrogation practices including polygraph programs and the contemporary surveillance state ironically stemmed from nationalizing (some) procedural due process standards, the next watershed moment in the development of this new constitutional order came, equally ironically, in the form of a federal law that institutionalized the constrained conceptions of privacy rights and the categories of people to whom procedural fairness is due, while apparently protecting (some classes of) people’s privacy and due process rights (in some contexts).

There are two chief commonalities, which I examine in greater detail in Chapter 3, between the *Miranda* moment and its ongoing latticework of retrenchment, and the EPPA moment and its analogous effects. First, both *Miranda* and EPPA look progressive, but actually represent significant conceptual and practical constraints on due process and the spiritual justification for privacy rights in the context of the right against self-incrimination. Just as *Miranda* constrained the Fifth Amendment right against self-incrimination by reading it instead as a privilege against self-accusation, laying the foundation for myriad legitimate exceptions, EPPA protected some employees’ privacy

in terms of recognizing the sanctity of their internal states, but simultaneously legitimated broad government and private sector exemptions to this protection. Second and relatedly, both *Miranda* and EPPA generated a long and continuing chain of backlash that generated mounting incentives for governmental (especially law enforcement) misconduct.

Chapter 2 has explored the first of two landmark moments reflecting efforts to constrain the use of polygraphy. Chapter 3 explores the second set of turning points. The first of these is the Employee Polygraph Protection Act (EPPA), which standardized privacy protections for some employees. This reflected a focus on what privacy rights individuals have over their internal states, and in what contexts. The second is the explosion of post-9/11 lie detection program growth.

All these landmark moments reflect efforts to constrain the use of polygraphy, but their very different emphases reflect responses to overarching trends of increasingly federalized governmental power and institutionalization of exceptions for security – trends that come to a head in the fourth, post-9/11 era of the surveillance state. The post-EPPA and post-9/11 eras in lie detection program growth share a key feature that makes them distinct from *Frye* and *Miranda* as landmarks. *Frye* justified a nationalization of standards regarding what evidence is, and *Miranda* justified a management role of the federal government in local and state police practices regarding what interrogation and coercion are. By contrast, EPPA institutionalized the use of a liberty-security opposition to carve out exceptions to the protections of civil rights and due process that it enshrined. This laid the foundation for more exceptions using the same logic. These exceptions were thus not a new, post-9/11 development.

Indeed, as Chapter 2 has demonstrated, they are in some ways a continuation of post-*Miranda* retrenchment. But as Chapter 3 will show, post-9/11 growth in lie detection programs has been distinctly massive: quantity in this case expresses a salient quality all its own. This shift to an explicit logic of security over liberty prioritization in the context of polygraphy is a significant political development, because lie detection embodies a particular construction of state power and social control in which data-driven – but not necessarily scientific – embodiments of the fact-value distinction enact domination of the objective over the subjective. We internalize this domination when it changes how we think about truth and justice.

Truth, Lies, and Polygraph Tape:
Later Development of the Contemporary Surveillance State

“The physiology of the detainees was usually horrible due to their lack of sleep, eating, depression, and extreme temperatures.” – Air Force internal document comment from an anonymous American polygrapher of Abu Ghraib detainees, on one reason why he believed polygraphs in this context to be inaccurate.

Introduction

Recall that this dissertation presents a sweeping analysis of administrative technology as an evolving way of understanding expertise and the Enlightenment fact-value distinction that has tremendous implications for politics from the construction of state power to the psychology of street-level bureaucrats. Chapter 2 argues that landmark efforts to protect individual rights and limit government power – although they succeeded in important ways – also created a moral-political landscape that eroded the parallel fact-value and public-private boundaries gradually farther and farther until the external shock of 9/11 turned that erosion into a landslide. Four landmarks mark the path of the developing federal legal regime surrounding lie detection. These landmarks among others delineate important conceptual shifts that illustrate and contribute to this key boundary erosion. As a result, they are also major turning points in the development of the contemporary surveillance state. Chapter 2 explored the first pair of these four turning points: the Progressive Era precedent *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923), and the Warren Court precedent *Miranda v. Arizona*, 384 U.S. 436 (1966). Chapter 3 now explores the second pair of these turning points: the Employee Polygraph Protection Act (EPPA) and the post 9/11 era.

Chapter 3 is structured as follows. First, I analyze EPPA as a key turning point in the development of polygraph programs and the surveillance state. Recall that

EPPA standardized privacy protections for some employees but not all, reflecting a focus on what privacy rights individuals have over their internal states and in what contexts. This focus reflects a different use of the fact-value distinction than in *Frye* and *Miranda*. In the post-EPPA and post-9/11 eras of lie detection, this distinction constrained some forms of state power. *Frye* justified a nationalization of standards regarding what evidence is, and *Miranda* justified a management role of the federal government in local and state police practices regarding what interrogation and coercion are. By contrast, EPPA institutionalized the use of a liberty-security opposition to carve out exceptions to the protections of civil rights and due process that it enshrined. This assumed a parallel association between squishy ethics (value) and material realities (fact) – using the same fact-value distinction to expand state power. EPPA matters because it laid the foundation for more exceptions using the same logic, and these exceptions were institutionalized in a landslide after 9/11. As Chapter 2 demonstrates, these exceptions in some ways continue post-*Miranda* retrenchment.

But as Chapter 3 shows in an examination of post-9/11 growth in lie detection programs, these exceptions have been distinctly massive. This shift to an explicit logic of security over liberty prioritization in the context of polygraphy is a significant political development, because lie detection embodies a particular construction of state power and social control in which data-driven – but not necessarily scientific – embodiments of the fact-value distinction enact domination of the objective over the subjective. We internalize this domination when it changes how we think about truth and justice. Big Data does not have to mean Big Government, but the alignment of the fact-value distinction with the security-liberty opposition presupposes that it does.

Protecting (Some) Employees: The Employee Polygraph Protection Act and the Bounds of Legitimate Government

Like *Frye* and *Miranda*, the Employee Polygraph Protection Act of 1988 (29 USC §2001 et seq.; 29 CFR Part 801; hereafter “EPPA”) was a move by independent oversight authority to check a broader problem of perceived procedural unfairness with a new national standard. That problem was unscientific expert testimony in *Frye*, inconsistent due process standards in *Miranda*, and growing demand for employee rights in the context of changing cultural conceptions of privacy in EPPA. All these watershed moments in the development of polygraph programs and surveillance state are politically significant because they make our contemporary new Constitutional regime a logical and under certain conditions inevitable outcome, even though it seems at first glance that they represent institutionalized constraints and pressures that would push the opposite way. In particular, Progressive Era reform pressures, the Warren Court due process revolution, and the apparent labor victory of EPPA all shared a notable, albeit counter-intuitive, anti-civil liberty element. This element cemented the foundations of the contemporary surveillance state.

In this section, first I restate more specifically what EPPA established as a law. Then, I review the effects it had on use of lie detection in practice. Because its aggregate market effects (i.e., on the overall use of polygraphy) are so counter-intuitive, this review leads to a further discussion of four effects EPPA did *not* have. Then I discuss four effects EPPA actually did have, highlighting its contributions to the political development of massive lie detection programs and the contemporary surveillance state. Finally, in the next section I discuss the post-9/11 developmental moment of explosion in the growth of

polygraphy and surveillance. If 9/11 was the match that lit this political developmental fuse, EPPA was the last plank in the tinderbox.

EPPA is a federal law Congress passed forbidding most private employers from requiring or requesting current or prospective employees to take lie detector tests, or penalizing them in any way for refusing to take such a test (U.S. Congress 1988; U.S. Department of Labor 2009). Lie detector tests include polygraphs, voice stress tests, and any other psychological stress measurement device used to diagnose truthfulness/deception or honesty/dishonesty. Like other civil rights laws, EPPA nationalizes a federal floor of civil rights protections which some state and local laws go beyond, providing greater protections. It also nationalizes exceptions to these protections (which state and local laws can trump). These exceptions permit lie detector tests for security-related private sector and governmental employees, and when there is an ongoing investigation. The security exemption covers armored car, alarm, guard, and pharmaceutical companies: people who handle guns, drugs, and large amounts of cash. Its *de facto* interpretation can be narrower, however, because polygraphers in private practice generally prefer to take on private (non-employee) and public sector (governmental) clients who pose less of a risk of potentially violating EPPA. EPPA also lets private employers polygraph employees under administrative investigation for action resulting in economic injury to the employer, like workplace theft. In such cases, EPPA requires the employer describe in detail the basis of his reasonable suspicion that the employee was involved in the activity being investigated – a requirement that might contribute to potentially systematic problems of confirmation bias in polygraphy by giving the polygrapher negative background information about the subject. EPPA also

establishes record-keeping requirements, but not reporting requirements, as well as federal enforcement teeth. The Employment Standards Administration’s Wage and Hour Division of the U.S. Department of Labor administers EPPA, and the Secretary of Labor may sue violators for up to \$10,000 per violation. Employees or recruits may also bring private suits under EPPA for additional relief. In this respect, EPPA was set up to socialize costs of violations, rather than making it incumbent on individual complainants to bring private suits. These liabilities and their socialized nature, again, limit private sector polygraphers’ use of the more obscure exceptions to EPPA. Perversely then, EPPA generated incentives for polygraphers to seek out clients with no due process protections (e.g., criminals and governmental employees) rather than clients who might claim some.

An informed observer might have seen EPPA as a death knell for lie detection in general and polygraphy in particular. It placed significant limits on private sector employment-related polygraphs at a time when the majority of polygraphs fell under that umbrella. But public sector use of polygraphy expanded quickly and consistently after EPPA, so that the number of polygraph exams administered in the U.S. today (about 2.5 million) is greater than it was before the EPPA temporarily lowered it (2 million). As I detailed in the section on polygraph use above, lie detection is common, costly, and chronic. What, then, were EPPA’s effects and why does this legislation, which curiously failed to even reduce the long-term use of a practice it largely outlawed, matter at all?

Given the failure of sweeping federal legislation to reduce the overall use of the polygraph in the long term, it seems natural to briefly list the other effects that EPPA surprisingly failed to have. First, EPPA did not establish an oversight framework for exempt polygraph programs. It did nothing to institutionalize accountability or

transparency in government polygraph programs in particular, despite a long history of reported abuses. For example, in 1985, GAO reviewed the DoD’s polygrapher training. They reported it was conducting unauthorized training exercises without informed consent of participants, and instructing trainees in techniques that might violate DOD policies about individual rights and privacy (Government Accountability Office 1985). As I will discuss in greater detail in the next section, sources strongly suggest that such abuses are ongoing. These sources include interviews I conducted with people who administered and underwent federal governmental polygraphs, in combination with previously unreleased documents I obtained from multiple federal governmental sources. Lack of accountability and external oversight authority in domestic and international U.S. governmental polygraph programs seems curious given that an act of Congress significantly reined in the industry.

Why didn’t EPPA institutionalize oversight for governmental polygraph programs? The most obvious possible explanation for this oversight regarding oversight is a common explanatory thread in all the effects EPPA surprisingly failed to have. Interest-group liberalism in Lowi’s sense of pluralistic competition resulting from broad expansion of public programs caused the construction of a federal legal regime surrounding lie detection that privileges the old power of particular industries and sectors while creating new privileges or protections for other groups (Lowi 2010). This regime reflects diminished popular control as a result of Congressional delegation of its authority to interest groups. It contributes to the diminished rule of law that characterizes the contemporary surveillance state, replacing equal protection under the law with a codified stratification of due process privileges for some.

Accordingly, EPPA failed to treat all employers – or even all private employers – equally, to treat all classes of people equally, and to validate a privacy right as the basis for anti-polygraph legislation. These failures continued to lay the legal and political cultural framework within which the explosive post-9/11 growth of polygraph programs and the surveillance state was inevitable. In the remainder of this section, I detail these failures, why they might be surprising, and how they happened, before highlighting what EPPA *did* accomplish and why it matters.

EPPA’s failure to treat all employers alike represents a win for multiple interest groups. Existing governmental programs were exempt from the legislation, as were a number of private industries with notoriously powerful lobbies, including arms and pharmaceutical industries. As Lowi opined, “Stress on civil liberties is always likely to work to the benefit of those who already have the wealth and power to defend their liberties as well as their luxuries” (Lowi 2010). EPPA responded to the stress lie detection placed on civil liberties precisely by institutionalizing special status for these powerful interest groups. In this respect, EPPA exemplifies the use of legislation to shelter bureaucracy and practices that the private sector and the public would no longer permit, protecting government polygraph programs and their private-sector siblings – security-related polygraph programs staffed overwhelmingly by former government polygraphers – from market pressures. At the same time, EPPA was a tremendous political win for organized labor. Not only did it protect the vast majority of private sector workers from routine employment polygraph screenings, but it also explicitly preserved the right of workers in unprotected sectors to use collective bargaining to better protect themselves. However, the nature of EPPA’s exceptions simultaneously laid the

foundation for continuing lie detection and surveillance program growth by legitimizing the appearance of a trade-off between security and liberty – despite insufficient evidence establishing the efficacy of these programs. Conceptually, debate about such a trade-off sets the terms of public discourse about privacy in a fashion that privileges the conceptualization of privacy rights as instrumental or procedural rather than more deeply moral or spiritual.

EPPA’s failure to treat all employees alike thus matters conceptually because it further erodes the Fifth Amendment right against self-incrimination, cementing it instead as a privilege reserved for some classes of people. This leaves no redress for due process violations for the people omitted from the protective umbrella. Thus, like *Miranda*, EPPA diminishes due process protections while seeming to institutionalize them, prefiguring subsequent retrenchment of the very programs and practices it purported to limit. For example, it fails to provide a remedy to governmental employees when national security-related polygraphs cost them jobs in relation to perceived Title VII violations (i.e., on the basis of discrimination relating to race, color, religion, sex, or national origin). This means cleared employees have no legal recourse against discrimination that violates widely shared legal and ethical norms (Mayer 2000). Thus, as a consequence of EPPA’s failure to treat all employees alike, government polygraph programs are not subject to equal opportunity law. This matters, because lack of diversity can harm organizational efficiency and thus jeopardize national security agencies’ aims (Callum 2001; Richard 2000; Sommers 2006).

Even in light of the inegalitarian consequences of EPPA’s failure to give all employees equal due process protections, EPPA’s most profound failure may have been

its conceptual one. It does not establish a privacy right as the basis for its protections. Instead, it establishes privileges for certain classes of people in certain contexts. In this way, like *Miranda*, EPPA further entrenched a narrow reading of Fifth Amendment due process protections, interpreting them as privileges stemming from the practical utility of procedural fairness instead of rights stemming from a deeper source. That deeper source might be spiritual, in the sense of the hermit’s rationale for a right against self-incrimination, relating to ideas about sanctity of inner state and obligation to share that state only with God and not with the state (O’Brien 1978). It might just as easily be scientific, in the sense of establishing a standard of proof that must be met before courts evaluate or abstain from evaluating assertions that the interests of security trump those of liberty. It serves security interests poorly, for instance, for polygraph programs to operate outside the domain of equal opportunity law if they do not actually deter misconduct in the first place. In any event, this diminished conceptualization of the justification of privacy rights is politically significant because it weakens the juridical and political cultural discourse on civil liberties, reducing it to an imagined balancing act between the ostensibly competing values of liberty and security.

EPPA’s failures as a whole, and in particular the failure to institute oversight in relation to government polygraph programs, might at first glance appear surprising. After all, multiple Congressional evaluations of government polygraph programs concluded there was insufficient evidence supporting these programs’ efficacy. In 1963, a Congressional committee headed by John Moss also reported polygraphy itself lacked research establishing its validity. In 1974, a House Committee recommended all government agencies stop using polygraphs for all purposes. And in 1983, the Office of

Technology Assessment (OTA) produced an evaluation of polygraphy at the request of Congress after the President asked to increase government use of polygraphy in response to leaks. Much like the National Academy of Sciences in 2003 when Congress requested another evaluation of polygraphy, OTA reiterated prior Congressional conclusions that existing evidence did not support government polygraph programs’ efficacy and polygraphy itself lacked insufficient scientific basis. Despite this long and consistent chain of Congressional condemnation of governmental polygraph programs in particular, EPPA exempted such programs completely.

Interest-group liberalism matters, but it only partly explains this outcome. Its significance continues to drive important phenomena such as private sector capture of public policy, as exemplified by the International Association of Chiefs of Police using federal grant money in 1995 to write a model national polygraph policy for law enforcement in collaboration with the main private interest group of the polygraph industry, the American Polygraph Association. Another, necessary but not sufficient condition is that government polygraph programs lack oversight because the bureaucracies within which they are situated lack effective accountability or oversight in a much broader context. These programs are situated in bureaucracies under the executive branch (e.g., CIA, DIA, NSA, and FBI) that lack effective accountability and oversight more generally. Previously unreleased documents I obtained under the Freedom of Information Act show this lack of external oversight of government polygraph programs is ongoing, is associated with a pattern of complaints, and dovetails with ongoing and repeated failures to produce scientifically valid studies assessing these programs’ efficacy. Alleged polygraph program mismanagement has affected

governmental employees’ lives since EPPA so significantly that multiple cases of suicidality relating to polygraph complaints are documented. But it is not particularly surprising that Congressional and judicial limits on lie detection programs have failed to limit government polygraph programs. Rather, it would be surprising if these programs were an exception to the overwhelming rule of executive agency autonomy that has grown stronger over much if not all of American history, especially since the birth of the regulatory, then carceral, and then surveillance states in the last hundred years.

EPPA, then, had a number of failures with important political developmental consequences, but none of these failures was surprising. What EPPA succeeded in accomplishing in this context was no small feat: it standardized privacy protections for large classes of employees, simultaneously institutionalizing exceptions to these protections. In this way, EPPA marked a resurgence of concern over individuals’ privacy rights with respect to their internal states, in contrast to conservative post-*Miranda* retrenchment that tended to constrain these rights. At the same time, EPPA was concerned only with some individuals’ privacy rights – private sector employees not working in security-related sectors and not under administrative investigation – while *Miranda* protected only criminal defendants. One marked continuity, then, between EPPA and *Miranda* is this limitation on the very due process rights they promote. This limitation is both conceptual and practical. Conceptually, just as *Miranda* read a relatively narrow interpretation of the Fifth Amendment into law, so did EPPA institutionalize a narrow conception of what process is due employees as a normative matter. Practically, just as *Miranda* defined the classes of people to whom due process applied relatively narrowly, by not recognizing everyone as a potential criminal suspect

or defendant, so too did EPPA define particular classes of people to whom its protections apply.

Overall, the very different emphases of *Frye*, *Miranda*, and EPPA – what counts as expert scientific testimony, what counts as coercion or interrogation, and what process is due private employees as a matter of right versus privilege, respectively – reflect changing institutional responses to overarching trends of increasingly federalized governmental power and formalized exceptions for security. These trends come to a head in the fourth, post-9/11 era of the surveillance state.

Out of *Frye* and Into the Fire: Expertise, Privacy, and Due Process Post-9/11

The previously analyzed *Frye*, *Miranda*, and EPPA moments of conceptual shift in the development of the U.S. legal regime surrounding lie detection all established counter-intuitive constraints on the very due process protections they seemed to promote. These constraints cemented the foundations of the legal and institutional architecture of widespread modern-day lie detection programs and the surveillance state as a whole. The post-September 11, 2001 (9/11) political historical moment of explosive growth in these lie detection and surveillance programs is different from its precedents because it is marked by considerable growth in lie detection in the absence of a path-changing legal precedent or law. The existential threat of 9/11 had a range of political effects that such threats usually have, including lack of rule of law, aggressive punitiveness toward outgroups, and other manifestations of elevated authoritarianism (Feldman and Stenner 1997; Hetherington 2009).

And yet the explosive post-9/11 growth of polygraphy and related deception detection programs detailed below also illustrates deep continuity between this new political historical moment and the preceding eras discussed above. In this section, I argue that this ongoing fourth era, which other scholars have identified as one of constitutional regime change, is the logical convergence of unintended consequences of these preceding moments under the right (or wrong, as the case may be) political historical conditions (Balkin 2008; Balkin and Levinson 2006). First, I briefly explore its continuity with the old surveillance state legal regime established across all branches of the federal government in legislation such as the Foreign Intelligence Surveillance Act (FISA) of 1978, precedents such as *California v. Ciraolo*, 476 U.S. 207 (1986), and the autonomous political behavior of bureaucratic agencies under the executive. Second, I delineate its continuity with the legal lineage of the watershed moments of focus analyzed above – *Frye*, *Miranda*, and EPPA. Third, I describe the widespread use and growth of lie detection and its descendants today.

Although “surveillance state” is a relatively new term in the legal literature referring to governmental operations such as mass communications metadata and data collection programs that relatively new technologies have made possible in recent decades, in some ways the new surveillance state is the same as the old surveillance state. As the Church Committee documented, every U.S. President since Eisenhower has abused domestic intelligence-gathering means to spy on Americans for political purposes. The resultant Foreign Intelligence Surveillance Act (FISA) of 1978 established that this was fine by Congress as long as these were Americans suspected of espionage or terrorism. In *California v. Ciraolo*, SCOTUS ruled that aerial surveillance of people's

literal backyards was also Constitutional. Bureaucratic agencies under the executive have operated with increasing autonomy since the Progressive Era (Carpenter 2001). The close collaboration between governmental and private companies that characterizes the new surveillance state also goes way back. For instance, Western Union allowed government access to all private telegrams it transmitted in the 1950s and 1960s, and AT&T has similarly long allowed such access to its customers’ calls (Greenwald 2011).

In addition to expanding on these surveillance precedents through expanding means of surveillance, the post-9/11 surveillance state builds on the legal genealogy of *Frye*, *Miranda*, and EPPA by reapplying these eras’ constraints on due process to the same new socio-political contexts created by expanding means of surveillance. Specifically, *Frye* and the Progressive Era professionalization and standardization reform pressures it exemplified foreshadowed the professionalization and standardization of U.S. law enforcement, intelligence, and military interrogation practices including various forms of lie detection. After 9/11, these practices were explicitly expanded to again include torture, and this expansion was planned as part of the post-9/11 federal legal regime. We know this was planned because of documentation of meetings including military and CIA lawyers discussing the need for documentation to protect Americans who were to engage in the proposed coercive interrogation practices (Siems 2011).

We know this planned combination of coercive interrogation practices and the professionalization needed to make them at least appear somewhat rational-legal was then executed on the ground because of documentation of torture from a variety of internal U.S. governmental sources, including feedback from Air Force and NSA polygraphers who interrogated thousands of detainees in Iraq and Afghanistan following

U.S. invasions. Polygraphers noted they were giving “DI” (deception indicated) exams, sometimes by structuring polygraph “tests” so that detainees who interrogators considered guilty were bound to “fail,” and sometimes by lying to detainees about test results in order to use the polygraph as a “hammer” against detainees in interrogations. Some complained that military interrogators were feeding them erroneous and damning information. Multiple polygraphers expressed misgivings about the accuracy of polygraphs conducted in these environments. Some cited the loud noises from nearby demolitions and lack of soundproofing in polygraph interrogation rooms, while others explicitly referred to the effects of torture on their subjects. Regardless, U.S. military and intelligence agencies continued contracting with polygraphers to perform testing on detainees overseas. Anecdotal reports suggest this practice continues and has expanded in recent years to include use of handheld “lie detectors” by U.S. troops in Iraq and Afghanistan. These devices, which are even less reliable than traditional polygraphs, do not produce data that would be admissible in any U.S. court of law. This disjuncture between legal and extra-legal fact-finding standards reflects continuity of the two-tiered approach institutionalized in *Frye*.

Relatedly, *Miranda*’s standardization of what counts as a coerced confession stretches a long shadow over the modern-day surveillance state, particularly with respect to increasingly widespread lie detection programs that can affect everyday freedom of movement, such as FAST and AVATAR. “Real evidence” such as physiological data does not require consent under *Miranda* and other due process precedents of the avowedly radical Warren Court. This limited conception of what the right against self-incrimination means assumes that the right is about procedural fairness, effectively

functioning as a privilege triggered only under certain conditions, rather than internal sanctity, functioning as a natural right that it is the state’s primary *raison d’être* to protect. This conception permits security agencies to interrogate employees and recruits about personal issues including racial and political beliefs and practices, sexuality, sexual abuse and trauma, psychological problems, and decades-old drug experimentation – with no threat of legal repercussions, because due process protections including anti-discrimination laws are not triggered. Senior officials at multiple security agencies have reportedly told polygraphers that if they are not getting complaints, they are not doing their jobs.

They are getting complaints. Documents I obtained from the Department of Defense Office of the Inspector General reflect an ongoing pattern of complaints, allegations of wrongdoing, and requests for assistance related to alleged federal polygraph program waste, fraud, and abuse as recorded in the Defense Case Activity Tracking System (D-CATS). Several complainants allege polygraphers behaved in “unprofessional” and “nasty” ways or violated non-discrimination laws; there is no evidence indicating these complaints were investigated. One black applicant from Baltimore whose FBI polygraph led to his disqualification for a federal job complained: “I was called a lazy, lying, drug dealing junkie by a man who doesn’t know me, my stellar background or my societal contributions... Just because I am young and black does not automatically denote that I have ever used any illegal drugs.” Private sector complaints under EPPA concluded since fiscal year 2007 – completely separate from these public sector complaints – total over 1.25 million (U.S. Department of Labor 2013).

Polygraphers and high-level officials have a pattern of complaining, too. One 2008 whistleblower complaint to a Department of Defense hotline resulted in the supervisor of the CIA polygraph program in Reston, Virginia receiving a written reprimand for partially substantiated allegations of mismanagement, creating a hostile work environment, and failure to take action when potential security violations are discovered as a result of polygrapher activities. No further action or investigation was taken. Another whistleblower complaint in 2013 concerned alleged polygraph program-related contracting crimes, failure to report criminal confessions including child molestation to law enforcement, and top-down polygraph interrogation of employees about sensitive personal issues such as childhood sexual abuse. The complaint was made by four senior NRO officials to the Inspector General of the NRO, and resulted in the agency’s deputy director conducting an auspicious search to find and fire the officials, and threatening to prosecute them as leakers. (NRO’s polygraph program is managed and staffed by CIA and Air Force polygraphers.) NRO’s Inspector General subsequently sent a “seven-day letter” to Congress, a very uncommonly used mechanism for reporting serious agency misconduct (Taylor 2013). No corresponding Congressional action is documented.

Finally, EPPA further institutionalized the post-*Miranda* retrenchment of due process protections by enshrining extensive security-related exceptions to privacy rights in federal law. Just as private security-related and public sector employers may still require recruits or employees to submit to routine polygraph screening under EPPA, so too are exceptions to due process privileges on security grounds increasingly widespread in the post-9/11 era. This reflects a natural extension of the idea that an individual’s

internal state is unprotected outside of criminal proceedings that trigger due process privileges. Limits on due process – particularly limits that presume the normative and logical validity of prioritizing security over liberty when security benefits of due process limits have not been factually established – have generated additional limits on due process by denigrating rule of law. One possible common cause of this denigration across many carceral programs including lie detection is increasing authoritarianism in response to perceived threats. Authoritarianism is associated with aggressive punitiveness to outgroups, which characterizes many state responses to marginalized groups such as Muslims, immigrants, and political activists after 9/11 (*Patriot Acts: Narratives of Post-9/11 Injustice* 2011). Outgroups are who we tend to disproportionately think are lying, regardless of actual truth-telling (Levine et al. 2000). In this way, lie detection programs arguably make aggression toward outgroups seem scientific and neutral. In parallel, mass surveillance programs more generally systematize outgroup aggression in apparently professional but actually discretionary ways. They do this by making conformity with ingroup norms seem like an important value to internalize, because privacy rights do not protect ordinary people’s internal states, records, personal correspondence, movements, or beliefs. In addition to exemplifying this massive moral-political shift, polygraph programs matter in and of themselves because of their widespread and increasing use, as I describe in greater detail in the following section.

Current Uses of Lie Detection

Polygraphs might seem exotic, limited to the worlds of spies and criminals. Nothing could be farther from the truth. Millions of exams are conducted in the U.S. annually,

with billions in direct costs – the majority of which come from taxpayers for various public sector polygraph uses. Far from being a niche concern, polygraphs are common and growing in several realms.

EPPA significantly limited private sector employment polygraph use. This means most private employers can no longer require applicants or employees to submit to psychophysiological tests of deception as a condition of employment or to avoid discrimination. Many predicted EPPA would kill the polygraph industry. At the time of the law’s passage, about 75% of polygraphs were employment-related (Jussim 1985). The majority of polygraph tests were for private employers, followed in pre-EPPA frequency by polygraph use in local criminal justice investigations, and then by federal governmental use (Office of Technology Assessment 1983).

But instead of dying off after EPPA’s enactment, polygraphy has thrived. Evidence suggests that large and steady expansions in public sector uses of polygraphy have carried annual domestic polygraph administrations well beyond their pre-EPPA peak of two million (Sening 1989).² For example, the number of U.S. polygraph schools accredited by the practice’s primary professional organization (the American Polygraph Association) rose from 11 in 1995, to 17 today (Barland 1995). APA membership is also rising significantly because of foreign membership increases.

The U.S. federal government is the single largest polygraphy client globally, driving market demand for the services polygraphers certified by these growing schools

² I estimate their use at about 2.5 million polygraph exams administered annually in the U.S. This is a guessing game, because systematic data is not available, for instance, from intelligence agencies on their polygraph use frequency in pre-employment and employee screenings, or from local and state police departments on their polygraph use in administrative and criminal investigations. One recent estimate asserts more than a million polygraph exams are given in the U.S. each year (Gardner 2009). This estimate says it includes both criminal justice and private employment uses, but it appears to exclude federal governmental polygraph exams – i.e., the vast majority of such exams.

will provide. But this insight collapses the heterogeneity in a wide variety of diverse public sector polygraph programs affecting the lives and life chances of a staggering array of federal public servants, citizens applying for those positions, U.S. allies, local and state-level law enforcement officers and applicants, criminal suspects, convicted criminals, immigrants seeking official status – and potentially anyone traveling through a busy airport or train station. Behavioral, verbal, and especially psychophysiological threat, credibility, or deception assessment systems that extend polygraph screening methods to transportation and border security screening contexts touch the realms of everyday freedom of movement, expression, association, and privacy for all.

Despite limited efficacy data – none involving controlled trials under field conditions – and public outcry, deception detection screenings using these systems are already at varying stages of regular use (Furnas 2012). The Department of Homeland Security’s Transportation Security Administration uses a program called SPOT (Screening of Passengers by Observation Techniques) in some airports and other transportation hubs. Building on behavioral analysis techniques long taught in federal polygrapher training, SPOT operates on the premise that nonverbal, behavioral cues can be used to reliably detect deception. Drawing from techniques long taught as part of polygraph interview and interrogation at the federal polygraph training institute, SPOT relies on about 2,800 behavior detection officers at 176 U.S. airports to interpret visual cues such as fidgeting, sweating, and negative affect to assess possible security threats. SPOT costs \$200 million/year, and TSA has spent around \$900 million on the program since its inception in 2007 (U.S. Government Accountability Office 2013). SPOT has never identified a security threat, although in one year it resulted in 199 arrests (for, e.g.,

drug trafficking or immigration violations) from 2,214 police referrals out of 37,370 passengers screened. TSA has yet to produce sufficient scientific evidence on the program’s efficacy, and after issuing multiple reports criticizing this shortcoming, the Government Accountability Office has called for limits to its future funding. Specifically, a May 2010 GAO report noted an absence of scientific evidence for using passenger behavior and appearance to identify security threats (U.S. Government Accountability Office 2010). A June 2011 GAO report noted DHS’s efforts to validate SPOT continued to be significantly hindered by methodological design flaws (U.S. Government Accountability Office 2011). And in November 2013, another GAO report conducted an extensive literature review on the science of lie detection, concluding that SPOT’s operating premise – that nonverbal behavioral indicators can be used to detect deception – was not supported by peer-reviewed research (U.S. Government Accountability Office 2013). Rather, behavioral lie detection success rates appear to be about as good as chance (54 percent). The most recent GAO report addressing SPOT also criticized DHS’s continued failure to evaluate the program’s efficacy, including a 2011 attempt to validate SPOT using unreliable data and poorly defined outcome measures.

SPOT is not *sui generis*. Rather, it is part of a group of well-funded, ongoing governmental programs using techniques from polygraphy to screen the general public. DHS has also field-tested a next-generation polygraph for this type of public security screening setting called FAST (Future Attribute Screening Technologies). FAST incorporates some of the same measures of physiological arousal as polygraphs, such as changes in heart rate and breathing patterns, body movements, and eye contact, in addition to some new ones, such as body heat fluctuations, eye movements (i.e., blink

rate and pupil dilation), and changes in speech patterns (i.e., vocal intonation, pitch, and meter). Like a lie detector, FAST measures a variety of physiological indicators, ranging from heart rate to the steadiness of a person's gaze, to judge a subject's state of mind. But there are major differences from the polygraph. FAST relies on non-contact sensors, so it can measure indicators as someone walks through a corridor at an airport unaware that they are being evaluated in this way, and it does not depend on active questioning of the subject. It is thus like a wireless polygraph machine with more measures. DHS has collected personally identifiable data on the public using FAST, including video, audio, and psychophysiological recordings, since at least June 2010 (U.S. Department of Homeland Security 2010). FAST is shrouded in greater secrecy than SPOT, and cost estimates on the program do not currently appear to be publicly available. Some officials say it is based largely on Paul Ekman's discredited ideas about microexpressions and that TSA has no plans to deploy it (Higginbotham 2013). But Ekman's work actually relies on a far narrower range of measures and technologies than FAST, and other sources say it is in increasingly wide field use but concerns about public opposition to the program loom large.

Similarly, AVATAR (the Automated Virtual Agent for Truth Assessments in Real-time) asks travelers simple questions, assessing responses for signs of deception – changes in the voice or heart rate, body language, and facial expression – that may indicate stress. AVATAR has been in field use at a border crossing in Nogales, Arizona since 2011 (O'Reilly 2012; Greenemeier 2012). Despite DHS paying for its development and field testing, CBP (a DHS component) is purchasing AVATAR and its implementation as a product and service for an estimated \$6.3 million in 2012 alone

(Nunamaker and Golob 2012). In 2013, AVATAR was slated to expand to the northern border, generating an additional projected \$12.4 million in product sales and \$3.6 million in annual services. In 2014, AVATAR is slated to expand to international air travelers, generating a projected \$12.9 million in product sales and \$3.8 million in annual services. That is just for the Trusted Traveler program. When CBP upgrades only the 50 busiest American airports' basic self-service Global Entry kiosks to AVATAR kiosks, that move will generate a projected \$45 million in product sales and \$13.5 million in annual services. Over the next few years, AVATAR developers also hope to tap the government insider threat and private sector security markets. All told, estimated revenue potential for AVATAR from CBP alone is \$115.6 million in product sales and \$34.2 million in annual services. In addition, the DoD has at least 500 priority military bases that would require an average of 10 AVATAR systems (valued at \$500 million) plus annual services such as implementation support (valued at \$150 million). The private screening market value of AVATAR is the polygraph market, which AVATAR's creators estimate at \$3.6 billion in annual revenue. Having completed an international field test in cooperation with EU border guards at a Romanian airport in December 2013, AVATAR appears to be rolling out ahead of schedule (BORDERS National Center for Border Security and Immigration 2013).

If its infamous reign over air travelers' shoes, liquids, and genitals is the best-known and broadest-reaching visible manifestation of the surveillance or national security state, the role of the sprawling federal security apparatus as an employer of millions is the next step down in notoriety and size. Roughly 854,000 people hold top-secret security clearances as federal law enforcement or security state employees (Priest

and Arkin 2010; Priest and Arkin 2011). Obtaining and maintaining a top secret clearance usually requires at least one and routinely more polygraphs, although there is a TS/SCI clearance without a polygraph. The number of positions requiring a TS/SCI clearance with polygraph has greatly increased since 9/11, and over 50% of candidates for all cleared jobs are rejected through polygraph exams (Geracimos 2002; A. Becker 2013). Lower security clearance levels can require polygraphs as well. As a result, recruit and employee security screening polygraphs rose nearly 750% in the FBI from 2002 to 2005 (U.S. Department of Justice 2006). Much to the chagrin of the scientific community, polygraph programs also play increasingly prominent roles at national labs (Aftergood 2000; Department of Energy 2006).

Yet neither an increase in the number of applicants for an increasing number of post-9/11 security state jobs, nor increasing frequency of polygraphing existing employees (or vetting of foreign sources) under existing programs, completely accounts for this sort of astronomical increase. Rather, sweeping new polygraph requirements have also contributed to the fantastic growth of post-9/11 federal employee polygraph programs. For instance, in late 2008 about 5% of U.S. Customs and Border Protection (CBP) applicants were polygraphed. This figure increased 10% in 2009, even though CBP polygraph failure rates were substantial, with estimates ranging from 60% to 99.7%, or 340 of 341 (Archibold 2010; Kalish 2010). Despite this marked trend, Congress passed the Anti-Border Corruption Act of 2010, mandating that by 2013, all CBP applicants be polygraph-tested before hiring (U.S. Congress 2010). If a 95% increase in polygraph exams sounds quite large in percent terms, translating it into raw numbers does only unscours its magnitude. CBP began a hiring surge in 2006 that would eventually add

around 17,000 employees – making it the single largest U.S. law enforcement agency (A. Becker 2013). GAO recently encouraged CBP to follow through on its plans to assess the feasibility of expanding polygraph screenings from all prospective to current officers and agents, in a 2013 report conducted to help evaluate and enhance integrity-related CBP programs in response to DHS statements that drug-trafficking organizations attempt to infiltrate or otherwise corrupt CBP (Government Accountability Office 2013).

Federal governmental use of polygraphs also involves aggressively exporting lie detection programs to strategic allies. Thus there is a large and consistently expanding global network of U.S. polygraph training and equipment recipients in the form of other governments. This network is both formal and informal. Formally, polygraph programs are required as part of U.S. sponsored anti-corruption programs such as Plan Colombia, the Mérida Initiative in Mexico, and others in the Bahamas, Bolivia, Guatemala, Honduras, and Iraq (U.S. Government Accountability Office 2010; U.S. Department of State 2010). Informally, the U.S. supports strategic allies through selectively favorable polygraph export policies, as in the cases of Singapore, South Korea, Japan, Australia, and Taiwan. The American Polygraph Association accredits as many polygraph training programs outside the U.S. as it does within. Conversely, we antagonize our strategic competitors, such as Russia and China, by forbidding the export of polygraph equipment to these countries, ostensibly because of fears of human rights abuses. As a whole, these international export patterns reflect the exercise of U.S. federal governmental power through concerted efforts to shape the bureaucracies of client states and competitor powers.

In addition to shaping bureaucracies in the federal security state and abroad, the trend of consistently growing polygraph programs applies more broadly at home. Local and state police increasingly use polygraphs for administrative inquiries, criminal investigations (including on alleged victims and witnesses), and to screen recruits. This trend has generated some push-back. For example, the 2005 Violence Against Women Act threatened police departments that continued to make the victim-witness passing a polygraph a condition of proceeding with a rape investigation with loss of some federal funding (Archambault and Lonsway 2008). In a national survey, about 37% of rape crisis centers reported police asked victims to take a polygraph before they would conduct an investigation – leading most often to the withdrawal of victim cooperation, in other cases to police conducting no investigation after the victim “failed” or refused the polygraph, and in at least one case to police arresting the victim for failing the test (Sloan 1995). But VAWA did not authorize systematic data collection on the prevalence of this type of use of polygraphs, so the continuing prevalence of this type of polygraph use is hard to gauge.

The best systematic data on police use of polygraphs comes from the Law Enforcement Management and Administrative Statistics Survey (LEMAS), a Bureau of Justice Statistics survey that collects nationally representative, state and local law enforcement agency data about every three years. LEMAS has collected data on pre-employment polygraph use from 1997-2007. Nationally representative state and local law enforcement agency use of polygraphs has increased nearly 10 percent over this 10 year period. About 41 percent of departments now require recruits to pass a polygraph screening before hiring. Another form of lie detection program, voice stress analysis, also

grew steadily in police recruit screening use during this period. The overall growth rate of voice stress programs was nearly 6 percent, but their starting prevalence was much smaller than that of polygraphs. In 2007, about 9 percent of agencies used voice stress to screen recruits. The existence of internal affairs units and authorized collective bargaining in law enforcement agencies negatively determine lie detection programs, suggesting unionized police departments and those that already have an organized way of dealing with misconduct tend not to institute such programs as much as other departments.

Although the best available systematic data on police use of polygraphs deals with pre-employment screening, polygraphs are used in a diverse array of additional carceral/surveillance contexts. For example, federal detainment and interrogation practices from CIA black sites to immigration detention centers are heavily influenced by polygraphers’ supposed expertise in interrogation methods and sometimes include polygraphs (Dow 2004; Priest 2007). Previously unreleased government documents I obtained under the Freedom of Information Act show that, despite polygraphers officially registering concerns about the worth of polygraph interrogations conducted with interpreters under unstable conditions, the U.S. military interrogated thousands of post-9/11 overseas detainees using polygraphs from at least 2004-2008 (National Security Counselors 2013).³ Thus illegal immigrants and suspected terrorists are increasingly subject to polygraph testing. In addition, convicted criminals, especially those subject to

³ I began submitting Freedom of Information Act (FOIA) requests to multiple federal agencies for documents and data relating to polygraph programs in 2009. Over the next few years, I discovered what organizations like the ACLU and EFF already knew: requesters frequently have to sue to obtain data under FOIA. I could not so much as obtain processing notes on my own FOIAs, much less responsive records. My requests and related files were repeatedly ignored or lost altogether, before I retained a national security lawyer with expertise in FOIA and polygraphs. All subsequently obtained documents are available online at http://nationalsecuritylaw.org/document_vault.html.

post-conviction sex offender treatment programs as a condition of parole, are increasingly subject to polygraph testing that includes interrogation about their sexual fantasies – a literal inquiry into thoughtcrime (J. V. Becker and Murphy 1998; Branaman and Gallagher 2005; Meijer et al. 2008). (Polygraph interrogations of sex offenders that included required questions about sexual fantasies date back at least to the 1980s. It is the number of people subjected to this questioning, not the scope of the questioning, that has recently changed.) Finally, immigrants seeking asylum or other special status can introduce polygraph evidence of their truthfulness in some jurisdictions. Illegal immigrants, suspected terrorists, convicted sex offenders, asylum seekers – low-status, deviant groups without the full protection of the law. Along with some high-status public servants including, for instance, FBI agents, these people have fewer due process protections than ordinary criminal suspects or defendants. But they have not made a willing trade-off of one privilege (e.g., privacy of internal state) for another (e.g., serving their country). So the classes of exceptions to the protection EPPA extended most private sector employees are diverse and numerous.

Polygraphs are not only common and chronic, but also costly. While computerized polygraph and voice stress scoring algorithms have recently multiplied, computerized scoring methods and other forms of automation of the screening process are not the norm in practice. Thus the costs of conducting polygraph tests have not decreased in response to evolving technology. And polygraph testing is not cheap. Cost estimates tend to be three-digit for an individual. At the organizational level, CBP alone spends about \$3 million a year implementing its polygraph program, citing an \$800/exam cost (A. Becker 2013). Extending this cost estimate to the approximately 2.5 million

domestic annual polygraph exams, polygraphy within the U.S. alone is a \$2 billion/year industry. This estimate does not include larger deception detection programs such as SPOT, FAST, and AVATAR. With SPOT’s \$200 million/year operating budget and numerous other such programs under development or in use, and with alternate lie detection technologies such as voice stress proliferating, deception detection is a \$3-4 billion/year domestic industry.

The (Tempting) Wrong Puzzle: Puzzling Over Lack of Evidence-Based Evidence

As the section above detailed, polygraph programs and their next-generation deception or threat detection tools and techniques are in widespread and increasing use across the American federal government, local and state policing, and an array of other legal, administrative, and security contexts. Their use is common, chronic, and costly.

Polygraphs are more common than even their widespread use suggests on its surface: approximately 2.5 million polygraph exams are conducted annually in the U.S., and anyone who has traveled through a major U.S. airport in the past few years has likely been subjected to at least one next-generation deception detection screening program. Polygraph programs are also chronic, in the sense that their overall use seems only to grow. This growth pattern stands in contrast to nearly 100 years of judicial, legislative, and executive branch powers constraining these programs due to their insufficient evidence base and history of abuse. In particular, post-9/11 growth of public sector polygraph programs has been tremendous and sustained. And these programs are costly: by my conservative estimate, lie detection is an over \$3 billion/year industry within the U.S. alone, and the vast majority of associated expenses are paid by taxpayers.

Yet, as scientists have lamented since the inception of modern lie detection, there is at best insufficient evidence supporting the efficacy of polygraphs in detecting or deterring crime or misconduct (Iacono and Lykken 1997; Iacono 2001; National Research Council 2003; Lykken 1998). Indeed, there is insufficient scientific basis for asserting that there exists a psychophysiological, behavioral, or other known deception response to detect in the first place – or that it is even possible to conduct a proper validity study of lie detection under real-world conditions (Saxe and Ben-Shakhar 1999; Saxe 1991; C. F. Bond and DePaulo 2008; C. F. Bond et al. 1992). Little wonder, then, that people tend to detect deception at rates well below chance – and trained experts tend to fare notably worse than uninitiates (Kassin and Fong 1999; Meissner and Kassin 2002; Levine, Kim, and Blair 2010; Vrij and Semin 1996; Zuckerman, Koestner, and Alton 1984; Zuckerman, Koestner, and Driver 1981). Voice stress analysis correctly identifies only 15% of deceptive subjects on average (Damphousse et al. 2007). Published field studies on polygraphy claim to detect deception at rates far greater than chance, but Monte Carlo simulation shows the confidence interval for the positive predictive value of polygraphy in the same field studies is actually 45%-60% (Zelicoff 2007).

Lie detection lacks sufficient evidence supporting its use in security screenings and criminal proceedings, but it fits right in with other forensic procedures that lack scientific basis (National Research Council 2009). Legally credible forms of forensic evidence (including DNA fingerprinting, fingerprinting, arson "signs," rape kits, interrogation confessions under the commonly utilized Reid technique, and traffic radar) have fundamental validity problems similar to those of polygraphy. First, individual human beings interpret all these forms of evidence. It is a common fallacy that forensics

are objective in the sense that they are somehow self-interpreting and immune to biases in perception. Second, many commonly known "facts" such as the unique nature of fingerprints, the particular "signs" of arson and rape, and the physiological arousal measures that are taken to proxy for deception in polygraphy, have not actually been established as scientific fact. Third, standardized training and accreditation/certification or continuing education programs are not mandatory and in some cases do not exist for forensic scientists, rendering the administration of forensic tests such as polygraphs significantly heterogeneous.

Polygraph evidence, along with other forensics such as ballistics, tool mark, and bite mark evidence, is generally considered less legally credible than many other forms of forensic evidence. In most states, polygraph charts and chart interpretations are not usually admissible in criminal court, and there military courts ban polygraph evidence completely. However, polygraphs can be utilized in a wide variety of contexts explored in greater detail in the following section, including in criminal and administrative investigations to interrogate suspects, witnesses, and victims, and in sex offender treatment and parole decisions, insurance fraud and child custody civil litigation, and asylum/immigration hearings. In addition, polygraph interrogation confessions (like most interrogation confessions) are admissible in criminal court. Polygraphs on the whole are more widely used than other any other single forensic tool as law enforcement, military, and intelligence/security agencies and numerous foreign governments routinely use them to interrogate criminal suspects, witnesses/sources, parolees, applicants for employment, and employees. In this way, polygraphs exemplify the limitations and broad significance of unscientific forensic evidence.

Little is known about how this disconnect between forensics and science came about and why it persists. One might well ask, therefore, why lie and threat detection programs based on observation of physiological responses are widespread and growing. On the surface, this seems like a classic political developmental paradox: Why would a way of producing evidence that is not itself evidence-based experience consistent, and in some contexts staggering, growth?

As public policy scholars and others know, however, a question about why interventions without scientific basis persist and often grow in widespread use is the wrong question to ask about any public policy intervention. Sadly, social and political interventions are simply not evidence-based as a general rule (Campbell 1998; T. D. Wilson 2011). Just as medical interventions in days of yore were not tested using randomized experimental trials to generate reliable and valid estimates of their causal effects before their widespread implementation, so is public policy today based more on common sense than empirical testing. Indeed, an embarrassing array of expensive, widely-implemented policies and programs have recently been shown to have no effect on the outcomes they try to influence – or, worse, to actually work at cross-purposes to their intended goals.

For instance, many relief agencies and first responder organizations such as police departments commonly use a technique of emotional first-aid or psychological debriefing called Critical Incident Stress Debriefing (CISD) to help people deal with trauma. But controlled studies of CISD suggest it actually impedes the natural trauma recovery process, increasing the likelihood of developing PTSD following exposure to a traumatic event (Bisson et al. 1997; McNally, Bryant, and Ehlers 2003; Lilienfeld 2007; Sijbrandij

et al. 2006). Similarly, the well-known drug and alcohol abuse prevention program DARE (Drug Abuse Resistance Education), which has been implemented in 75 percent of American schools, has no statistically significant effects on students’ drug or alcohol use in controlled trials (Rosenbaum and Hanson 1998; S. L. West and O’Neal 2004; U.S. Government Accountability Office 2003). This is not to suggest that the people implementing these policies – such as counselors volunteering their services after communities experience violence, or educators striving to keep their students safe from substance abuse – have anything but good intentions. Rather, it is a fact of the political ecosystem. There is no FDA of laws or social programs.

So the fact that polygraphs have insufficient evidence to justify the overwhelming majority of their current uses does not by itself make these programs exceptional. Asking why polygraph programs grow is thus not the most interesting question to ask about them or any other administrative decision-making technology, many of which have better evidence supporting their use. Rather, the growth of these programs can be most usefully understood as a moral-political event through a series of watershed moments in the development of a particular legal regime interpreting evolving applications of the fact-value distinction. These applications construed that distinction as being central to what counts as expertise (in *Frye*) and what forms of self-incrimination versus privacy the Constitution protects (in *Miranda*) – without then consistently applying the same basic principle of evidentiary credibility as a precondition of the security-liberty prioritization EPPA enacted and post-9/11 surveillance state programs operate under today.

Conclusion

Lie detection programs are common, costly, and chronic. They affect millions of Americans, are global U.S. governmental exports, and cost billions of taxpayer dollars annually – despite insufficient evidence supporting their theory (deception detection) or practice (in terms of efficacy). Over the past century, courts, legislatures, and executive branch powers have tried to limited lie detection and related interrogation practices. However, their overall use seems only to grow. This disconnect, far from making polygraphy an outlier among programs with strong public sector growth patterns, places lie detection in good company with widespread interventions such as D.A.R.E. that do not work and may do harm, but were implemented in a large scale before being scientifically tested.

What does make the political development of lie programs exceptional is the convergence of watershed moments in the development and growth of polygraph programs with major turning points in the political development of the surveillance state. As explored in the preceding sections, these convergences and their interplay shed novel light on the evolution of concepts and practices surrounding core politic concerns such as the construction of evidence (and thus of truth), the definition of coercion and political institutional attitudes toward it, and the limits of legitimate government with respect to what substantive versus procedural rights due process implies and for whom. The legal history of interrogation is thus the legal history of the development of the contemporary surveillance state.

g Factors for G Men? How Law Enforcement Hiring Tools
Affect Police Department Diversity and Brutality

“Bureaucracy was the counterpart of cancer, it grew bigger and destroyed everything except itself” (Puckoon, Spike Milligan).

Introduction

This dissertation presents a sweeping analysis of administrative technology as an evolving way of understanding expertise and the Enlightenment fact-value distinction that has tremendous implications for politics from the construction of state power to the psychology of street-level bureaucrats. The last chapter examined the development of the federal legal regime surrounding polygraph programs, and the relationship between this history and the political development of the surveillance state. This chapter examines the next level of analysis down in a sequence of descending levels of analysis, estimating the institutional-level effects of some (more and less) technology-mediated screening tools including polygraph programs.

Local and state police departments nation-wide have recently increased their use of several pre-employment screening tools including polygraph testing as part of efforts to create more highly skilled, professional workforces. Such measures may impact other characteristics as well, such as the organizational diversity of police departments, and the rate at which citizens complain about excessive use of force. But the effects of these tools on departmental diversity and brutality have not been established. Police diversity matters, because representativeness of political institutions is a measure of equal opportunity. Police brutality is important both as an extreme abuse of public trust, and as a proxy measure of police misconduct. Thus these measures, which matter normatively on their own, also matter as measures of whether pre-employment screening tools that police forces implement in order to increase professionalism are effective at doing so. Analysis of nationally representative police survey data using coarsened

exact matching and difference-in-differences regression generates novel evidence of the causal effects of these tools on police diversity and brutality. My findings suggest some programs are better than others at increasing diversity or decreasing brutality. But better measures of brutality are needed to better assess the latter set of effects.

The socio-historical context of increasing police use of pre-employment tools is ironically tied up in public trust (or mistrust) of governmental power. Bureaucracies are not popular in America today, and have not been for some time (Kaufman 1981; Hibbing 1995; Rourke 1987). Like all systems of authority, they have always been subject to some degree of condemnation and mockery. But being bureaucracies, they have tended to respond in a relatively organized fashion. One characteristic response involves greater apparent quantification and standardization (Porter 1996; Guldi 2011). This standardization of bureaucratic decision-making, as if to limit the discretionary power of individual bureaucrats and thus enhance administrative competence and neutrality overall, is an old trick with some new instantiations in the form of administrative decision-making technologies that actually do increase the legal-rational quality and scientific accuracy of otherwise bias-prone discretionary decisions in some contexts – as I establish through six original survey experiments in the next chapter. At the same time, these modes of standardization are also themselves vulnerable to bias. For example, in the next chapter I present survey experimental research suggesting that “lie detectors” (or polygraphs), increasingly used to screen police recruits, are vulnerable to confirmation bias, or bias stemming from the interpreter receiving prior extraneous information about a subject, like reading the subject’s background investigation. More broadly, any bias these modes of standardization institutionalize is particularly dangerous because it is difficult for its victims to render it visible, precisely because the systematic bias and unfair outcomes it may generate stem from apparently

procedurally fair, neutral, and even scientific processes. In this way, new tools might elevate old biases to the level of (seeming) science, or at least fairness in the sense of freedom from prejudice (as distinct from actual accuracy). This reproduction of power through apparent neutrality and especially through the credibility of science has long been a problem in the production of knowledge in general and the operation of bureaucracies in particular (Lipsky 1980; Unger 1976; Scott 1998). What is new here is the fine-grained empirical test of hypotheses stemming from the possibility that some of these biases – if present in employee screening tools – might adversely affect the quality of the very bureaucratic workforce they are intended to improve.

Such effects would be particularly worrisome in light of the transformative twentieth- and twenty-first century growth of the “law and order” state. This highly racialized state is unprecedented in scale and punitiveness among contemporary nations, changing the nature of American elections, censuses, and civic life in general (Gottschalk 2006; Gottschalk 2008; A. Lerman and Weaver 2014; Lerman and Weaver 2012; Weaver and Lerman 2010). The growth of state and local law enforcement personnel reflects this broader trend of carceral/surveillance state growth. Between 1992 and 2008, the number of sworn officers grew annually at an average rate of 1.6% – higher than the average annual 1.2% growth rate of the U.S. population (Reaves 2012). Yet, scholars have not evaluated how law enforcement screening tools affect departmental diversity or brutality. This is surprising insofar as police brutality proxies for police misconduct broadly, providing a test in this context of which screening tools best deter or weed out “bad apples” – information that would help police departments improve their budgetary as well as hiring decisions. The troubling possibilities that screening tools that seem to standardize hiring decisions might actually institutionalize bias and degrade the quality of the workforces they are

meant to improve motivate the question: How do law enforcement screening tools affect police department diversity and brutality? Do they have inegalitarian or perverse effects? Do they seem to measure g factors for G men – underlying general ability factors relevant to job performance – improving workforce quality as intended by measurably decreasing brutality? Or do they lack significant effects at all?

Data

The core data used to address this question are from the Law Enforcement Management and Administrative Statistics Survey (LEMAS). LEMAS collects nationally representative, state and local law enforcement agency data. It is sponsored by the U. S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics and authorized by the Omnibus Crime Control and Safe Streets Act of 1968, as amended (42 USC 3732). LEMAS has been conducted every 3-4 years since 1987. Data on the independent variables of interest (law enforcement selection tools) has been collected since 1997, meaning that it was collected in 1997, 1999, 2000, 2003, and 2007.

LEMAS uses a stratified random sample, so the largest municipal departments which together employ over 50% of American police are overwhelmingly likely to be consistently sampled. This facet of LEMAS underscores the heterogeneity of American law enforcement agencies. To account for variation in departmental characteristics such as size that could influence outcome measure changes, key covariates are added to a matching framework using data from the FBI’s Uniform Crime Reports and the Census Bureau’s American Community Surveys. These variables, the matching framework, and the methodological approach within which they are situated are discussed in greater detail in the following section.

Aggregate LEMAS response rates exceed 90%. A respondent from a randomly selected agency within this sample frame takes about three hours to answer the survey, usually by mail although electronic submission is encouraged.⁴ Key survey measures include use of screening techniques in selecting new officer recruits, number of sworn full-time officers and their racial and gender characteristics (diversity), and sustained and total citizen complaints of excessive use of force received during previous year (brutality). LEMAS, like any national survey, is an evolving instrument. The survey began measuring use of screening techniques in 1997, and brutality in 2003. Analyses are subsequently limited to 1997-2003 (for descriptive statistics regarding screening tool use trends) and 2003-2007 (for estimates of the effects of these tools).⁵

The question wordings for diversity and brutality variables are straightforward. The 2003 survey instrument requests a numerical value for "actual FULL-TIME SWORN personnel with general arrest powers (3a) by RACE and GENDER," with columns for male/female and rows for racial subgroups. The 2007 instrument has nearly identical question wording, but disaggregates the response prompts so that racial and gender subgroups are not linked.

Similarly, in relation to the brutality measures, the survey requests a numerical value for "TOTAL use of force complaints received," along with a numerical breakdown of those complaints in the following categories: unfounded, exonerated, not sustained, sustained (defined as sufficient evidence to justify disciplinary action against the officer(s)), pending, and other dispositions. The 2007 use of force question wording is similar ("Enter the current dispositions for all formal citizen complaints received during 2006 regarding use of force"), with response

⁴ BJS uses median value or ratio imputation to minimize losses, e.g., imputing the median value of a ratio reported in the contemporary survey administration by other agencies in the same sample cell for numeric items to which an agency did not supply a response. Estimates of sampling error, chance variation from surveying a sample instead of the entire population, are measured by an estimated single-digit standard error that varies every survey administration by size of estimate and of base population.

⁵ The necessary geographic identifiers to match law enforcement agencies for the diversity analysis are also only available in the 2003 and 2007 LEMAS datasets. So although the diversity data was also collected in 1997 and 2000, my analyses remain by necessity limited to 2003 and 2007 for both dependent variables of interest.

options in this administration specified as sustained (defined the same as in 2003), other disposition (collapsing unfounded, exonerated, not sustained, and withdrawn from the 2003 version), pending, and total. LEMAS data on police brutality were only collected in the 2003 and 2007 administrations, so analysis of this variable is confined to that period.

Both outcome variables are probably systematically biased due to their self-reported nature, but this bias gives part of the analysis more rather than less leverage. Conservative measures yield more meaningful statistical results, because they make null hypothesis tests tougher. This means that evidence against the null – if it exists – can subsequently be even more persuasive. In other words, because the conservative nature of the measures generates estimates that tend toward underestimation, tests that produce significant results produce more powerful evidence of true effects.

Again, such systematic bias does not affect the validity of the analyses comparing differences in otherwise similar departments that implemented a given screening program with those that did not. I will return to the details of the methods in the following section. However, it is worth noting that substantively, the diversity variable might tend to be an over-estimate and the brutality variable an under-estimate. This is because local and state-level law enforcement agencies might worry that low diversity and high brutality could attract negative attention from the federal government, which sponsors the survey collecting this data and funds a lot of local and state law enforcement education, training, equipment acquisition, and other activities. The brutality measure is likely to be more significantly skewed, because it is several links down a communication chain involving the very authorities implicated in the abuse. Scholars have noted that this type of complaint data yields a significantly limited, proxy measure of how police harassment and violence, and the attendant complaint processes themselves, affect citizen trust of

police (Tyler 2006). Nonetheless, LEMAS provides the only national measure of excessive police use of force currently available. The need for better measures is a pressing research priority. But given available data constraints, analyzing these imperfect measures lets me conduct a first assessment of any apparent relationship between police selection tools and departmental diversity or brutality. Future research should use better brutality measures in particular to tease out mechanisms of any results that do exist, as I discuss in greater detail later in this chapter.

The outcome measures of diversity and brutality are log-transformed in analyses so that the data are more normally distributed and therefore more in accord with estimating assumptions. Specifically, diversity is operationalized through a set of measures as the log of the proportion of full-time sworn officers in various minority groups (e.g., black, female) out of the total number of full-time sworn officers in a given agency. Brutality is operationalized through a pair of measures as the log of sustained citizen complaints of excessive officer use of force, and as the log of total citizen complaints of excessive officer use of force. Constructing multiple measures for each variable enables more finely honed analyses of causal effects.

Police diversity matters, because lack of diversity can harm organizational efficiency and thus jeopardize the law enforcement mission to protect and serve (Callum 2001; Richard 2000; Sommers 2006). Moreover, representation of women and racial minorities in law enforcement agencies since the 1960s is a story of hard-won, significant gains despite continued underrepresentation, followed by plateauing or even regression since the mid-late 1990s (Sklansky 2006). The immediate historical context of these recent trends are the urban riots of the 1960s, which led to the 1967 President's Commission and 1968 Kerner Commission reports recommending increasing hiring of minority police officers (Hickman and Piquero 2009). Yet

despite strong social and political pressures, American police tend to not be representative of the populations they serve, and gains in police diversity over the past fifty years have been tenuous.

Below I present a brief, novel update of aggregate trends in police diversity using LEMAS data since 1997. The trends shown here are consistent with what is already known about minority representation in American police forces. Minorities and women remain under-represented in state and local law enforcement agencies. Because the distributions of proportions of minorities in police departments are non-normal, I log-transform them in analyses but report raw mean proportions for descriptive statistical clarity. Table 1 below also reports their variance, medians, and interquartile range in addition to their means. I issue a note of caution regarding interpreting minority trends other than African-American over time, particularly Hispanic trends, due to changing definitions of ethnicities. Standard errors run between .000 and .075.

Table 1

Racial/Ethnic and Gender Diversity in Nationally Representative Law Enforcement Agencies

		1993	1997	2003	2007
Female Proportion	Mean	0.052	0.053	0.063	0.090
	Variance	0.002	0.003	0.003	0.019
	Median	0	0	0	0.036
	Interquartile Range	[0, 0.081]	[0, .083]	[0, .100]	[0, .104]
Af.-American Proportion	Mean	0.051	0.083	0.042	0.050
	Variance	0.003	0.007	0.003	0.004
	Median	0	0	0	0
	Interquartile Range	[0, 0.023]	[0, .031]	[0, .023]	[0, .294]
Hispanic Proportion	Mean	0.040	0.031	0.032	0.037
	Variance	0.003	0.003	0.003	0.003
	Median	0	0	0	0
	Interquartile Range	[0,0]	[0,0]	[0,0]	[0, .016]
Other Minority Proportion	Mean	0.016	0.015	0.018	0.018
	Variance	0.002	0.002	0.003	0.003
	Median	0	0	0	0
	Interquartile Range	[0,0]	[0,0]	[0,0]	[0,0]

Source: LEMAS 1997, 2000, 2003, and 2007. The sample size for 2007 was 2809 for all racial category proportions, but 2807 for female proportion. Minor variation in sample size for the proportion of female officers in 2007 stems from missing data.

Police diversity as a dependent variable is important in its own right as a reflection of equal opportunity and a possible factor in the perceived legitimacy of law enforcement within the communities it serves. Despite significant progress, severe outcome disparities between blacks and whites persist in relation to employment, education, and housing (National Research Council 2001; National Research Council 2002; Sidanius and Pratto 2001). Racial disparities are perhaps most stark in criminal justice contexts: racial minorities account for approximately 60% of total jail and prison populations. Black males are incarcerated over 6 times more frequently than white males and 2.6 times more frequently than Hispanic males (H. C. West 2010). Law enforcement selection tools that institutionalize bias against blacks might in various ways help explain continuing macro-structural racial disparities such as these, which contrast with formal equal opportunity at the micro level.

Police diversity might also be important as a factor in organizational culture and thus police behavior, although this point is debated. Evidence more strongly supports a relationship between police gender diversity and lower rates of police misconduct than between police racial diversity and that outcome. It might be surprising that police departmental racial diversity does not have a well-established relationship with police brutality. One might suppose that a more diverse police force would tend to behave in more tolerant ways because, for instance, of well-established associations between racial intolerance and aggressive punitiveness toward outgroups. However, existing studies on diversity in police forces are not able to determine whether police racial diversity significantly affects disciplinary problems including excessive use of force (Aamodt 2004; Sklansky 2006; Walker 2007; Hickman and Piquero 2008). In fact, racial minorities and women have *lower* police academy averages and performance ratings (Aamodt 2004). But they receive no fewer condemnations, have no more disciplinary problems, and miss

no more days of work, leading some researchers to speculate that lower minority police performance ratings could be due at least in part to supervisor bias or other factors unrelated to actual performance (Aamodt 2004).

But when it comes to police performance, not all forms of diversity are created equal. Male officers are generally over-represented in use of force complaints, just as men are over-represented as perpetrators (and victims) of violent crime in the general population. For example, women were about 16% of the NYPD in 1998, but comprised less than 10% of officers investigated for alleged brutality by the Civilian Complaint Review Board (A. S. Green 1996). Although the inverse relationship between police gender diversity and brutality is poorly understood, it provides suggestive evidence that some forms of police diversity influence police behavior in politically and socially important ways.

Like diversity, police brutality has weighty political and social significance. It is the most extreme form of police misconduct. It impairs not only the physical and emotional well-being of those subject to it, but also the trust that citizens in the affected community more broadly have in authority – and thus the respect they enact for the law by observing rules and cooperating with authorities (Tyler 2006). Thus, unjustly harsh policing, far from making citizens safer by scaring them into compliance with the law, degrades rule of law.

As previously discussed, law enforcement agency self-reports of citizen complaints of excessive use of force are probably extremely conservative and otherwise problematic as measures of police misconduct or brutality. However, they are currently the best available measures of national police brutality. LEMAS indicates the vast majority of departments report no sustained complaints. The variance, however, is quite large. And the variance of total complaints is substantially larger than the variance of sustained complaints – about 20 times

larger – suggesting that the number of sustained citizen complaints of excessive officer use of force is probably a particularly conservative measure of an extreme subset of police misconduct.

Methods

Using a recently developed matching technique in combination with difference-in-differences regression analysis (DID), I compare similar police departments that enacted and did not enact a variety of pre-employment screening tools including polygraphy to estimate the causal effects of these tools on departmental diversity and brutality. The combination of this matching technique with DID produces more accurate and less model-dependent causal inferences than traditional, commonly used modeling approaches. In this section, I describe these methods, why and how I use them, and their limits.

Matching Technique

Coarsened exact matching (CEM) facilitates the ability to find a quasi-experimental match by coarsening or collapsing the covariates, and then uses an exact matching procedure on the coarsened variables. Coarsening is generally part of measurement. For example, survey researchers often coarsen respondent age according to cutpoints such as 18-24 years old, 25-35 years old, and so on.

CEM differs from other common matching methods by specifying the degree of permissible imbalance between quasi-experimental treatment and control groups before matching (Blackwell, Iacus, King, and Porro 2009; Iacus, King, and Porro 2009; King et al. 2011). In this way, the procedure negates the need for post hoc balance checking by ensuring balance within the sample. Thus instead of comparing the composition of the treatment and

control groups as is appropriate when reporting results using other matching techniques, I talk in greater detail about matching variable selection and strata (or bin) definition.

After constructing the matching framework, CEM consists of three steps: coarsening, grouping, and dropping. First, matching variables are coarsened into intervals. CEM co-author King offers the analogy of the Columbus egg – an egg Columbus made stand on its head by blunting the end on the table. Like that egg, matching variables have their differences blunted, or coarsened, in CEM. Second, observations are grouped into these coarsened bins. Finally, bins that are not populated by members of both quasi-experimental treatment and control groups are dropped.

The matching variables I use in this analysis are: agency type, region, whether the department authorizes collective bargaining, whether the department has an internal affairs unit, racial diversity in the population served, violent crime rate in the area served, size of population served, and size of law enforcement agency. Agency type refers to whether the law enforcement agency is a sheriff’s office, local police department, or state law enforcement agency. These agencies’ resources and other core characteristics systematically vary. Region is included as three dummy categorical variables. The region categories are those used in U.S. Census and American National Election Studies: northeast, north central, west, and south.⁶ Finally, collective bargaining is an established influence on the political development of some selection tools, such as polygraph programs. Unions and civil liberty groups were influential in the passage of the federal Employee Polygraph Protection Act of 1988, which forbids most private sector employers from requiring current or prospective employees to take polygraphs (U.S. Congress 1988; Westin 2003). Thus it makes sense to control for whether sworn officers are authorized to

⁶ The Census/ANES region groupings are: northeast (CT, ME, MA, NH, NJ, NY, PA, RI, VT), north central (IL, IN, IA, KS, MI, MN, MO, NE, ND, OH, SD, WI), south (AL, AR, DE, D.C., FL, GA, KY, LA, MD, MS, NC, OK, SC, TN, TX, VA, WV), and west (AK, AZ, CA, CO, HI, ID, MT, NV, NM, OR, UT, WA, WY).

engage in collective bargaining. Similarly, selection tools might be thought of as attempts to professionalize police forces. Whether a department has a specially designated internal affairs unit can similarly be seen as an attempt to professionalize the force. Thus whether a department has an internal affairs unit is included in the matching framework as a proxy measure of agency professionalization. Racial diversity in the population served is operationalized as ratio of non-whites to total population according to county-level 2003 and 2007 intercensal data collected in the Census Bureau’s American Community Surveys. For obvious reasons, this covariate might affect departmental diversity. It also might affect brutality by changing the nature of police-community relations due to socio-historical factors relating to race in America. Violent crime rate in the area served is operationalized as city-level violent crime rate as collected in the FBI’s Uniform Crime Reports. The size of the population served proxies for urbanness and correlates highly with racial/ethnic diversity as well as violent crime rates, according to both survey and geographic data. The size of the law enforcement agency is operationalized as the number of full-time sworn officers. In addition, CEM treats missing values as values to match on.

For each measure, law enforcement agencies are split into strata or bins, and matches are only selected within bins. Most of the matching variables are easy to break into bins in ways that require minimal explanation. For instance, region has four bins because it has four categories as defined by the Census and ANES. Similarly, agency type has six variable values in the 2003 LEMAS codebook and three in the 2007 codebook, collapsing its possible values. So agency type in the matching framework has only three possible values, and these values (sheriff, local, and state law enforcement agency types) are used as the bins. Finally, the matching framework contains two binary variables: presence or absence of authorized collective bargaining and an internal affairs unit. Those variables’ two possible values define the bins.

The matching variables that require more thoughtful bin definition then, because they are subject to coarsening, are: racial diversity in the population served, violent crime rate in the area served, size of population served, and size of law enforcement agency. The bin size is computed using automated histogram methods according to the Freedman-Diaconis rule. The rule is:

$$\text{Bin size} = 2\text{IQR}(x)n^{-1/3}$$

where $\text{IQR}(x)$ is the interquartile range of the data and n is the number of observations in the sample x . This treats the histogram as a density estimator (Freedman and Diaconis 1981). It is generally more accurate than the default algorithm, Sturge’s rule, which works well for $n=200$ but less well in larger datasets like the ones used in this analysis.

The results of this matching framework are summarized in Table 2 below. I report balance checks based on the L statistic that coarsened exact matching’s authors recommend (Iacus, King, and Porro 2011). Lower L statistics mean lower imbalance. These statistics thus show that this method improves balance, although in some cases it also makes the number of observations smaller than is ideal for significance tests. This reduction in sample size is the primary cost of exact matching. Another feature that could be considered a cost is that by reducing the number of matched cases to exact matches, this method changes the scope conditions of the analysis with each application. Future research should thus investigate the relationship between scope conditions of estimates across models with different independent variables. One benefit is that there is little basis for questioning the quasi-experimental treatment and control groups’ comparability on the matched variables. Another benefit is that the cost ensures conservative effect estimates, enhancing the confidence one can place in estimated effects that do achieve significance.

Table 2

Matching Balance

Treatment	N Treated	N Untreated	Pre-match L	Post-match L
Unmatched Data				
Problem-Solving	130	463	0.149	0.031
Diverse Culture	101	478	0.010	0.066
Background Investigation	7	421	0.393	0.139
Credit Check	73	529	0.267	0.021
Criminal History Check	5	405	0.443	0.059
Driving Record	8	420	0.412	0.147
Drug Test	65	523	0.438	0.073
Mediation Skills	107	472	0.159	0.068
Medical Exam	29	547	0.297	0.048
Personal Interview	7	257	0.375	0.079
Personality Inventory	172	421	0.12	0.031
Physical Agility Test	68	511	0.164	0.049
Polygraph Exam	32	549	0.305	0.029
Psychological Evaluation	41	540	0.285	0.021
Second Language Test	25	556	0.209	0.157
Voice Stress Analyzer	31	542	0.135	0.067
Volunteer Service	116	483	0.23	0.054
Written Aptitude	89	503	0.167	0.042

Source: LEMAS 2003 and 2007.

CEM is the best procedure for this context, because it allows comparison of local and state law enforcement agencies that are extremely similar except for their implementation of particular selection tools.⁷ This comparison accounts for time invariant selection bias and minimizes model dependence.⁸ Controlled field experiments using nationally representative law enforcement agencies blocked on the variables used in this matching framework, randomly assigning those matched agencies to use or not use individual selection tools, would provide higher quality evidence for simpler analyses assessing the causal effects of these tools. In the absence of the availability of that form of evidence, using CEM before DID addresses the third variable problem typically of paramount concern in modeling approaches to causal inference. It does this by reducing model dependence, average treatment effect estimation error, and internal validity threats from measurement error. In the following subsections, I discuss in greater detail first DID as applied here and then the third variable problem in this context.

While it is the best procedure for this analysis, given the constraints of available data, CEM is not a panacea for the problems of causal inference in observational data. In addition to the potential third variable problem, it requires a trade-off between narrowing scope conditions and more exact matching, because adding more covariates to the matching framework causes more observations to drop. Thus identifying trends in the characteristics of what observations tend to drop is important for specifying the types of agencies to which I can make strong causal claims about generalization. This also means the established scope conditions will tend to be

⁷ Alternative matching algorithms include propensity score matching and nearest neighbor matching. A future article version of this dissertation chapter might use all three methods in order to assess how robust results are across these techniques.

⁸ Critics of matching contest this model dependence reduction claim (Miller 2013). But this vein of criticism is really about possible abuses of matching techniques, and not the methodological merits or limits of the techniques themselves. The intended use of CEM is for analysts to try different frameworks and see which one maximizes balance. The analyst then selects for balancing before proceeding to the analysis. This separates design from estimation. Following this intended use pattern, I fit for the best matching framework before proceeding to estimation. Adhering to separation of design and analysis in this way, making design decisions before analysis occurs, addresses concerns about possible model dependence in incorrect uses of CEM.

different for each analysis (i.e., of the effect of a particular independent variable on a particular dependent variable), because different agencies will drop. So the ability of this type of analysis to speak to known relative causal effects of different selection tools is limited by the extent to which different departments drop from different analyses, suggesting potentially different scope conditions of the resultant effect estimates. Overall, then, CEM requires a trade-off between how many covariates we balance on versus how many cases we retain. This trade-off limits the strength of direct comparative claims we can make stemming from causal effect estimates it helps establish. What this limitation requires is simply a note of caution with respect to comparability across changing scope conditions.

Difference-in-Differences (DID)

Difference-in-differences regression analysis (DID) compares the difference between the treatment group across time periods (pre and post) and the analogous difference between the control group across time periods. This is a common econometric technique with many recent applications across social sciences. One of the best-known DID studies compared fast food sector employment in New Jersey and Pennsylvania between February and November of 1992, before and after a minimum-wage raise was implemented in New Jersey only (Card and Krueger 1994). Including Pennsylvania controlled for potentially omitted variables, such as the national recession. This addressed the paramount methodological concern with most observational studies that purport to make causal inferences. Rather than decreasing employment as some economic theories would have predicted, the raise increased employment according to Card and Krueger's analysis. This finding, and other findings using this method, is contentious. However, in the absence of a more collaborative and rigorously empirical approach to implementing public

policies as randomized experiments, as Campbell proposed, DID provides an opportunity to more critically evaluate the evidence on how policies affects people’s lives (Campbell 1998).

Typical OLS model assumptions apply to DID regression analysis, as does a parallel trend assumption. This means that DID is vulnerable to omitted variable bias as a potential violation of the parallel trend assumption, in the event that some omitted variable causes non-parallel trends between quasi-experimental treatment and control groups. This omitted variable bias, or third variable problem, is the typical Achilles’ heel of modeling techniques in general. However, it only threatens DID estimation in the event of *time variant confounds* that systematically vary in different ways across the treatment and control groups (i.e., across agencies with and without particular selection tool use patterns or programs). It is unlikely but possible that such confounds systematically affect only police departments that do or do not institute particular programs. This possible problem is further discussed in the next section. In expectation, random errors cancel and systematic self-report errors do not affect difference-in-differences estimates.

Critics of DID have recently focused on another possible problem with common implementations of the technique. Because of serial correlation of standard errors, OLS estimates of the standard deviation of treatment effects tend to be biased downward. This can in turn generate false-positive results by causing overestimation of t-statistics and significance levels (Bertrand, Duflo, and Mullainathan 2004). Ignoring serial dependence in the disturbances means underestimation of the true variance. Estimating Newey-West standard errors, in addition to collapsing the information into pre and post periods, corrects for this.

A strong case can be made for drawing causal inferences from observational data analysis using the best applicable, recently developed matching technique (CEM) in combination with

DID. But this technique still has two significant limits. First, it estimates causal effects without determining mechanisms. Second, it remains vulnerable to the third variable problem. The question of mechanisms is one that future research should address with respect to the particular effects these analyses estimate. The third variable problem is addressed in greater detail below.

Third Variable Problem

One possible criticism of any observational analysis of causal effects of interventions including public policies is that, without random assignment of units of interest (in this case, police departments) to receive or not receive the treatment (here, various selection programs), some third variable could be driving observed relationships between independent and dependent variables. The question this problem implicitly poses is: Are these interventions (the implementation of various police selection tools, in this case) as good as random? Or are they endogenous to other, time variant trends?

Well-known examples of this so-called third variable problem abound. For example, we used to think post-menopausal women who took hormone replacement therapy had fewer cardiovascular events and overall better health. But randomized controlled trials, the sine qua non of experimental research, showed that correlation was worse than spurious. HRT harmed many women before these trials established its dangers. Some third variable or combination of variables, such as medical care access, self-care, socio-economic status, education, and the interplay of these factors, probably caused the women who received HRT before these trials to have better outcomes than the women who did not. We still do not know exactly what that third variable was.

Proponents of methods like CEM and DID argue that these methods solve the third variable problem. Critics argue that causal inferences are best made using simpler analytical methods on gold-standard evidence harnessing the power of random assignment. I take a cautiously optimistic middle ground. By using the best available methods on the best available data to estimate causal effects, my analyses advance the literature on law enforcement selection tools. They are a step forward. Even a reader who does not wish to entertain a causal claim on the basis of observational data analysis can appreciate the established associations as suggestive of future direction for research with potentially broad social and political impacts. Randomized controlled trials on these tools as interventions in nationally representative law enforcement agencies would generate better evidence for what may well be more accurate analyses of these tools' causal effects. Future research should conduct these trials.

In addition to taking this compromise position on the issue of how big a threat to causal inference the third variable problem is in the context of CEM and DID, I identify a few specific possible instantiations of the third variable problem in the context of this research. I suggest they do not threaten causal inference. First, self-selection of particular kinds of recruits into (or out of) particular kinds of departments might bias estimates of diversity or brutality. For example, African-Americans might stay away from known racist departments, and bullies might be attracted to departments with reputations for excessive use of force. However, these forms of self-selection do not pose probable internal validity threats to difference-in-differences estimation of causal effects, because time invariant confounders are not a threat to this type of analysis. That is, as long as the self-selection factor is constant over time, this design is not vulnerable to that threat. Since my comparisons are over a relatively short period of time (i.e., 2003 to 2007), it seems unlikely that systematic time variant confounds between treatment and

control groups threaten the analysis. What this method does not protect against is selection effects that occur with the implementation of selection tool programs and that also happen to produce or contribute to the estimated effects on diversity and brutality. For example, departments that are trending towards characteristics that produce more or less diversity or brutality might also be trending in a time-variant way toward implementing selection tool programs that then appear to produce more or less diversity or brutality, but actually do not. Again, this seems possible but unlikely. So self-selection at the level of individual recruit behavior probably does not significantly threaten results’ validity. But systematic self-selection at the level of police departments with particular (changing) characteristics implementing particular selection tool programs might be a form of the third variable problem that poses a threat to the internal validity of these analyses.

This section summarized the methods I employ in this analysis, why and how I use them, and their limitations. In the following section, I identify trends in police use of the pre-employment screening tools that LEMAS tracks and that thus comprise the interventions of interest. Then, I estimate the causal effects of these interventions on departments’ diversity and brutality using CEM and DID.

Increasing Use of Police Pre-Employment Screening Tools

Descriptive statistical analysis of nationally representative state and local law enforcement agency (LEMAS) survey data shows a trend of increasing public sector use of many pre-employment screening tools in nationally representative law enforcement agencies between 1997 and 2007. All such tools for which there are data show increased usage, as Table 3 shows.

Table 3 – *Frequency and Percentage of Representative National Law Enforcement Agency Use of Police Recruit Screening Tools, 1997-2007:*

	Frequency with percentage in parentheses:				
	1997	2000	2003	2007	% change
College degree (2-4 year)	306 (10.09)	279 (9.35)	255 (8.92)	305 (10.74)	.65
Interview	2,978 (98.12)	2,865 (95.98)	2,804 (98.21)	2,817 (99.23)	1.11
Second language ability test	NA	38 (1.27)	35 (1.23)	97 (3.42)	2.15
Analytical problem-solving	NA	NA	883 (30.96)	953 (33.57)	2.61
Criminal history check	2,948 (97.13)	2,908 (97.42)	2,828 (99.05)	2,835 (99.86)	2.73
Background investigation	2,912 (95.95)	2,884 (96.62)	2,799 (98.04)	2,820 (99.33)	3.38
Driving history check	2,843 (93.67)	2,827 (94.71)	2,771 (97.06)	2,819 (99.30)	5.63
Cultural diversity understanding	NA	NA	399 (13.98)	559 (19.69)	5.71
Credit history	NA	NA	1,870 (65.50)	2,022 (71.22)	5.72
Voice stress analysis	95 (3.13)	144 (4.82)	179 (6.27)	253 (8.91)	5.78
Medical examination	2,656 (87.51)	2,597 (87.00)	2,545 (89.14)	2,667 (93.94)	6.43
Psychological evaluation	2,220 (73.15)	2,113 (70.79)	2,143 (75.06)	2,324 (81.86)	8.17
Conflict management	NA	NA	258 (9.04)	514 (18.10)	9.06
Writing aptitude test	NA	1,650 (55.28)	1,628 (57.02)	1,829 (64.42)	9.14
Polygraph	951 (31.33)	1,028 (34.44)	1,109 (38.84)	1,159 (40.82)	9.49
Drug test	2,285 (75.34)	2,230 (74.71)	2,296 (80.42)	2,526 (88.97)	13.63
Volunteer history check	NA	204 (6.83)	237 (8.30)	589 (20.75)	13.92
Personality inventory	NA	1,119 (37.49)	944 (33.06)	1,471 (51.81)	14.32
Physical test	1,709 (56.31)	1,666 (55.81)	1,727 (60.49)	2,027 (71.40)	15.09
Sample size	3,033-3,035	2,985	2,855-2,859	2,839-2,840	

Source: LEMAS 1997, 2000, 2003, and 2007.

Results are ordered top-down from smallest to largest percent change. Table 3 yields several insights. Unsurprisingly, the smallest gains were made by tools with a very high existing rate of usage, such as criminal history checks and background investigations. These tools did not have much room to gain in use.

Surprisingly, what is at least on the face of it the lowest-cost screening mechanism – requiring a college degree – has one of the lowest use rates in addition to the lowest growth rate among screening tools. This stagnant and low use pattern contrasts with the uniquely proven value of education, mediated by experience, to predict all performance criteria except for commendations and injuries in a meta-analysis of law enforcement selection tools (Aamodt 2004). That is, officers with a four-year college degree – regardless of whether it is in criminal

justice or a related field – vastly outperform lesser-educated officers after two years on the job. Similarly, a study conducted by the International Association of Chiefs of Police, Police Administration Committee, showed that the 58% of Florida police officers with the lowest education level (which in the state of Florida means a high school diploma) accounted for 75% of all disciplinary cases and 77% of all certification revocations – the most severe administrative discipline the state can issue (Cunningham 2006). Law enforcement professionals are generally aware of this relationship between higher recruit education level and higher performance, and lament the infrequent use of education level as a recruit screening tool (K. Johnson 2006).

Given strong evidence supporting education as a police screening tool, and support among leading law enforcement professionals for the use of this tool in particular, why is its use low and barely increasing in comparison with other such tools? Market conditions on both sides probably explain this puzzle. On the supply side, generally steady unemployment rates during this time period among people who have attained a college education generate market conditions that make hiring them harder and more expensive as compared to candidates without a college degree (U.S. Department of Commerce, 1997-2007). On the demand side, police ranks have grown at a tremendous pace over the same time period, with the number of sworn officers rising at an average annual rate that exceeds the growth rate of the U.S. population (Reaves 2012).

The only other police selection tool in widespread use with a good evidence base for improving job performance – cognitive ability testing – is represented in LEMAS as the analytical problem solving assessment. Like education requirements, cognitive ability testing appears to be surprisingly under-utilized in police selection given its proven efficacy at raising the quality of workforces. In general, intelligence strongly predicts future job performance (Hunter and Schmidt 1996). In the law enforcement selection tool context, it bears

acknowledging that education and cognitive ability are related but not redundant. Research suggests that using a combination of both types of selection tools is more valid in terms of incremental validity than cognitive ability alone (Aamodt 2004).

The under-utilization of these proven effective screening tools deepens a parallel mystery: the increasing use of other selection tools which are not proven effective at improving the quality of police forces. Why are local and state law enforcement agencies increasingly employing a wide variety of recruit screening tools that cost them in time, money, and recruits, without proven benefits? One plausible contributing factor is that decreasing marginal costs may contribute to some of the overall increase in selection program use. For example, criminal and credit history checks have become easier to conduct due to advances in electronic record-keeping and police access to relevant databases. In theory, this explanation might relate to lie detection programs as well, because computerized polygraph and voice stress scoring algorithms have multiplied over this time period. However, computerized scoring methods are not the norm for polygraph tests in practice, and so the costs of conducting such tests do not appear to have decreased. Thus, it is a mystery why police screening tools with relatively well-established evidence base have low and stagnant use patterns, while tools with insufficient evidence supporting their use in pre-employment screenings have relatively high and increasing use patterns. But does this mystery matter in terms of impacting law enforcement agencies as political institutions? How do different police screening tools affect police diversity and brutality?

Results

According to analyses using coarsened exact matching and difference-in-differences regression per the framework described above, several law enforcement selection tools have statistically significant effects on departmental diversity and brutality. Some of these effects are counter-intuitive and potentially normatively concerning. However, all of them are open to interpretation in terms of mechanism, organizational and social effects, and – in the case of the brutality measures – what is being measured at the level of the outcome variables’ meanings.

Before presenting and interpreting results, I emphasize a distinction between departmental and individual selection effects in possible mechanisms of all estimated effects. Mechanisms relating to characteristics of departments that select particular tools (a form of the third variable problem discussed above at greater length) are possible but comparatively unlikely in this type of analysis. They hinge on some unknown but salient incompleteness of the matching framework, or on a component of time variance in those departmental characteristics. Causally salient time variance in this short time period is implausible. Conversely, mechanisms relating to police recruits’ characteristics need not be time variant. As a result, this type of mechanism of all estimated effects is most plausible. However, none of the possible mechanisms I suggest are mutually exclusive.

As Table 4 below shows, only psychological evaluations as a selection tool systematically affect departmental gender diversity. Departments that adopt this screening tool tend to then have more women as a log proportion of sworn full-time officers. The estimated effect size is .019 with a standard error of 0.009, $p = 0.040$. Multiple explanations for this effect exist. A relatively greater degree of psychological health (e.g., in terms of self-efficacy) might be required for women to self-select into law enforcement careers, because that is a counter-gender

normative career choice. Inversely, men who self-select into law enforcement might be at some aggregate comparative gender psychological disadvantage, or might experience a gendered form of stereotype threat under psychological testing. Alternately, perhaps departments that choose to implement psychological evaluation of police recruits as a screening tend to be friendlier to female candidates. However, the latter explanation would require a time variant friendly gender climate. It is not obvious why a change in departmental attitudes toward female police recruits would correlate with implementation of psychological evaluations. Moreover, although this analysis does not establish causal mechanisms, possible individual selection-side mechanisms are generally more plausible than possible department selection-side mechanisms for methodological reasons detailed above.

Table 4: *Treatment Effects – Diversity*

<i>Selection Tool</i>	<i>Female Effect (SE)</i>	<i>African-American Effect (SE)</i>	<i>Hispanic Effect (SE)</i>	<i>Other Effect (SE)</i>
Problem-Solving	0.007 (0.005)	-.012 (.005)*	.004 (.004)	-.001 (.002)
Diverse Culture	0.003 (0.006)	-.020* (0.006)	.001 (.004)	-.003 (.003)
Background Investigation	0.04 (0.021)	.044 (.030)	.013 (.017)	.012 (.014)
Credit Check	-0.001 (0.007)	-.002 (.007)	.000 (.005)	-.001 (.003)
Criminal History Check	-0.005 (0.026)	-.005 (0.25)	.005 (.017)	.001 (.013)
Driving Record	-0.006 (0.019)	.028 (.026)	.007 (.016)	.001 (.013)
Drug Test	0.002 (0.008)	.002 (.007)	.003 (.006)	-.000 (.004)
Mediation Skills	0.008 (0.005)	-.015* (.006)	-.004 (.004)	-.003 (.003)
Medical Exam	0.007 (0.011)	-.016 (.013)	.008 (.008)	-.007 (.006)
Personal Interview	0.002 (0.018)	.011 (.015)	-.015 (.014)	.003 (.007)
Personality Inventory	0.008 (0.005)	-.005 (.005)	-.003 (.004)	.002 (.002)
Physical Agility Test	0.012 (0.008)	-.002 (.008)	-.008 (.005)	-.007 (.004)
Polygraph Exam	-0.016 (0.011)	.009 (.012)	-.004 (.008)	-.002 (.006)
Psychological Evaluation	.019* (0.009)	-.010 (.011)	.008 (.007)	-.005 (.005)
Second Language Test	0.009 (0.009)	-.006 (.008)	-.013 (.007)	.002 (.004)
Voice Stress Analyzer	0.001 (0.009)	-.019 (.010)	.006 (.007)	-.003 (.005)
Volunteer Service	-0.002 (0.005)	-.008 (.005)	-.005 (.004)	-.004 (.003)
Written Aptitude	0.006 (0.006)	-.005 (.007)	.002 (.004)	-.002 (.003)

Note: Each reported treatment effect is difference-in-differences in representation measured as proportion of total full time force. These measures are log-transformed so that the data are more normally distributed and therefore more in accord with estimating assumptions.

* $p < .05$.

Three selection tools systematically affect departmental racial diversity, and all these effects adversely impact the representation of African-Americans on police forces. I argue a stereotype threat mechanism most likely explains all these effects. The use of analytical problem-solving ability assessments to screen prospective police decreases the log proportion of African-Americans as sworn full-time officers with an estimated effect size of $-.012$, standard error $.005$, $p = 0.016$. This finding is consistent with research on racial testing disparities, and may be explained by stereotype threat causing performance decrements (Steele and Aronson 1995; Nguyen and Ryan 2008). Such a possible mechanism suggests that the use of analytical problem-solving ability assessments as a screening tool might incorporate research on combatting stereotype threat in order to avoid institutionalizing racial disparities (G. L. Cohen et al. 2009).

Assessment of understanding of diverse cultural populations to screen prospective police also decreases the log proportion of African-Americans as sworn full-time officers. It does so with an estimated effect size of $-.020$, standard error $.006$, $p = 0.002$. This is a counter-intuitive finding insofar as it might be interpreted as suggesting that some racial minorities are screened out by tests that are intended to increase organizational capacity to interact well with diverse populations. Departments with particularly hostile environments toward African-Americans might tend to adopt this particular selection tool at a higher rate than other departments, and then hire fewer African-Americans for a variety of possible reasons. For example, perhaps they are simply not able to attract as many quality African-American candidates as departments that do not have known problems with hostile environments. However, this explanation would require the potential confound of hostile environment be time variant – an implausible feature over such a short period of time. Another possible explanation that does not require time variance is that stereotype threat might affect African-American recruits’ performance on these assessments just

as it may affect their performance on analytical problem-solving tests. While counter-intuitive, this explanation is consistent with what we know about explicit stereotype threat cues most affecting racial minorities’ test performance in other context. A test about understanding diversity (perhaps one written and scored by other-raced evaluators) is about as explicit as racial identity cues come. Again, this analysis does not establish causal mechanisms, but such possible individual selection-side mechanisms are more plausible than possible department selection-side mechanisms.

Conflict management skills assessment also decreases the log proportion of African-Americans as sworn full-time officers. It does so with an estimated effect size of $-.015$, standard error $.006$, $p = 0.014$. The stereotype threat explanation of comparatively poor African-American test performance remains plausible in this context. Conflict management skills assessment might cue racial stereotypes about African-Americans and aggression, causing cognitive decrements that – without an intervention to address stereotype threat – may be adversely impacting African-Americans’ performance on these tests, and thus their representation on police forces. But again, other explanations remain plausible. For example, departments that tend to self-select into implementing conflict management skills assessment may again be or appear to be less hospitable environments to racial minorities as compared to departments that tend not to implement these programs. Again, however, that environmental factor would have to be time variant to affect this analysis. Stereotype threat is thus again a more plausible possible mechanism of the effect.

As Table 5 shows, only polygraph tests as a selection tool systematically affect the log proportion of sustained citizen complaints of excessive officer use of force. Departments that adopt this screening tool tend to then have fewer such sustained complaints. The estimated effect

size is -15.57 with a standard error of 4.86, $p = 0.002$. Polygraph tests have no systematic effect on the total number of complaints. There are several possible explanations for this combination of estimated effects. First, perhaps polygraph proponents are correct that polygraph testing detects “bad apples” or deters future misconduct, and so there are fewer well-founded complaints of police brutality against recruits who pass polygraph tests, because these recruits go on to engage in less misconduct. However, we would expect a true decrease in the measured entity (complaints) to correlate across measures (sustained and total complaints). Indeed, if either measure could be said to be more fragile, it is the sustained complaint measure, because it quantifies a rare subset of a rare event.

A second possible group of explanations for this contrasting pair of results is that police recruits who successfully undergo polygraph screenings might be better at lying or at not admitting wrongdoing under interrogation than recruits who do not pass these tests, thus causing fewer sustained but not fewer total complaints of police brutality against recruits who pass polygraph tests. Some research suggests that deceptive subjects with significantly lower scores on the Socialization Scale of the California Psychological Inventory are more likely to pass polygraphs while truthful subjects with higher socializations scores are more likely to fail (Waid, Orne, and Wilson 1979). Selecting individuals with lower socialization through polygraph tests might thus help explain at a finer-grained level the disconnect between null effects on total citizen complaints about use of force, and a decrease in sustained complaints. Future research should hone and test competing possible psychological mechanisms of such a possible individual-side selection effect. For example, selecting for people with a black-and-white cognitive style who are less likely to believe they have done wrong (versus people with a more ambiguity-tolerant cognitive style who are more likely to experience doubt and guilt) could be

seen as distinct from selecting for people who are simply less likely to confess to having committed misconduct regardless of their moral beliefs.

Finally, police departments that self-select into implementing polygraph testing programs might systematically treat brutality complaints differently than departments that do not implement these programs. The third explanation is less likely because, like other self-selection sorts of explanations discussed above, it requires time variance of the environmental confound. Again, this time variance is implausible. In light of the methodological implausibility of a department-side selection effect and the inconsistency between the null total and significant sustained complaint results, the second explanation I suggest seems most plausible. Police polygraph programs probably cause an artefactual decrease in sustained citizen of complaints of excessive use of force, but not in total complaints, because they select on recruit characteristics that change complaint outcomes without changing brutality rates themselves.

Three selection tools systematically decrease the log of total citizen complaints of excessive officer use of force. None of these tools also affects the log of sustained citizen complaints of excessive officer use of force. These tools are analytical problem-solving ability assessments, credit history checks, and conflict management skills assessment. Of these, only credit history checks has no apparent adverse effect on departmental diversity – which is surprising, because credit history might proxy for (racialized) socio-economic status and disadvantage recruits with fewer socio-economic resources. But, unlike the other tools that appear to decrease total citizen complaints of excessive officer use of force, it does not appear to also harm departmental diversity.

Table 5: *Treatment Effects – Brutality*

<i>Selection Tool</i>	<i>Sustained Complaints Effect (SE)</i>	<i>Total Complaints Effect (SE)</i>
Problem-Solving	-2.70 (2.18)	-53.15* (23.46)
Diverse Culture	-2.03 (2.73)	2.05 (28.30)
Background Investigation	2.23 (13.6)	-23.98 (71.63)
Credit Check	-2.36 (2.93)	-102.28* (36.51)
Criminal History Check	3.14 (13.7)	31.03 (167.54)
Driving Record	2.29 (11.6)	8.13 (60.07)
Drug Test	1.20 (4.26)	10.70 (32.22)
Mediation Skills	-3.30 (2.79)	-57.14* (27.81)
Medical Exam	9.44 (7.21)	27.28 (65.99)
Personal Interview	1.66 (7.54)	-43.37 (39.95)
Personality Inventory	.277 (2.27)	18.54 (24.52)
Physical Agility Test	.372 (3.77)	8.05 (43.58)
Polygraph Exam	-15.57* (4.86)	-62.34 (48.73)
Psychological Evaluation	-.025 (5.02)	9.72 (69.36)
Second Language Test	.720 (4.59)	15.62 (46.10)
Voice Stress Analyzer	-9.26 (4.90)	-28.74 (48.05)
Volunteer Service	-1.56 (2.80)	-44.23 (27.91)
Written Aptitude	-3.26 (2.65)	-16.65 (29.10)

* $p < .05$.

Analytical problem-solving ability assessments of police recruits decrease the log proportion of total citizen complaints of excessive officer use of force. They do so with an estimated effect size of -53.15 , standard error 23.46 , $p = 0.024$. This finding is consistent with research showing that cognitive ability testing tends to raise the quality of workforces, intelligence predicts future job performance, and law enforcement selection tools relating to cognitive ability tend to improve workforce performance (Aamodt 2004; Schmidt and Hunter 1998). However, if this form of recruit screening is truly measuring a *g* factor – a single underlying general ability factor that is relevant to job performance – it remains unclear what exactly that factor is, why it does not cause a subsequent decrease in sustained complaints in addition to total complaints, and why it causes a decrease in African-American representation on police forces.

Credit history checks of police recruits decrease the log proportion of total citizen complaints of excessive officer use of force. They do so with an estimated effect size of -102.28 , standard error 36.51 , $p = 0.005$. Credit history might proxy for impulse control or some other *g* factor that logically predicts police likelihood to use excessive force. It is not obvious, however, why this decrease does not correlate with an attendant decrease in sustained complaints of excessive officers use of force. Perhaps factors other than a complaint’s veracity influence its disposition.

Conflict management skills assessment also decreases the log proportion of total citizen complaints of excessive officer use of force. It does so with an estimated effect size of -57.14 , standard error 27.81 , $p = 0.041$. There is some face plausibility that conflict management skills might decrease both police use of force and citizen perceptions of excessive use of force.

Overall, one notable trend these results show is null results for most police selection tools. Screening techniques that elevate the quality of the workforce might be expected to decrease police misconduct including brutality. But most tools do not appear to affect the best available measures of police brutality, and no tool affects both sustained and total citizen complaints of excessive officer use of force. Moreover, better brutality measures are needed to better establish what is being measured at the level of those outcome variables’ meanings. Future research would ideally include both better brutality measures, and randomized experimental implementation of the screening tools that are quasi-experimentally analyzed here.

Conclusion

Nationally representative local and state police departments have been increasing use of pre-employment screening tools, but the effects of these tools on diversity and brutality remain unclear. Using coarsened exact matching and difference-in-differences regression, I estimated the causal effects of these tools on police diversity and brutality. Most of the examined selection tools do not adversely affect police diversity, and none adversely affect police brutality by increasing citizen complaints of excessive use of force. But nor do any of these tools decrease both sustained *and* total citizen complaints of excessive officer use of force – a set of effects that one might expect a screening tool that identifies a g factor for G men to have.

As in the previous chapter’s examination of the development of the federal legal regime surrounding polygraph programs and the surveillance state, the evidence analyzed here suggests that polygraph programs have distinct effects as compared to other sorts of programs. They are the only screening tool examined that systematically decreases sustained citizen complaints of excessive officer use of force. But they do not also decrease total complaints. So the possibility

exists that they work – in the sense of detecting or deterring true misconduct. But given available evidence, it is more likely that polygraph programs instead populate police forces with people who behave differently after excessively using force – perhaps people who feel less guilt, see less moral ambiguity, or are otherwise better at withholding derogatory information during questioning. More research is necessary to test possible mechanisms of all estimated effects. This analysis was an important first step in estimating what effects warrant further study.

In the next chapter, I descend again to a lower level of analysis, testing for bias in technology-mediated public administrative tools including polygraphs at the level of individual behavioral decision-making. In anticipation of that analysis, this chapter also established that polygraphs do not appear to institutionalize racial bias at the organizational level under field conditions, providing a useful reference point for assessing the generalizability of those survey experimental results. In addition, this chapter presents novel, fine-grained empirical test of hypotheses stemming from the possibility that some common cognitive biases – if present in employee screening tools – might adversely affect the quality of the very bureaucratic workforce they are intended to improve.

Neutralizing Prejudice or Smuggling Bias in a Trojan Horse?
Race, Fairness, and Technology-Mediated Administrative Decisions

“A great many people think they are thinking when they are merely rearranging prejudices”

- William James

Introduction

This dissertation presents a sweeping analysis of administrative technology as an evolving way of understanding expertise and the Enlightenment fact-value distinction that has tremendous implications for politics from the construction of state power to the psychology of street-level bureaucrats. The preceding empirical chapters examined first the development of the federal legal regime surrounding polygraph programs – and the relationship between this history and the political development of the surveillance state – and then the institutional-level effects of some (more and less) technology-mediated screening tools including polygraph tests. Continuing down to the next-lower level of analysis, this chapter assesses when, whether, and how some common cognitive biases might affect the behavioral decision-making of street-level bureaucrats making administrative decisions with the help of technology. In a series of six original survey experiments, I analyze novel evidence on the neutrality of administrative decision-making technologies with respect to potential racial and confirmation bias.

The growth of the vast administrative state we know today is one of the core developments of contemporary American politics. The liberal claim of administrative neutrality (also known as neutral competence) – in which bureaucrats apparently pursue democratically sanctioned ends through rule-bound, rational-legal means – legitimates this apparatus (Kaufman 1956). While some internal critics of liberalism themselves

question this claim, the ideal of administrative neutrality as a facet of the liberal state toward which we want to progress is a common feature of legitimate liberal government in political liberalism (Smith 1999; Dworkin 2002; Rawls 2005; Klosko and Wall 2003). Unsurprisingly, then, decision-making technologies, designed to decrease error by increasing objective application of neutral rules, are in increasingly widespread use. Such technologies include polygraphs in policing, medical diagnosis tools in healthcare, and welfare benefits administration tools in social work. Despite their increasingly widespread use, insufficient evidence exists to establish the neutrality of these technologies. I test this neutrality at the behavioral decision-making level with respect to racial and confirmation bias. I find racial stereotypes do not tend to influence the technology-mediated decisions examined. But I find suggestive evidence across four survey experiments on polygraph chart interpretation that negative background information can be used as a decision-making heuristic in the interpretation of ambiguous stimuli, confirming pre-conceived notions while appearing to produce independent evidence. This chapter thus contributes fresh insights into administrative neutrality, race, and fairness at the individual behavioral decision-making level.

Background

Bureaucratic agents' construction of state power and social control through apparently rational-legal, but fundamentally discretionary, decisions has been a central political science concern since Weber (Weber 2004b). Yet Progressive Era thinkers including Woodrow Wilson reincarnated Saint-Simon's vision of ideal government as an “administration of things,” legitimating expansions of public administration by advancing

the ideal of neutral, apolitical policy implementation (W. Wilson 1887). Administrative decision-making technologies seem to constrain bureaucratic discretion, and indeed the appeal of apparently merit-based, systematizing, scientific, rationalizing, standardizing sorts of bureaucratic processes has characterized institutional responses to backlash against growing state power for decades if not centuries (Kaufman 1969; Porter 1996; Ross 1992; Guldi 2011). However, existing evidence does not establish the neutrality of those technologies. Indeed, much evidence on racial and intersectional (or additional outgroup status) bias suggests that, whether consciously endorsed and intended or not – and sometimes particularly when race is not a conscious consideration at all – prejudicial attitudes powerfully affect decisions and behaviors in a wide array of political and social contexts (e.g., Mendelberg 2001; Mendelberg 2008; Winter 2008; Rabinowitz et al. 2009). This evidence base motivates the question: do administrative decision-making technologies themselves institutionalize bias?

Error in decision-making technologies such as polygraphs can have life-or-death consequences, as Shabaka WaQlimi discovered in 1973. “My remembrance of that polygraph is that I was sat in the chair, certain wires were attached to my fingers and around my chest, and I was asked certain questions. And I was eager to answer it. Cause I knowed I hadn’t done a damn thing.”⁹ Shabaka, formerly known as Joseph Green Brown, was active in the Black Panthers when police arrested him for the robbery, rape, and murder of Earlene Barksdale in Tampa, Florida. Although the polygraph was ostensibly used only as an investigative tool, police stopped investigating the murder victim’s husband after he passed his polygraph. They then took Shabaka’s polygraph failure as confirmation that they had the right man, and investigated no other leads. Shabaka was

⁹ Shabaka WaQlimi, interview, June 2011.

convicted of Earlene’s murder. During his incarceration, his brother died for want of a kidney that prison officials denied him permission to donate. Shabaka spent over thirteen years on death row and came within fifteen hours of scheduled execution before his exoneration. An appeals court ruled the prosecution had knowingly allowed false testimony at trial, and the state decided to not retry the case.

Do decision-making technologies like polygraphs act as vehicles of racial and confirmation bias, and their potential interplay in so-called intersectional bias, giving old prejudices a patina of scientific validity? Or can they neutralize these biases in public administrative decisions that affect tens of millions of everyday Americans’ lives? These questions are central to the study of race in American politics and fairness in public administration. They also speak to central political developmental and normative concerns about what state centralization means in the late modern era, and how value-neutrality functions ideologically in increasingly technological state apparatus.

The Debate: Liberal Reformers versus Post-Structuralist Critics

One of political liberalism’s central tenets is a commitment to egalitarian norms, historically stemming from and conceptually associated with the Enlightenment and Progressive Era ideals of progress and objectivity. One of the central debates in liberalism, dividing classical and modern liberals, is about what that commitment to egalitarianism means in practice – e.g., whether it entails formal or effective equal opportunity. In some narratives, the answer to that question stems strictly from principle. For instance, some classical liberals argue that redistributive public policies violate a fundamental justice principle that privileges the fairness of process over the equality of

outcomes (Nozick 1974). But in the dominant liberal democratic narrative at least since President Johnson justified affirmative action with the analogy, it depends on how fair the race of life has been before the starting line (L. B. Johnson 1965). In the administrative context, this question of procedural fairness – of whether the bureaucratic process by which decisions with weighty consequences for equality in distributive terms are reached – becomes a question of whether discretionary power institutionalizes prejudice at the micro level of street-level bureaucrats’ decisions. A related question operating at a level of analysis one step lower is whether this micro-level neutrality (or its appearance at the level of procedural fairness) affects people’s attitudes about the fairness of distributional outcomes, and related racial attitudes. This chapter addresses the micro-level, behavioral decision-making question, while the next chapter addresses the latter question at the macro level of administrative neutrality’s possible attitudinal effects.

Against the Enlightenment ideal of rational-legal government replacing relatively venal church and political hierarchies, scholars as early as Weber have recognized this ideal precisely as an ideal-type unlikely to be found in its pure form in the wild (Kaufman 2006; Weber 2004a). Post-structuralist critics are only the most recent and vocal critics questioning the assumption that a functionally objective state apparatus is capable of neutrality, rather than reproducing dominant ideologies and inequalities (Adorno and Horkheimer 2002). Neutrality is construed in this context as the absence of systematic bias. Social theorists broadly agree state power is constructed by locking people into routines that produce outcomes the state wants, whether or not these routines and outcomes are functional. They give this process various names, including habitus, capillary power, caging, and routinization (Bourdieu 1977; Bourdieu 1990; Foucault

1995; Foucault 1980; Mann 1997; Mann 1993; Weber 2002). These routines are created by the maintenance of logics of appropriateness, the understandings by which people make decisions based on their role and its script in a given encounter (March and Olsen 1989).

The Danger: Technology Might Institutionalize Bias

What we know about the empirical construction of state power through these routines, from research across diverse literatures including political and social psychology, bias and equity in policing, and policy and law research suggests that they are often characterized by the use of mental shortcuts including racial stereotypes and substereotypes. Political science research suggests it is sometimes intersectional racial and outgroup substereotypes, not necessarily stereotypes about a racial group as a whole, that most strongly activate prejudice (Gilens 2000; Hancock 2007b; Hancock 2007a; Hurwitz and Peffley 1997; Peffley, Hurwitz, and Sniderman 1997; Hurwitz and Peffley 2005; Winter 2008; Mendelberg 2008). Moreover, empirical evidence abounds demonstrating the potential of cognitive biases – including racial, confirmation, and intersectional bias – to affect seemingly scientific forensic data interpretation (Lynch et al. 2008; Dao 2005; DiFonzo 2005; Giannelli 2011; Giannelli 2006; Meissner and Kassin 2002; Rozelle 2007; Thompson 2009; Whitman and Koppl 2010). Racial bias also affects simulated police shooting decisions (Correll et al. 2007). Racial bias seems to undergo a compounding interaction with confirmation bias in the form of felon status in labor markets (Pager 2007). It is not known whether racial and confirmation bias similarly affect technology-mediated administrative decisions such as polygraph chart

interpretation, either alone or in combination. A survey I conducted of state-licensed polygraphers in Virginia from Nov. 2010 – March 2011 found that about 20% of polygraphers believe some groups are more likely to fail polygraphs than others.¹⁰ So while technologies that appear to limit street-level bureaucrats’ discretionary power seem to offer a way to circumvent or overcome the normative perils of mental shortcuts, existing evidence suggests they might instead institutionalize prejudices, particularly racial and intersectional bias.

Such possible bias matters practically because polygraphy in the criminal justice/security context, medical diagnosis tools in the parapublic realm, and welfare benefits administration tools in the classic public administration setting are examples of a broad range of increasingly technology-mediated evidentiary routines through which bureaucrats and citizen-subjects have performed state power in new, apparently neutral and scientific ways since Progressive Era reforms. These reforms were part of the 19th to mid-20th century trend toward professional groups’ embrace of “technologies of distance” (Porter 1996; Guldi 2011). These apparently standardizing, quantifying, and mechanically objective tools legitimated elites, sometimes institutionalizing discretion in neutral-seeming ways. A long set of traditions of science, technology, and society, social and cognitive psychology, survey research methodology, and media studies work about related but seemingly disparate topics including standardization and state-building,

¹⁰ The response rate in this by-mail survey was approximately 33%, or $N=89$ out of 273 Virginia licensed polygraphers. The National Citizen Survey routinely obtains a comparable 33.33% response rate to their by-mail survey without follow-up. The measure mentioned here is whether respondents indicated any subgroup in response to the question: “Physiologists disagree about whether some groups may be more likely to fail polygraphs than others. What about you? Do you think any of these groups are more likely to show deception? Check all that apply.” Listed categories were: veterans, students, immigrants, foreigners, blacks, whites, Hispanics, Jews, Christians, Muslims, women, men, and homosexuals. Immigrants were the most frequently selected category. About 20% (18/89) of substantive responders indicated they thought some groups tend to fail polygraphs more than others.

attribution bias and prejudice, and bogus pipelines argues that these sorts of seemingly rational-legal decisions tend to be value-laden, as against the liberal ideal of neutrality. Alternatively, some combination of stronger egalitarian norms with new and better technologies might actually solve the problem of (irrationally) shortcut-prone discretion in administration.

Micro-Level Results: Technology-Mediated Decisions Appear Neutral

Leading theories of implicit and explicit racial attitudes predict that racial bias against blacks and particularly against negative stereotype-conforming blacks is pervasive, harmful to minorities, and affects decisions particularly strongly under conditions including ambiguity and the ability to attribute decisions to factors other than prejudice (attributional bias). By contrast, my findings show that at the micro level, apparently neutral decision-making technologies neutralize, or refrain from institutionalizing at the level of decisions, racial bias. This finding holds across three case studies of different administrative technologies in large-scale online survey experiments. In the keystone case study of polygraphy, the finding holds across a broader array of experimental conditions and data sources. Results suggest the positivist conception of a functionally objective state apparatus is an empirically valid vision with respect to the seemingly pervasive and entrenched use of mental shortcuts in the form of racial and intersectional bias.

Experimental Designs: Overview

The micro-level survey experiments this article summarizes occurred online between Feb. 13, 2012, and Oct. 6, 2013. The online administration mode avoided contamination from possible, unintended race and gender effects from an individual survey administrator (Gosling et al. 2004; Orne 2009; Osborne 2001). Participants were workers on Amazon.com’s Mechanical Turk (MTurk) platform located in the U.S. MTurk is an Internet survey platform that facilitates simple, inexpensive recruitment and payment of subjects. Individuals undertake “Human Intelligence Tasks” (HITs) on MTurk, which can include survey experiments. MTurk data compares favorably with typical experimental political science and psychology data collection methods, which are local convenience or student samples, in terms of internal and external validity. Research indicates MTurk samples replicate diverse experimental findings, even though their characteristics differ from the general population’s (Berinsky, Huber, and Lenz 2012; Buhrmester, Kwang, and Gosling 2011; Gosling et al. 2004) . Moreover, it is replication and triangulation, not random selection, which establish generalizability of experimental findings. Results replicate across survey experiments and across diverse data sources. Thus, it is reasonable to assume that the inferential statistics obtained from this series of experiments tend to be low in bias.

All six micro-level survey experimental designs incorporated random assignment to different race and background information variable values. Participants were randomly assigned to view a vignette consisting of some pertinent text alongside a photo that conveyed the race of the mock subject (i.e., the ostensible polygraph subject, patient, or food stamp applicant). The photos were normed along relevant dimensions: age,

familiarity, mood, memorability, and picture quality (Kennedy, Hope, and Raz 2009). Photos come from the Center for Vital Longevity Face Database. They feature neutral facial expressions and gray backgrounds (Minear and D. C. Park 2004).

All materials were pretested in Charlottesville, Virginia between January 2012 and September 2013, are available in the University Library (LIBRA) repository as a supplement, and will be available in full along with the associated data online upon publication. All recorded observations were utilized in the reported analyses, with the following pre-determined exclusions: (1) non-U.S. respondent location according to IP address, (2) repeated study completions from the same IP address, (3) non-compliance with quality control measures, established through automated checks, and (4) failure to enter a valid response ID (a code all respondents received at the end of the survey) when prompted.

Experimental Designs: Polygraph Case Study

In a series of four polygraph chart interpretation experiments (Experiments 1-4), the dependent variable is polygraph chart interpretation. In Experiments 1-3, the dependent variable values are whether participants judged a polygraph chart as indicating deception (“deception indicated,” valued at 1 for analysis), or not (“no deception indicated,” valued at 0). In Experiment 4, additional dependent variable operationalizations are added to this measure. Those operationalizations are whether participants would rather the polygraph be set to interpret charts in suspicious or friendly mode (to enable technology-mediated Bayesian updating), and whether participants assess the chart as correctly indicating whether someone is lying.

Like a typical polygraph chart, the chart (Figure 1 below and in appended survey instruments) consisted of a static graph representing changes in electrodermal responses (a measure relating to sweat on skin), cardiovascular responses (heart rate and a proxy measure of blood volume), and respiration rate and depth. Using complete polygraph charts from the field was undesirable due to their length and confidentiality issues. Instead, a sampling of open-source polygraph charts was compiled, and slightly modified to be relatively uniform in appearance other than controlled changes. Experienced federal polygraphers thought charts generated in this way, including the one used in this survey experiment, were real. Instructions to participants indicated how to read the charts and were based on federal polygraph chart interpretation methods (Department of Defense, Counterintelligence Field Activity 2006; Maschke and Scalabrini 2005; Psychophysiological Detection of Deception Program 2006; Sullivan 2010).¹¹ The chart was designed to be ambiguous, and results confirm its ambiguity.

¹¹ Instructions, also presented in supplemental form as part of the full survey instruments, read: Now, score the subject's polygraph chart using the following guidelines.

- Where the reaction appears greater following a relevant question (R) than following a control question (C), deception is indicated.
- Where the reaction appears greater following a control question (C) than following a relevant question (R), no deception is indicated.
- Look for the greater reaction for each parameter, where blue is respiration, green is galvanic skin response, and red is heart rate.

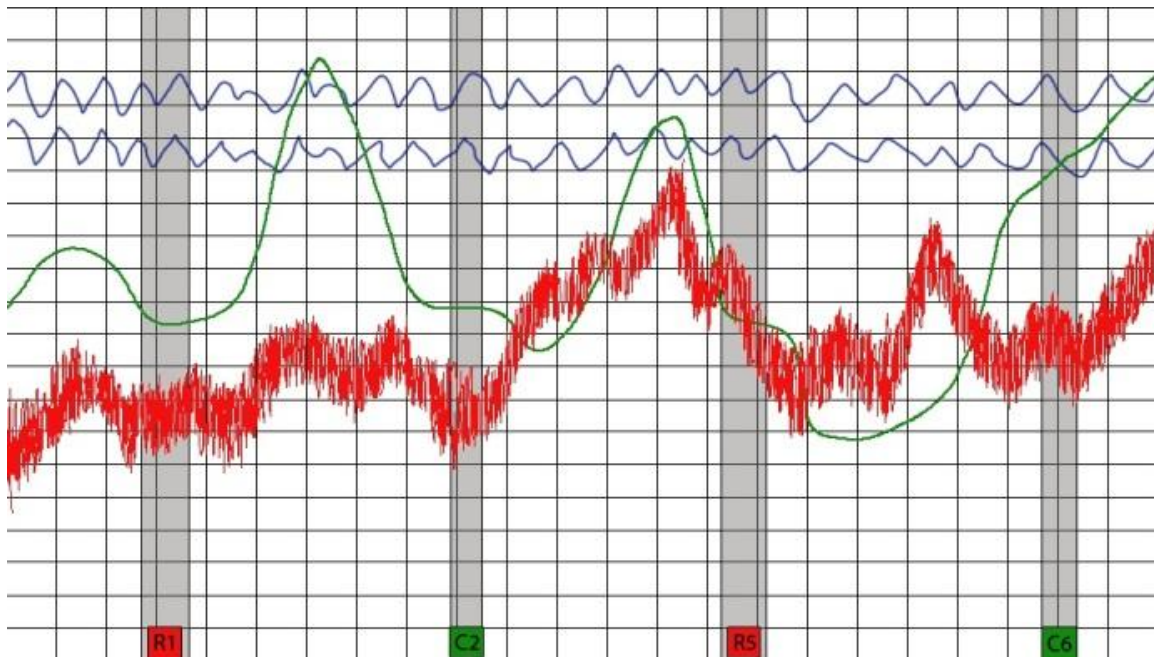


Figure 1: polygraph chart.

In Experiment 1, the treatment variables are polygraph subject race and background information. The experimental design is a 5x2 fully crossed matrix randomly varying race and background information. Race here has five values: dark-skinned black, light-skinned black, dark-skinned Hispanic, light-skinned Hispanic, and white. The black and Hispanic skin color variations were created digitally from one medium skin-toned black and Hispanic photo, respectively. This controlled for all features other than skin color within those pairs. Background information has two values: negative and neutral. These values are operationalized as varying information under the subheadings employment, medical, criminal, family, and credit history.

This experimental vignette text is presented to subjects as a background investigation they need to familiarize themselves with before scoring the polygraph chart, like a real polygrapher would. These treatment conditions all have good mundane

realism, because polygraphers see polygraph subjects’ race and familiarize themselves with background information, frequently in the form of background investigations structured like those used here, before conducting polygraphs.

Experiment 2, like Experiment 1, varies polygraph subject race and background information. The experimental design is a 2x3 fully crossed matrix randomly varying race and background information. It simplifies the race variable values to black and white. The photograph used for the black variable value is the original (non-morphed) photo used to generate the dark-skinned and light-skinned black photos in Experiment 1. The photograph used for the white variable value is the same photo used in Experiment 1. This set of simplified race variable values is then used again in Experiments 3 and 4. The background investigation variable values in Experiment 2 include a negative, neutral, and added positive background investigation treatment condition.

Experiments 3 and 4 both implement a 2x2x2 fully crossed matrix. Both retest race (black/white) and background information (negative/neutral) manipulations and their interactions. They also randomly vary novel contextual conditions. In Experiment 3, those conditions are the presence or absence of threat, emotion, and time pressure (“TEP”) as a combined set of conditions that might increase reliance on mental shortcuts in general and implicit bias in particular by encouraging automatic as opposed to controlled cognitive processing (Bargh 1994; Devine et al. 2002). The TEP control condition omits threat, emotion, and time pressure cues. In the TEP treatment condition, all three of those features are present, providing a tough test of the hypothesis that these contextual conditions systematically influence polygraph chart interpretation. Participants randomly assigned to the TEP condition view an introductory paragraph containing

information about violent crime rates. The information is framed as justification for why collecting evidence from suspects themselves is important. This vignette contains information needed to answer a quality control question in which participants must acknowledge violent crime can affect “Anyone, including me and my loved ones,” to proceed with the study. Participants are also asked to think of the last time they were the victim or witness of a violent crime such as a robbery or assault, or were afraid they were going to be, and rate their fear. This question invites participants to personalize the threat of violent crime presented in the vignette. Participants are also asked to think of the last time they heard about a violent crime, and rate their anger. Finally, time pressure in the TEP condition is operationalized through the addition of the following vignette immediately preceding the polygraph chart interpretation task: “Speed can be important in criminal investigations. So, to further simulate a real polygrapher's job in a criminal case, please work as fast as you can, without compromising accuracy. The two workers who correctly interpret the chart the fastest will each receive a bonus of \$5.” This treatment condition cues both intrinsic and extrinsic motivation for speed, generating time pressure.

In Experiment 4, the novel contextual condition randomly varied along with race and background information in the fully crossed 2x2x2 matrix is probability framing. All participants view a table applying Bayes’ rule to polygraph testing under Suspicious versus Friendly testing modes. Then they view a set of statements reframing the same likelihoods expressed as frequencies in the table as probabilities. In one probability framing treatment condition, participants view these probabilities in terms of accuracy rates. In the other, these probabilities are presented in terms of error rates.

Results: Polygraph Case Study

Negative background information systematically influenced polygraph chart interpretation across the first three of four experiments. In the fourth experiment, it significantly affected another interpretive scoring choice offered only in this experiment that is not typically available in polygraph chart interpretation, but might be in the future as technology catches up with statistics and Bayesian updating becomes increasingly automated. This overall collection of results suggests that confirmation bias, or the application of prior information to inform seemingly independent judgments, affects polygraphy.

Interestingly, this negative background information effect did not interact with the polygraph subject's race. Rather than undergoing a compounding interaction with confirmation bias, racial bias did not systematically affect polygraph chart interpretation either by itself or in combination with confirmation bias. Sample sizes are large enough that it is reasonable to expect a practically significant interaction effect between race and background information would have been detected with statistical significance if it existed. Nonsignificant interaction findings across cases including polygraphy, medical decision tools, and welfare benefits administration tools – as explored in greater detail in the following section – thus yield the important insight that race and background information probably do not undergo a compounding interaction in technology-mediated administrative decisions the way that leading theories about intersectional bias might lead one to expect. The credibility of this counter-intuitive finding is bolstered by its replication across six experiments (four in relation to the polygraph case study discussed

in this section, and two more in relation to secondary case studies of similar technologies that operate in different realms).

Table 1 summarizing Experiment 1 shows that the race and skin color of the polygraph subject do not affect the interpretation of the chart as indicating deception or not. The log odds of the chart being interpreted as indicating deception are significantly heightened when it is associated with a negative background investigation. The odds are about two times higher that a negative background investigation is associated with a guilty chart reading ($p = 0.013$).

Table 2 shows that Experiment 2 replicates and extends these results. In this smaller sample ($N = 241$ as compared to Experiment 1's $N = 1208$), the confirmation bias effect is only marginally significant ($p = 0.07$). The race of the polygraph subject still does not appear to systematically influence chart interpretation, even when combined with positive racial substereotypes in a new background information condition.

Table 3 presents results from Experiment 3, again replicating and extending these results. Again, race does not help explain variation in chart interpretation. This remains true even under conditions that we might expect to magnify racial and confirmation bias by increasing reliance on automatic cognitive processing – namely, threat, emotion, and time pressure. But confirmation bias, as operationalized by the negative background information variable value, does significantly help explain variation in chart interpretation ($p = 0.043$).

Table 4 shows that when probability frames are added, the results replicate and extend. Specifically, race and substereotype (race interacting with background information) does not help explain variation in chart interpretation. Under these framing

conditions emphasizing either accuracy or error rates, viewing negative background information causes interpreters to be more likely to select the suspicious instead of friendly testing mode given this option that was not offered in earlier experiments. So given a way to effectively delegate to technology what appears to be rational Bayesian updating of guilt in response to negative background information, interpreters systematically tend to take that option and then refrain from using that background information again to inform their chart readings. This result is highly statistically significant ($p = 0.012$). On one hand, it shows that confirmation bias in polygraph chart interpretation replicates and in a way extend to information environments in which probability-updating enables interpreters to make more fully neutral decisions. On the other hand, it shows that this same confirmation bias can be rechanneled in a rational way.

Table 1: Polygraph Chart Interpretation – Deceptive/non-deceptive, Experiment 1	
	Deception Indicated (SE)
Dark-skinned African-American	-0.118 (0.255)
Light-skinned African-American	-0.083 (0.258)
Dark-skinned Hispanic	-0.256 (0.268)
Light-skinned Hispanic	0.061 (0.255)
Negative background information	0.665* (0.266)
Dark-skinned African-American X negative background information	-0.582 (0.363)
Light-skinned African-American X negative background information	-0.147 (0.380)
Dark-skinned Hispanic X negative background information	-0.269 (0.378)
Light-skinned Hispanic X negative background information	0.024 (0.370)
Constant	-0.105 (0.188)
N = 1208	

* = $p < 0.05$, *** = $p < 0.000$. Results reflect coefficients from a logistic regression model, with standard errors in parentheses.

Table 2: Polygraph Chart Interpretation – Deceptive/non-deceptive, Experiment 2	
	Coeff. (SE)
African-American	-0.299 (0.436)
Negative background information	0.822[†] (0.453)
Positive background information	0.406 (0.433)
African-American X negative background information	-0.272 (0.654)
African-American X positive background information	0.191 (0.621)
Constant	-0.251 (0.291)
N = 241	

[†] signals $p = 0.07$. Results reflect coefficients from a logistic regression model, with standard errors in parentheses.

Table 3: Polygraph Chart Interpretation – Deceptive/non-deceptive, Experiment 3	
	Coeff. (SE)
African-American	-0.087 (0.371)
Negative background information	0.731* (0.361)
African-American X negative background information (substereotype)	-0.063 (0.524)
Threat, time pressure, emotion (TEP)	0.492 (0.350)
TEP X African-American	-0.455 (0.508)
TEP X negative background information	-0.328 (0.523)
TEP X substereotype	0.190 (0.745)
Constant	-0.405 (0.253)
N = 480	

* = $p < 0.05$, *** = $p < 0.000$. Results reflect coefficients from a logistic regression model, with standard errors in parentheses.

Table 4: Polygraph Chart Interpretation – Characteristics of judgments produced, Experiment 4			
	Deceptive chart interpretation	Suspicious (versus friendly) mode	Assessment of chart as correctly indicating deception
African-American	-0.357 (0.357)	0.008 (0.415)	-0.043 (0.474)
Negative background information	-0.028 (0.353)	1.843*** (0.391)	0.629 (0.429)
African-American X negative background information	0.927 (0.538)	-0.079 (0.587)	-0.284 (0.666)
Probability frame (error rate focus)	0.119 (0.349)	0.957* (0.379)	-0.065 (0.474)
Probability frame X African-American	-0.001 (0.510)	-0.615 (0.561)	0.386 (0.669)
Probability frame X negative background information	0.104 (0.526)	-0.578 (0.573)	-0.924 (0.711)
Probability frame X substereotype	-0.250 (0.748)	0.642 (0.813)	0.472 (0.972)
Constant	-0.154 (0.227)	-1.204*** (0.269)	-1.61*** (0.304)
N = 482			

* = $p < 0.05$, *** = $p < 0.000$. Results reflect coefficients from a logistic regression model, with standard errors in parentheses.

Experimental Designs: Medical Diagnosis and Welfare Benefits Administration Tool Case Studies

In a set of two additional survey experiments (Experiments 5 and 6), participants similarly see photos embedded in vignettes and then make technology-mediated administrative decisions. In an exploration of the generalizability of the polygraph chart interpretation results across diverse administrative realms, different technologies are employed in these secondary case studies. The first is a medical diagnosis decision tool and the second is a welfare benefits administration tool. In both experiments, the design

is a 2x2x2 fully crossed matrix randomly varying race, substereotype-relevant and domain-specific background information, and stereotype threat.

As in the polygraph chart interpretation experiments, the vignettes that operationalize the race and background information variables use information that relevant street-level bureaucrats (e.g., doctors and social workers) would really have under field conditions. Thus the treatment and control conditions in Experiments 5 and 6 have good mundane realism. The tools utilized in these secondary case study experiments are themselves used in field settings. The medical diagnosis tool used in this research is one of the top such tools in terms of its incorporation of evidence-based guidelines and content (W. F. Bond et al. 2010). This type of tool can improve diagnosis accuracy and efficiency by outperforming doctors (Graber, Tompkins, and Holland 2009; Thomas et al. 2008). It differs from more mainstream tools like WebMD or Mayo Clinic Symptom Checkers by producing probability-ranked differential diagnoses in response to user-entered symptoms.

The stereotype threat operationalization, which is added in these secondary case studies, is identical across Experiments 5 and 6. It also has good mundane realism, because regular people use the tools examined in these experiments (unlike polygraphs) by themselves under conditions that frequently cue group identities including gender and race. Stereotype threat is a type of normative threat in which cueing group membership in general or expected subpar group performance in particular can cause performance decrements and even physiological changes consistent with threat responses (Blascovich et al. 2001; Osborne 2006; Steele and Aronson 1995; Steele, Spencer, and Aronson 2002). Assessing whether stereotype threat generates bias in technology-mediated

administrative decisions is important, because stereotype threat might affect both street-level bureaucrats using technologies, and citizen-subjects themselves using increasingly popularized technologies like these. Indeed, the idea of reorganizing hierarchical decision-making to decrease bias and increase accuracy builds on the long-standing American pragmatist critique of expertise. Dewey famously suggested the shoe-wearer rather than the cobbler knows best how a shoe fits (or fails to fit), and so the democratic public is in some ways expert on the policies that affect it (Dewey 1954). Similarly, patients often have private information essential to their diagnosis, and welfare recipients have information needed to tailor benefits access to their needs. For this reason, technologies such as medical diagnosis tools and food stamp calculators are increasingly available to and used by citizen-subjects as well as the street-level bureaucrats who serve them. Moreover, research assessing possible bias in public administration must evaluate both bias toward subjects and the implications of being a target of bias, in order to more completely address the question of how bias affects these decisions under field conditions. Interpreter bias and stereotype threat are two sides of the same coin of potential bias in the dyadic interactions that street-level bureaucrats such as police, doctors, and social workers engage in with their subjects or clients. Stereotype threat might even itself generate interpreter bias in these dynamic interactions, because expectations can change behaviors in social interactions in self-fulfilling or cyclical ways (Snyder, Tanke, and Berscheid 1977).

In the stereotype threat treatment group, a set of four simple demographic questions about participants' own age, gender, ethnicity, and race comes before the technology-mediated decision task. The same questions come after that task in the control

group. The wording of these and other survey questions comes from extensively pretested survey instruments such as the U.S. Census. These basic demographic questions are likely to be asked in healthcare and welfare contexts under field conditions. This subtle stereotype threat prime, operationalized simply as the sequence of the demographic survey questions, was used because it has good mundane realism and was likely to cause the strongest performance deficits in women (Danaher and Crandall 2008). Thus it provided a tough test of whether stereotype threat affects accuracy in technology-mediated decisions.

In Experiment 5, the dependent variable is the diagnosis generated for a patient through use of a diagnosis decision tool. Its values are whether participants generated the correct diagnosis using the tool (valued at 0 for no, 1 for yes), whether the first diagnosis the tool listed given the symptoms participants entered for the patient was psychiatric in nature (valued at 0 for no, 1 for yes), and whether participants thought the correct diagnosis was psychiatric in nature (valued at 0 for no, 1 for yes). Operationalizing diagnostic accuracy through this array of objective and subjective, process and outcome-oriented measures strengthens the outcome variable’s construct validity.

Assessing whether the first diagnosis the tool listed was psychiatric in nature was a particularly tough test of the intersectional bias hypothesis, that a black patient on welfare/disability presenting with symptoms of a very difficult diagnosis (Huntingdon’s chorea) would be assessed with the same degree of accuracy as a white patient not on welfare/disability presenting with the same symptoms. Racial and intersectional health disparities are amply documented (National Research Council 2003; NRC 2010). This is a fiscal as well as a social justice crisis: the Institute of Medicine cites a \$76 billion

annual cost of racial disparities in U.S. healthcare (LaVeist, Gaskin, and Richard 2009; National Research Council 2003). Less agreement exists around what causes these disparities, with experimental and observational evidence suggesting contributing factors include differential clinician attributions, implicit bias, structural incentives, patient subcultural preferences, perceived discrimination, and stereotype threat (A. R. Green et al. 2007; Brownlee 2008; Ayanian et al. 1999; Byrne et al. 2011; Williams and Mohammed 2008; Trierweiler et al. 2000; Strakowski et al. 1996). Psychiatric misdiagnosis of African-Americans with low socio-economic status might be particularly common, because blacks’ mental health has been politicized since the antebellum era (Harris-Perry 2011). Rather than receding, this politicization arguably escalated in the mid-late 20th century. During the Civil Rights era, police and doctors came to view politically and socially active black men as paranoid and dangerous (Metzl 2009). The psychiatric discourse reframed schizophrenia, previously a disease of disorganized and disobedient housewives, in terms of racialized aggression. Today, schizophrenia is disproportionately diagnosed in poor racial minorities, especially migrant communities from developing countries, across the U.S., Britain, Western Europe, and Israel (Bresnahan et al. 2007; Selten and Cantor-Graae 2005; Littlewood and Lipsedge 1992).

More broadly, medical diagnosis is significant because missed diagnoses involving “a principal underlying disease or primary cause of death” occur at an estimated rate of 8-24% in contemporary U.S. hospitals (Shojania 2003). Adverse events and serious errors are also common and often result from misdiagnosis (Rothschild et al. 2005). Misdiagnosis is common, costly, and disproportionately affects disenfranchised groups.

The survey instrument in Experiment 5 has three parts: the medical puzzle, consisting of the patient scenario and online diagnosis tool; post-test survey questions on patient assessments and user experience with the tool; and pre- or post-test demographic questions. In the first part, as in the other experiments, the race variable values were black or white, operationalized through photographs embedded in an experimental vignette. In this experiment, the faces were morphed to be slightly more similar. The background information variable values are being on welfare and having applied for disability, or not. The text operationalizing this variable in the welfare/disability condition reads: “He had provided an emergency contact, and gave the doctor permission to call her. The emergency contact was his social worker. She didn't know anything about his medical history, except that he recently applied for disability. She had referred him to a service organization for assistance completing the application due to his poor organization.” In the non-welfare/disability condition, the same portion of the text instead read: “... The emergency contact was his neighbor. She didn't know anything about his medical history. She sometimes helped him with paperwork due to his poor organization.” The welfare/disability treatments systematically covaried to maximize the strength of tests for disconfirming evidence. This is realistic and internally consistent, because a poor person who is ill would have a greater need to apply for disability than a person who does not financially need to work. This covariation also generates a strong test of the hypothesis that racial and class biases compound, since there are particular substereotypes about lazy, undeserving blacks seeking public assistance (Gilens 2000; Mendelberg 2001). Being on welfare and applying for disability can both be conceived as public assistance-seeking behaviors.

This vignette appears in the experiment as a case description in the medical puzzle which participants solve using the diagnosis tool. The vignette is an excerpt of a published case report written for a lay audience, minimally edited for length, clarity, and the experimental manipulation detailed above (Sanders 2005). This origin further bolsters results’ external validity, because the case represented was a real one, and the description was written by a relevant expert.

Overall, these independent variable operationalizations have good mundane realism. Healthcare workers see patients’ race and often know what insurance they have, clueing them into patients’ welfare/disability status. When treating a very sick patient, they frequently call the emergency contact. Patients routinely answer the same demographic questions asked in this study.

Experiment 6, like Experiments 1-5, varies mock experimental client-subject race and background information. Like Experiment 5, its design is a 2x2x2 fully crossed matrix. This design randomly varies race, family status, and stereotype threat. As in Experiments 2-5, the race variable values are black and white. The gender of the client-subject here is female, keying into stereotypes and substereotypes about mothers who are on welfare, particularly those who are single, and particularly those who are both single and black. Participants read a social work client scenario and enter relevant information from it into an online Food Stamp Calculator.¹² Then they complete a survey about this

¹² The Oregon Department of Human Services Food Stamp benefit calculator is designed for public use, freely available online (<https://apps.state.or.us/fsestimate/>), and provides estimates of food stamp benefits based on household income and expenses. This calculator is typical of online calculators. For instance, most states do not consider household assets, so for Experiment 6, I used a calculator that does not consider them. The farm bill currently under consideration by Congress would change this particular aspect of most states’ food stamp administration. But a series of failed compromises resulted in no farm bill being passed by Congress in 2013, so whether this change becomes law remains an open question.

task, including their opinions about the calculator and the client, and answer some questions about their own characteristics (e.g., demographics).

Welfare benefits administration tools including food stamp calculators like this one are in extremely widespread use. According to the Center on Budget and Policy Priorities, 49 states have food stamp webpages that link to eligibility screening tools, benefit calculators, or link to the USDA’s benefit calculator (Center on Budget and Policy Priorities 2012). In related contexts, such as housing assistance, federalist administration patterns make private-sector administration of public assistance commonplace, increasing decentralization and diversity in administrative methods. But at the same time, the use of similar administrative technologies, such as HAPPY Software for housing benefit eligibility calculation, is increasingly widespread.

The family status variable values in this survey experiment are married or unwed. These values are operationalized as the children’s father either being the mother’s husband and serving overseas, or the children’s father living out of state and the parents never having been married. Specifically, the relevant part of the vignette read: “The social worker asked, ‘How many adults live in the home?’ ” In the married/overseas condition, the vignette continued: “ ‘My husband, James, is stationed overseas,’ the client said. ‘He sends money back, but not enough to pay rent and food.’ She touched her wedding band and looked distracted. ‘Maybe \$750 a month?’ ” In the unmarried/out of state condition, that portion of the scenario read instead: “ ‘How many adults live in the home?’ ‘Just me. James, their father, lives out of state. He pays child support, but not enough to pay rent and food,’ the client said. ‘Usually \$750 a month.’ She touched her face and looked distracted. ‘I always thought we would get married.’ ” The covariation of

the black client’s picture and the unmarried text constituted a strong test of the hypothesis that racial and family status biases compound because of substereotypes about single black welfare mothers.

Finally, as in Experiment 5, stereotype threat is operationalized in Experiment 6 by randomly placing four demographic questions before or after the diagnosis task. These questions again ask about age, gender, Hispanic origin, and race/ethnicity.

Overall, the independent variable operationalizations in Experiment 6 had good mundane realism. Social workers collect or have data on and usually see clients’ race, either in person or in the client’s file. They ask how many adults live in the household when evaluating benefits eligibility. Clients routinely answer the same demographic questions asked in this study.

The dependent variable was estimated SNAP benefit (food stamp) dollar amount. Additional data was collected and analyzed on judgments of the client and benefit tool user experience, and benefit assessments. Multiple measures of each of these outcomes were gathered. For client judgments, participants indicated how needy and deserving they thought the client was, whether they thought the client was honestly reporting his/her resources and needs, and whether they thought the client was accurately reporting his/her resources and needs. Regarding benefit tool user experience, participants rated the tool’s overall ease of use, the adequacy of the information it yielded, their confidence in the estimated benefit eligibility amount it listed, whether they would want to use the tool in their own case or with their clients if they have received public assistance or worked in social services, and how they would want to use it (e.g., by themselves or with a case worker).

Results: Medical Diagnosis and Welfare Benefits Administration Tool Case Studies

Results from medical diagnosis and welfare benefits administration tool case studies suggest that findings generalize from the polygraph chart interpretation case. Table 5 shows that patient race and welfare/disability status and their interaction do not significantly add to the explanation of variability in dependent variable values relating to whether or not the correct diagnosis was produced using the medical diagnosis tool. Since it involved a wide variety of diagnoses, the data from which these binary dependent variables were constructed was more complex than in the other case studies of administrative technologies. Consequently, it makes sense to triangulate a few different constructions of a binary outcome from these data. At most, results are suggestive about the cognitive availability of racial stereotypes, because an African-American patient on welfare/disability is significantly more likely to be given a first possible diagnosis in a list of possible diagnoses generated from the tool that is psychiatric ($p = 0.022$) – when the correct diagnosis in this case (Huntingdon’s chorea) is physical, and psychiatric misdiagnosis can do harm. But, this apparent availability effect does not seem to affect outcomes that would make this effect practically meaningful in terms of potentially denigrating quality of care (e.g., whether the correct diagnosis is obtained from the tool, and whether the correct diagnosis is deemed to be psychiatric rather than physical).

Table 6 shows that results again replicate in the distinct technology-mediated administrative context of a welfare benefits calculator. In an ordinary least squares linear model treating the family status independent variable as factor level, results lead to failure to reject the null hypothesis of no bias. That is, the race and family status independent variables do not significantly add to the explanation of variability in

dependent variable values. This is true in the case of the primary dependent variable – how much a subject qualifies for in food stamps according to the benefits calculation tool – and also for more subjective measures of secondary interest, such as whether people thought the client was honest or deserving. The intercept in this model is highly significant while intercepts in the models using data from polygraph chart interpretation experiments were not, because there is at least a modal correct answer (benefit amount) here. By contrast, the polygraph chart was designed to be ambiguous.

Table 5: Medical Diagnosis – Characteristics of Diagnoses, Experiment 5 (SE)			
	Correct Diagnosis	First Diagnosis Psychiatric	Correct Diagnosis Psychiatric
African-American	-0.132 (0.333)	-0.360 (0.183)	0.066 (0.167)
Welfare/disability status	-0.067 (0.324)	-0.254 (0.179)	0.197 (0.166)
African-American X welfare/disability status	-0.360 (0.489)	0.582* (0.253)	-0.269 (0.232)
Constant	-2.558*** (0.227)	-0.688 (0.124)	0.055 (0.117)
N = 1204			

* = $p < 0.05$, *** = $p < 0.000$. Results reflect coefficients from logistic regression models, with standard errors in parentheses.

Table 6: Welfare Benefit Calculation – SNAP Benefit (Food Stamp) Dollar Amount, Experiment 6 (SE)	
	Coeff. (SE)
African-American	22.610 (15.340)
Single parent status	-16.309 (15.003)
African-American X single parent status	-1.485 (21.214)
Constant	438.766 (11.007)
N = 1201	

Results reflect coefficients from an OLS regression model, with standard errors in parentheses.

Micro-Level Results: Significance

Overall, this research both builds on and challenges existing social and cognitive psychology, political science, and sociology research on intergroup prejudice, attribution bias, and intersectional bias. Leading theories of implicit and explicit racial attitudes predict that racial bias against blacks and particularly against negative stereotype-conforming blacks is pervasive, harmful to minorities, and affects decisions, especially under conditions such as ambiguity and the ability to attribute decisions to factors other than prejudice.¹³ By contrast, I compile novel evidence suggesting that at the micro level, racial and intersectional bias do not systematically characterize technology-mediated decisions. This result is robust across survey experiments in the cases of polygraph chart interpretation, medical diagnosis decision tool use, and welfare benefits calculation. In

¹³ Working from these theories and suggestive pilot study results, I hypothesized in my original National Science Foundation Doctoral Dissertation Research Grant proposal that racial and confirmation bias would be present in polygraph chart interpretation, and that their effects would be compounding so that racial subtypes would generate more bias than either racial stereotypes or other background information alone. So results from Experiments 1-6 repeatedly disprove some of my initial hypotheses. This emphasizes the tough nature of the statistical tests, the counter-intuitive nature of the results, and the rigorous nature of the research.

the polygraph case, null racial and intersectional bias results hold across a variety of tough test conditions. For instance, they hold when the background information cues racial stereotypes that are positive instead of negative. This is surprising, because recent research suggests that positive stereotypes can activate uncorrected racism where negative stereotypes do not (Kay et al. 2013). Similarly, it holds under conditions of threat, emotion, and time pressure, which is counter-intuitive because such conditions tend to privilege automatic over controlled forms of cognitive processing that involve more reliance on mental shortcuts including stereotypes. It also holds under probability-focused framing treatments emphasizing either polygraph accuracy or error rates.

These counter-intuitive null racial and intersectional bias results generalize to lab conditions according to unrelated lab and field datasets analyzed using different methods. This triangulation of diverse sources and methods yields significant strength to the claim that these survey experimental results generalize to the real world. They are reviewed here in brief, and more in-depth exploration is undertaken elsewhere. The field data in particular was analyzed in greater detail in Chapter 4.

In a lab psychophysiology experiment, race (black/white) and liberal political activism (presence/absence) were systematically covaried (black liberal activist/white non-liberal activist) to enable a test of intersectional (or racial substereotype) bias from the subject side. Most bias studies look at attitudes and behaviors from the interpreter side of dyadic interactions, rather than examining them from the subject side. This leaves room for subject-side biases, such as performance decrements from stereotype threat in testing contexts, or aggregate physiological differences in subgroup responses, to institutionalize bias undetected. In this experiment evaluating the possibility of subject-

side intersectional bias in polygraphy, eighty college-age males recruited at the University of Virginia in April-May 2012 were randomly assigned to a stereotype threat or non-stereotype threat treatment group. Participants in this study did not interpret a polygraph chart. Rather, both groups heard a standardized recording of polygraph test instructions and questions based on federal and police polygraph protocols. Participants’ physiological responses were continuously measured using skin conductance and wrist-to-wrist EKG, and their verbal responses were also recorded. Because of the sensitive nature of typical polygraph questions (e.g., about illegal behaviors), and to avoid possible audience or other contamination effects (e.g., from race and gender of an individual study administrator), participants underwent this study alone in a soundproof room. In the stereotype threat group, part of the recording detailed possible bias against blacks and liberals in polygraphy. No systematic differences were identified across groups. This null racial and intersectional bias finding is consistent with null findings in the polygraph chart interpretation survey experiments. It adds to those interpreter-focused experiments the novel insight that these biases do not appear to present from the subject side of the interaction.

Similarly, in a quasi-experimental analysis of nationally representative police department survey data (the Law Enforcement Management and Administrative Statistics or “LEMAS”), I find that pre-employment polygraph screening programs do not affect departments’ racial or gender diversity. This result was presented and discussed in greater detail in Chapter 4. In brief, departments were matched on factors that might otherwise affect diversity, and difference-in-differences regression is employed to evaluate possible change from time zero (before some departments instituted polygraph programs) to time

one (after some departments instituted polygraph programs). This technique arguably allows analysts to establish causal effects of policy interventions as if those interventions were randomized treatments. Such claims are clearly stronger when the results are triangulated and appear consistent with results from experimental data sources, as in this case. Overall, the consistency of null racial and intersectional bias results across distinct survey experimental, lab psychophysiology experimental, and field quasi-experimental datasets and contexts strongly suggests that this finding is correct. Racial bias does not systematically affect polygraphy.

Despite its apparent robustness and ecological validity, two wrinkles complicate the simple null answer to the question of whether administrative technologies can institutionalize bias. At the micro level, some forms of bias do influence such decisions. In the polygraph case, results across four survey experiments suggest confirmation bias influences chart interpretation. Interpreters are more likely to judge a polygraph chart as indicating deception when it is associated with a negative history. The replication of this result across four experiments suggests it generalizes across a variety of contextual conditions, such as the presence of threat, emotion, and time pressure, and the availability of Bayesian updating accuracy/error rates for the polygraph test.

However, the generalizability of this robust confirmation bias finding to field conditions remains an open empirical question. Existing research on this question is limited and its results mixed. In 1986, CBS’s “60 Minutes” staged a theft and randomly selected four polygraphers to test four employee suspects. Each polygrapher was told that a different employee was probably the thief, and each examiner found that employee deceptive (Saxe 1991). These hypothesis-blind polygraphers appeared to exhibit

confirmation bias under field conditions. The study design was thus strong, yet the small sample size prohibits meaningful tests of statistical significance. Another field study staged an ostensible replication of a polygraph validity study with seven police polygraphers (Elaad, Ginton, and Ben-Shakhar 2006). This study was conducted by researchers employed at the time of the study by the same police force from which it used polygraphers, generating a possible conflict of interest. Polygraphers each judged four of eighteen different students who were ostensibly suspected of cheating. Polygraphers were told two students were probably cheaters and two were probably not. There were not enough charts scored deceptive to tell with statistical significance whether systematic differences existed across interpretations. A larger-scale between-subjects field experiment using hypothesis-blind polygrapher subjects is needed to settle the question of whether confirmation bias generalizes to field polygraph conditions.

Possible confirmation bias in polygraphy is politically significant in the context of bureaucratic discretion. It illustrates one way in which street-level bureaucrats can use apparently neutral means to construct evidence that fits their discretionary preferences. Confirmatory evidence-seeking is a well-documented problem in political institutions’ decision-making processes, affecting a broad range of judgments from interpretations of forensic evidence, to decisions about whether or not to go to war (Blix 2004; Jervis 2006; Thompson 2009). This common cognitive strategy for making sense of the world sits on a spectrum that includes relatively insidious actions, such as cherry-picking intelligence to justify a priori preferred policies, as well as the more common, automatic heuristic use that we all engage in on an everyday basis.

Understanding that confirmation bias can affect apparently neutral, scientific processes such as technology-mediated decisions is integral to combatting it, thereby strengthening the rational-legal basis of public administration in liberal democratic institutions. This is integral to keeping micro-level decisions consistent with egalitarian norms, because confirmation bias might otherwise compromise the neutrality of such decisions. In this regard, technology-mediated administrative decisions exemplify a means of procedural fairness that does not necessarily translate into distributive fairness.

At the same time, the potential reality of micro-level race-neutral public administrative decisions raises questions about the macro-level attitudinal effects of this neutrality. Do race-neutral polygraph tests legitimate racially disparate death penalty sentencing outcomes as neutral and fair? More broadly, how does neutral administration function as an ideology? This is an important question, because significant macro-level outcome disparities stemming from public policies contrast with potential micro-level public administrative neutrality. If micro-level neutrality legitimates macro-level inequality, it might not be so neutral after all. The next chapter thus addresses the question: Does technology-mediated administrative neutrality affect attitudes in ways that might be said to weaken egalitarian norms?

Conclusion

This chapter presented novel empirical evidence that supports the validity of the ideal of neutrality at the level of technology-mediated administrative judgment and decision-making with respect to racial stereotypes and subtypes. Street-level bureaucrats are likely to make race-neutral administrative decisions using technologies such as

polygraphs, medical diagnosis tools, and welfare benefits administration tools – even with client-subjects whose intersectional status (e.g., a criminal record or single parenthood) cues negative racial substereotypes that we might expect to compound bias. However, my evidence also delineates new scope conditions of this surprising administrative neutrality with respect to confirmation bias.

Overall, my results suggest that political liberals are partly right: at the micro level, objectivity and fairness are not only ideal-type but truly possible in public administration. Technologies for decision-making in diverse public realms across security, medical, and welfare contexts offer street-level bureaucrats opportunities to ensure fairness in terms of constraining potential racial stereotype and substereotype-driven prejudice in some decision-making processes, although such fairness does not necessarily correlate with accuracy or prevent use of all mental shortcuts.

Those caveats are substantial. Polygraphs, for example, are about as accurate at detecting deception as coin flips (Zelicoff 2007). And results from four novel survey experiments presented here suggest confirmation bias can systematically influence polygraph chart interpretation, institutionalizing hunches and judgments based on prior information (accurate or not) as the outcomes of apparently neutral, scientific evaluation.

At the same time, the ways in which race-related heuristics fail to systematically influence technology-mediated administrative decisions in the experiments summarized here thus suggest important scope conditions of existing theories about implicit, subtle, and aversive forms of racism in particular. Prejudice is widely believed to be most likely to seep into judgments including behaviors and attitudes when a valid excuse is available, e.g., “I’m not racist – this [black] subject must be guilty because he has a criminal

background” (Pettigrew 1979; Allport 1979; Dovidio et al 2005). My micro-level experimental results suggest scope conditions across diverse public administrative realms of otherwise robust and troubling findings about the cognitive gravity of racial prejudice.

At the meta level, I suggest the value prioritization that in some ways constitutes the political realm informs the selection of ends that apparently value-neutral means such as administrative decision-making technologies serve. Informed by American pragmatists such as Randolph Bourne and scholars working under the rubric of the new political sociology of science, I caution that the appearance of neutrality these technologies project is thus itself value-laden (Frickel and Moore 2006; Bourne 1992). For example, the use of lie detectors to inform police and administrative investigations presupposes a particular conception of guilt and innocence as capable of physiological, and implicitly mathematical, as opposed to moral, proof – a particular form of certainty rooted in the Enlightenment fact-value distinction (Locke 2003). Similarly, the use of food stamp calculators to inform welfare benefits estimates assumes a means-tested rationality of public assistance. These conceptions are heavily value-laden in ways that the neutrality of the technologies that assume them tends to render less visible by framing administration as apolitical.

This disconnect echoes Kaufman’s caution about the fear of bureaucracy (Kaufman 1981). It is not evil bureaucrats – perpetuating latent prejudices through the invisible means of technology-mediated administration – who reproduce inequalities, contributing to vast, racialized disparities in criminal justice, health, and socio-economic well-being. Rather, it is broader social dynamics that are in some respects autonomous (in the sense of operating independently of individual actors and their micro-level decisions),

whether one privileges market, dialectical, or other forces in their articulation and explanation. That this is not a presently popular social scientific brand of explanation reflects in part the very humanizing impulse that, redirected, drives many professionals to work for the justice, health, and welfare of their fellow human beings in the first place: that is, it is potentially disempowering to acknowledge that macro-structural forces larger than individual actors, and perhaps larger than entire professions, institutions, and the public policies they work together to implement, can still drive crucially important outcomes in liberal democratic societies.

Does Procedural Fairness Make Inequality Just?
Attitudinal Effects of the Ideology of Administrative Neutrality

Introduction

This dissertation presents a sweeping analysis of administrative technology as an evolving way of understanding expertise and the Enlightenment fact-value distinction that has tremendous implications for politics from the construction of state power to the psychology of street-level bureaucrats. The preceding empirical chapters examined the development of the federal legal regime surrounding polygraph programs and the surveillance state, the institutional-level effects of screening tools including polygraph tests, and how some common cognitive biases might affect the behavioral decision-making of street-level bureaucrats making administrative decisions with the help of technologies including polygraphs. In this chapter, I present results from a survey experiment testing whether the apparent neutrality of administrative decision-making technologies influences attitudes about racial and socio-economic inequalities in particular, and fairness in general. My findings suggest that this neutrality does not affect these attitudes. Overall, then, “thinking technologies” can help bureaucrats make fair, unprejudiced decisions with respect to race and sometimes background information. These tools do not appear to cause an increase in inegalitarian attitudes by promoting a rationalization that procedural fairness justifies inequality.

Chapters 4 and 5 established the aggregate race-neutrality of polygraph programs and other technology-mediated administrative decision-making tools. But post-structuralist critics of the liberal ideal of neutral administration might have another charge about the limits of the neutrality of these tools. Administrative neutrality might function

as an ideology. The very appearance of objectivity and fairness might legitimate structural inequalities and thereby magnify inegalitarian racial and socio-economic attitudes or just world beliefs more broadly. Addressing whether or not administrative neutrality functions as an ideology with attitudinal effects along these lines is an important question, because significant macro-level outcome disparities stemming from public policies contrast with potential micro-level public administrative neutrality. If micro-level neutrality at the level of behavioral decision-making legitimates macro-level inequality at the attitudinal level, then administrative neutrality might not be so neutral after all. This chapter thus addresses the question: Does technology-mediated administrative neutrality affect attitudes in ways that might be said to weaken egalitarian norms?

Attitudinal Effects of the Ideology of Administrative Neutrality

At a general cognitive level, the broad attitudinal hypothesis this survey experiment tests is that people are often tempted to overgeneralize. This over-generalization might take many particular forms, of which I identify and test four. Broadly, if technology-mediated administration can be truly neutral, this legitimates the ideology that says the “race” of life is fair, and so we are a post-racial society now. This ideology of administrative neutrality collapses formal and effective equal opportunity. In this way, truly neutral administrative technologies might legitimate and reproduce macro-structural inequalities that are anything but value-neutral. It is well known that judgments about neutrality, trust, and social standing independently impact judgments of procedural justice (Tyler 1989).

But it is not known whether belief in technology-mediated administrative neutrality legitimates racial, economic, and other macro-structural inequalities.

Experimental Design

Using a posttest-only randomized experimental design, this survey experiment compares a variety of attitudes expressed by respondents with or without exposure to a narrative vignette about how technologies are increasingly available to help administrators make fewer errors and more fair decisions, with less room for prejudice to influence their decisions.¹⁴ Respondents in the treatment group then answer five questions about these types of technologies. One is a quality control question that tests whether respondents read the vignettes and therefore received the treatment. The other treatment group-only questions address how respondents feel about these technologies (very positive to very negative), whether they use any technologies like this in their own work or personal life, and whether they think use of these technologies in government will continue to grow over the next ten years. These questions both helped the quality control question hide so that it was more likely to catch people who did not read the vignette, and reinforced the manipulation by making this type of technology more cognitively accessible.

¹⁴ This treatment vignette read:

Some people worry that administrators can use their discretionary power to make unfair decisions. For example, police evaluating whether someone is telling the truth, doctors diagnosing patients, and social workers calculating benefit amounts might let their prejudices influence their decisions – even without their meaning to.

But technologies are increasingly available that help professionals ensure their decisions are neutral. For example, lie detection technologies are sometimes used to help police evaluate credibility. Software that lists diseases to check for in a patient with certain symptoms is sometimes used to help doctors diagnose disease correctly. And food stamp calculators are sometimes used to help social workers calculate benefit amounts. Technologies like this help to ensure that administrative errors are lower, and all people are treated fairly.

This survey experiment’s posttest-only design, as opposed to one using an alternate narrative in the control group, minimizes the risk of alternative plausible explanations for differences across treatment and control groups given that there is no neat inverse of this administrative neutrality narrative. As in the micro-level survey experiments summarized in Chapter 5, the full instrument is available in the University Library (LIBRA) repository as a supplement, and will be available in full along with the associated data online upon publication. Since egalitarianism and racial egalitarianism are related but distinct sets of norms, the first two outcome variables assess attitudinal effects on beliefs about fairness broadly, while the second two assess racial attitudinal effects in particular.

The first of the broader measures is a four-factor measure of belief in a just world assessing procedural and distributive justice beliefs for self and others (Lucas, Zhdanova, and Alexander 2011). This measure grows out of research suggesting people tend to blame victims or losers in order to maintain just world beliefs (beliefs that people get what they deserve), that in turn maintain a delusion of control, thus perpetuating social injustice through perceptual bias (Furnham 2003; Lerner and Simmons 1966; Lerner 1980; Rubin and Peplau 1973; Rubin and Peplau 1975). The second of these broader measures assesses economic system justification, or the tendency to accommodate and rationalize existing economic inequality rather than contesting it on the basis of interest or identity (Jost and Thompson 2000). Economic system justification differs from belief in just world in three main ways. First, it focuses on the economic realm. Second, it assumes people legitimate the status quo as fair out of motivated reasoning rather than strong internal locus of control, primarily in order to reduce negative emotions and

uncertainty (Jost and Hunyady 2003; Jost et al.). Finally, it can be strongest among those with lowest status (Jost, Banaji, and Nosek 2004).

In measuring racial attitudes, I incorporate indirect and direct measures to account for possible social desirability bias while relying on explicit attitudinal reports. The indirect measure of racial egalitarianism takes the form of a list experiment, and the direct measure of racial attitudes takes the form of the Racial Resentment panel questions from the American National Election Study (DeBell et al. 2011). The indirect measure modifies a list experiment (a.k.a. item count) method famously used in racial attitudinal measurement by Sniderman et al. to overcome social desirability effects (Kuklinski, Cobb, and Gilens 1997; Kuklinski et al. 1997; Blair and Imai 2012; Imai 2011; Sniderman, Tetlock, and Piazza 1992). It does this by randomizing assignment for participants to see either three or four attitude statements, and asking them to indicate only the number of those statements with which they agree.

For example, Kuklinski et al. wanted to establish whether racial attitudes in the South had really undergone a sea change in the 1990s (Kuklinski, Cobb, and Gilens 1997). Interviewers read a list of three items, asking survey respondents to tell them only *how many* upset them (not *which items*). The items were: federal government increasing the tax on gasoline; professional athletes getting million-dollar contracts; and large corporations polluting the environment. Respondents in the test condition heard these baseline items plus a fourth item: a black family moving in next door. By comparing the average number of items named in the control and treatment conditions (three and four, respectively), Kuklinski et al. determined that widespread, but difficult to measure, racism continued to mark the South.

Just as changing political cultural factors led Kuklinski et al. to measure racism in a novel way decades ago, so do changing political cultural factors lead me to update this list experimental technique to get at racial attitudinal effects in this survey experiment. Specifically, the racial attitudinal statement my list experiment uses is a “coded” one. Explicit racism has become so socially unacceptable, and priming effects appear so powerful, that many race researchers have turned to studying such coded political communications (Mendelberg 2001; Mendelberg 2008). Using a coded racial attitudinal statement provides a “tough test” of the null of no bias, because leading survey research shows such language can have significant effects in the post-Willie Horton era (Hurwitz and Peffley 2005). Here, treatment group participants see the racially coded list item “police profiling of inner city youth,” while control group participants do not see that item in their list. The other three list items, seen by respondents in the treatment and control groups, were the same ones used in the Kuklinski et al. example described above.

Because what we think of as practically significant effect sizes in attitudinal dependent variables tend to be small relative to practically significant effect sizes of behavioral dependent variables, the data collection decision rule in the macro-level survey experiment assessing attitudes was 2400, or roughly twice the size of the largest three micro-level survey experiments. Technically, this survey experimental design is a 2x2 matrix randomly varying both exposure to the technology-mediated neutrality and fairness treatment and exposure to the list experimental racially coded item. However, the beliefs in just world and fairness measures come before the racial attitudinal measures in the instrument in order to prevent racial priming effects in the former dependent variable measurements. Thus, the broader attitudinal measures are really obtained in the simpler

two-cell matrix, and the experimental racially coded item exposure is omitted from models of those dependent variables.

Macro-Level Results

Bonferroni-corrected p -values indicate no design effects. In other words, exposure to the technology-mediated administrative neutrality and fairness treatment, combined with exposure to the four-item list experimental treatment, did not systematically affect the examined attitudinal measures. Neither treatment systematically affected indirectly or directly measured racial attitudes. The technology-mediated neutrality and fairness treatment also did not affect beliefs about fairness in the form of just world beliefs or economic system justification. Sample size is large enough that it is reasonable to expect that treatment effects of practically significant magnitudes on these measures would have been detected with statistical significance if they existed. Nonsignificant regression coefficients across attitudinal measures suggest that technology-mediated administration does not increase endorsement of inequalitarian racial, economic, or fairness norms. These results contribute to the literatures on subtle and aversive racism by delineating novel scope conditions of these supposedly pervasive and subtle biases. Null attitudinal effects of technology-mediated neutrality at the attitudinal level, as shown in this survey experiment, are moreover consistent with null behavioral effects of the mantle of administrative neutrality at the micro level, as shown in Experiments 1-6 in Chapter 5.

All dependent variables in this experiment are rescaled to range 0-1, so that their regression coefficients are relatively comparable as presented. For economic system justification, just world beliefs, and racial resentment measures, this variable is a rescaled

average of responses across the relevant scale items. Higher values indicate more of the outcome variable being measured. For economic system justification and just world beliefs, higher values indicate higher levels of endorsing fairness beliefs. For racial resentment, higher values indicate stronger endorsement of racially inegalitarian or illiberal beliefs. Finally, for indirectly measured racial egalitarianism as assessed through the item count technique, lower values mean respondents indicated fewer listed items were upsetting. Higher values indicate respondents said more of the listed items were upsetting, as a proportion. Thus in this case, the higher the number, the more upsetting the racially coded item (police profiling of inner city youth) was to members of the treatment group. As with the racial resentment variable, then, the lower number indicates less racial egalitarianism or liberalism and the higher number indicates more inegalitarian attitudes.

Table 1 summarizes results for outcome variables relating to fairness broadly. It suggests that exposure to a treatment articulating the ideology of administrative neutrality does not systematically affect fairness beliefs. This null result holds when fairness beliefs are measured in terms of procedural justice, distributive justice, justice for others, justice for self, and justice in the economic system.

Tables 2 and 3 summarize results for the indirect and direct racial attitudinal outcome variables, respectively. The list experiment indicator (the racial attitude treatment independent variable) is included in these analyses while it was excluded from analyses in Table 1, because the list experiment was sequenced after the questions establishing fairness measures in the survey experiment. This sequencing guards against the possibility that the racial attitude treatment could have acted as a racial prime in these

broader fairness measures. Similarly, the racial attitude outcome variable models presented in Table 3 include an interaction not included in Table 1, to address the question: Does exposure to the administrative neutrality treatment change reported racial resentment among those in the four-item list group? The answer to this question is no.

Being randomly assigned to the technology-mediated neutrality treatment and/or the racially coded item count treatment groups did not systematically affect respondents' racial attitudes as measured directly through racial resentment or indirectly through a list experiment. This leads to failure to reject (but not acceptance of) the null hypothesis of no effect of the technology-mediated neutrality and fairness vignette on fairness beliefs and racial attitudes alike. In other words, when given a technological curtain behind which to hide, respondents do not tend to express stronger beliefs that the world is just and subaltern groups should pull themselves up by their bootstraps.

Table 1: Ideology of Administrative Neutrality – Effects on Fairness Beliefs (SE)					
	Economic System Justification	Procedural Just World Beliefs	Distributive Just World Beliefs	Self-oriented Just World Beliefs	Other-oriented Just World Beliefs
Technology-mediated neutrality and fairness treatment	-0.005 (0.028)	0.006 (0.007)	0.003 (0.008)	0.013 (0.008)	0.010 (0.008)
Constant	5.004*** (0.020)	0.489*** (0.005)	0.519*** (0.005)	0.611*** (0.006)	0.523*** (0.007)
N = 2412					

*** = $p < 0.000$. Results reflect coefficients from an OLS regression model, with standard errors in parentheses.

Table 2: Ideology of Administrative Neutrality – Effects on Indirectly Solicited Racial Attitudes (SE)	
	Indirect Racial Attitudes
Technology-mediated neutrality and fairness treatment – sensitive item group	-0.082 (0.286)
Constant	0.478*** (0.210)
Technology-mediated neutrality and fairness treatment – control item group	0.040 (0.071)
Constant	0.739*** (0.051)
N = 2412	

*** = $p < 0.000$. Results reflect expected value on a transformed logit scale on the dependent variable for the treatment and control groups, modeling the presentation of recent item count technique researchers to report results of the list experiment (Imai 2011).

Table 3: Ideology of Administrative Neutrality – Effects on Directly Solicited Racial Attitudes (SE)	
	Racial Resentment
Technology-mediated neutrality and fairness treatment	-0.005 (0.018)
Item count racial attitude treatment	-0.017 (0.013)
Technology-mediated neutrality and fairness treatment X item count racial attitude treatments	0.024 (0.018)
Constant	0.488*** (0.009)
N = 2412	

*** = $p < 0.000$. Results reflect coefficients from an OLS regression model, with standard errors in parentheses.

Macro-Level Results: Significance

Results suggest that the ideology of technology-mediated administrative neutrality and fairness does *not* rationalize inegalitarian attitudes at the level of directly reported attitudes surrounding a variety of beliefs about fairness, economic justice, and race. On one hand, this suggests that administrative neutrality is truly possible and does not threaten liberal democratic values of equality and fairness at the level of public opinion. Perhaps, then, post-structuralist critics of the recurrent Enlightenment, Marxian, Progressive, and Second Restoration-Era dream of advancing freedom and equality through improving government as a more perfectly objective, neutral “administration of things” are too gloomy with respect to what is possible in politics.

On the other hand, at a larger level politics remains fundamentally value-laden. For example, perfectly neutral administration of U.S. policies that systematically advantage the rich by tying school budgets to local real estate taxes will reproduce class advantages that disproportionately accrue to the racial majority, thus perpetuating

racialized socio-economic inequality on a grand scale. The more perfectly neutral administrative procedures are – and it is unequivocally cause for celebrating that decision-making technologies can help make discretionary decisions and more perfectly neutral in normatively meaningful ways – the more difficult it becomes to localize inegalitarian phenomena such as racism in individual decisions that explicitly violate egalitarian norms. Technology can keep the trains running on time; it cannot tell us where to go or why. So far from generating pure empirical building blocks of a post-racial society, this invisibility of individual actors in perpetuating macro-structural inequalities might exacerbate a more insidious form of racism in two ways. First, it might render invisible the inegalitarian nature of apparently neutral policies like tying school budgets to local real estate taxes. Second, it might limit public normative focus to questions of procedural fairness, stunting the breadth of ethical vision that might otherwise include questions about broader social processes linked to inequalities and their reproduction.

Conclusion

This chapter presented novel empirical evidence showing that, at the attitudinal level, technology-mediated administrative neutrality does not influence racial attitudes or beliefs about fairness more broadly. Political liberals are thus right that at the micro level of behavioral decision-making *and* the macro level of attitudinal effects of these decisions, objectivity and fairness are not only ideal-type but truly possible in public administration and public opinion in relation to those administrative processes. Technologies for decision-making in diverse public realms across security, medical, and

welfare contexts thus offer street-level bureaucrats opportunities to ensure greater fairness in administration without compromising egalitarian norms.

But as Chapters 2 and 3 established in mapping the development of the federal legal regime surrounding polygraphs and surveillance, at a larger level level, administrative decision-making technologies embody a reliance on neutral means that can make value-laden processes seem apolitical. This appearance is an ideological one that reflects the construction of state power and social control through information technology. And yet, technology-mediated administrative neutrality does not exert the sorts of direct behavioral or attitudinal effects one would expect to see if the ideology of administrative neutrality rationalized inegalitarian beliefs in an explicit way.

Technology, Neutrality, and the Future of Politics:
Bias-Free Does Not Mean Value-Neutral

Are technology-mediated decisions Trojan horses that legitimize bias or perfectly neutral applications of rational rules? This dissertation used diverse sources of data and methods of analysis to tackle different facets of this question in five empirical chapters. It generated and synthesized novel evidence to suggest that the answer lies somewhere in-between. These technologies can ensure that in the aggregate, racial and intersectional biases do not affect decisions that implicit or explicit forms of prejudice might otherwise shape, such as polygraph chart interpretation, medical diagnosis, and welfare benefits administration. And they contribute in this way to procedural fairness without shaping attitudes about unequal distributional outcomes including racial and socio-economic disparities. At the same time, technology-mediated administrative decisions can legitimize some forms of bias by giving preconceived notions a veneer of independent evidentiary support where none exists. In addition, they manifest a particular construction of state power in sometimes literal ways. In these respects, the limits of technology-mediated neutrality, combined with the rapid and pervasive spread of decision-making technologies across a variety of realms, give this project significant theoretical and practical import. I established these limits on three main levels of analysis.

First, I traced the political development of the surveillance state through the legal history of lie detection. Despite their insufficient scientific basis, including a lack of methodologically sound efficacy data, lie detection programs cost billions of taxpayer dollars a year and affect most Americans. And despite nearly one hundred years of judicial, legislative, and executive branch efforts to limit these programs, they continue to

grow. In a two-part qualitative historical exploration (Chapters 2 and 3), I argue that landmark efforts to protect individual rights and limit government power succeeded in important ways. But counter-intuitively, these same efforts also created a moral-political landscape that eroded the parallel fact-value and public-private boundaries gradually farther and farther until the external shock of 9/11 turned that erosion into a landslide. And the landmarks that mark the path of the developing federal legal regime surrounding lie detection also simultaneously delineate important conceptual shifts that illustrate and contribute to this erosion. Thus, they are also major turning points in the development of the contemporary surveillance state. In this way, the growth of lie detection has changed how we think about core moral-political constructs including expertise, coercion, and the limits of legitimate government.

Second, I assessed the institutional-level effects of a variety of more and less technology-mediated police screening tools including polygraphs. Using descriptive and quasi-experimental statistical analyses of nationally representative local and state law enforcement survey data, I showed that the use of these tools can have unintended effects on police departmental diversity and intended effects on citizen complaints of police use of force. These analyses also established that some of my survey experimental results generalize to field conditions. Specifically, the null experimental racial bias results for technology-mediated administrative decisions across the board is consistent with estimated null institutional-level effects of police polygraph programs on diversity. These same polygraph programs, which are thought by their proponents to decrease or deter misconduct, are associated with a decrease in sustained complaints (complaints found upon investigation to have been made with cause) about excessive officer user of force.

However, polygraph tests have no systematic effect on the total number of complaints. I argue that we would expect a true decrease in the measured entity (complaints) to correlate across measures (sustained and total complaints). The most likely explanation, then, for this combination of effects is that police recruits who successfully undergo polygraph screenings are better at lying or at not admitting wrongdoing under interrogation than recruits who do not pass these tests, thus causing fewer sustained but not fewer total complaints of police brutality against recruits who pass polygraph tests.

Finally, in a series of original survey experiments, I show that technology-mediated decisions can be systematically influenced by background information in the context of polygraph chart interpretation. In this way, confirmation bias can color apparently neutral decisions. However, I also show that racial bias does not systematically influence such decisions, either by itself or in interaction with background information cueing racial stereotypes. This counter-intuitive, null racial bias result replicates across various experimental conditions including threat, emotion, and time pressure, and across different technologies including polygraphy, a medical diagnosis tool, and a welfare benefits administration tool. Moreover, the idea that technologies increasingly aid public servants in making fair decisions does not have inegalitarian racial or other attitudinal effects. So technology does not appear to be a Trojan horse for ascriptive bias in behavioral decision-making or attitudinal terms. In this respect, technology-mediated administrative decisions tend to be more neutral than leading theories of racial stereotypes and substereotypes in particular, and automatic cognitive processes in general, might predict.

The significant theoretical merit of this research emerges from its contribution to interdisciplinary debates. On one hand, broad and deep racial outcome disparities persist across realms including criminal justice, health, and social welfare. On the other, egalitarian racial attitudes seem increasingly to be the norm and equal opportunity a meaningful if constrained reality. My dissertation shows how apparently neutral decision-making tools can sometimes directly or indirectly account for this disconnect. For instance, my quasi-experimental analysis in Chapter 4 suggests that some police screening tools that seem race-neutral directly hurt the representation of African-Americans on police forces, and I argue the most plausible causal mechanisms for these effects have to do with stereotype threat. My interpretive analysis in Chapters 2 and 3 suggests that applications of the fact-value distinction that seem neutral are actually deeply-value-laden.

My dissertation's related, interdisciplinary conceptual contributions include novel synthesis of apparently contradictory expectations about bias in public administration from social psychology literature on prejudice on one hand, and information technology research that shows that increasing its rule-bound nature improves decision-making on the other. This research advances knowledge and understanding of validity problems in forensics in particular across diverse literatures including political and social psychology, bias and equity in policing, and law and policy. More generally, this research responds in methodologically robust and theoretically novel ways to growing calls to generate evidence evaluating the effectiveness (and possible side effects) of public policy interventions that are in many cases already widespread. If the proposed reconceptualization of neutrality as referring not only to technological means but also the

ends they serve is correct, then this research will support a massive shift in the way we think about public administration, for instance, focusing on the political nature of conceptions of expertise, coercion, and privacy that administrative technologies often take for granted. This is an important political science project, because the increasingly central role of imperfect but avowedly objective science and technology in justifying the reach of these systems is a key component of the contemporary experience of governmental power. Technology is changing the nature of citizen-state interactions.

This research also has significant practical merit based on its potential for broader impacts. Data, survey instruments, codebooks, and statistical code from this research will be available online post-publication to the greatest extent given the need to balance transparency with research participant confidentiality. Survey instruments are appended to this dissertation in the University Library (LIBRA) repository as a supplement. I aim to convey my results through peer-reviewed publication, conference presentation, translation in popular media (e.g., in newspaper opinion essays), and personal correspondence, to contribute to administrative reforms that enhance the quality of care patients and social welfare clients receive, and to promote police reforms that might increase departmental diversity and decrease police brutality.

This research is socially and politically important, because technology-mediated administrative decisions have widespread, direct effects on millions of Americans. Lie detection programs alone cost billions of taxpayer dollars annually. Evaluating their possible validity problems, such as confirmation bias in polygraph chart interpretation, and quantifying their institutional-level effects, such as increasing police brutality, are especially important in light of current and developing security screening systems already

in use at many transportation hubs and other public places. A wide variety of diverse public sector polygraph programs affect the lives and life chances of a staggering array of federal public servants, citizens applying for those positions, local and state-level law enforcement officers and applicants, crime victims/witnesses, criminal suspects, convicted criminals, immigrants seeking official status, U.S. allies – and potentially anyone traveling through a busy airport or train station. Behavioral, verbal, and especially psychophysiological threat, credibility, or deception assessment systems that extend polygraph screening methods to transportation and border security screening contexts touch the realms of everyday freedom of movement, expression, association, and privacy for all. At a larger level, lie detection programs are thus one particular instantiation of the broader trend toward carceral and invasive institutions and policies. As such, lie detection exemplifies the new constitutional order of the surveillance state, and the associated new mode of interaction between civilians and state actors. It redefines the realm of the political.

Overall, this evidence suggests decision-making technologies can truly make administration more neutral and fair with respect to ascriptive biases, promoting some forms of procedural fairness – without simultaneously promoting the potentially inegalitarian logic that procedural fairness legitimizes distributional unfairness. At the same time, I establish some novel limits of this neutrality. At the micro level of individual decision-making, Chapter 5 shows a negative background investigation can systematically affect interpretation of an associated polygraph chart interpretation under a variety of conditions. At the macro level of institutional uses of tools including polygraphs to make administrative decisions such as hiring, Chapter 4 shows that these

tools – like other social interventions frequently employed on a large scale prior to rigorous testing of their efficacy or effects – can have perverse, unintended consequences. For example, tests that assess understanding of cultural diversity that are sometimes used to screen prospective recruits to police forces actually decrease African-American representation on police forces. And at the meta level of deeply political choices about not just technological means but the value-laden ends they serve, Chapters 2 and 3 suggest that the reliance on rule-bound processes embedded in technology-mediated decisions itself reflects a particular construction of state power and social control. Power dynamics shape decisions about what can be standardized and how to standardize it. These decisions thus constitute power. Far from making administration apolitical, technology-mediated administration can actually expand governmental power. While technology-mediated decisions are thus not Trojan horses legitimizing ascriptive bias or perfectly neutral applications of rational rules, they may yet be covert carriers of a new concept of the political in which an appearance of rule-bound administration replaces deeper democratic responsiveness or discourse about prioritizing values and defining the legitimate ends of government action.

To better assess that possibility, future research should address the question: How does everyday state surveillance affect people’s trust, well-being, and use of public services? Research on discrimination and stress suggests that chronic and severe exposure to unfairness and trauma physiologically impair the body’s ability to recover and function normally. Scholarship on urban violence similarly suggests that communities suffer health costs from gun violence. We know that injustice kills on an individual basis and that communities share that health burden, but we do not know

whether the physiological mechanism for the former carries over to group-level effects, especially when injustice is mediated by a particular public policy. I posit a new theory of procedural injustice stress that bridges these literatures and fills this gap by testing three interrelated hypotheses. Secondary data I propose to analyze in postdoctoral research will show whether high-intensity policing, a policing model involving intensive deployment of officers to high-crime areas, causes health problems across five U.S. cities in the form of public health service use levels. Primary survey and psychophysiological data will assess the effects of such policing on allostatic load (a normally healthy stress response that stops being elastic and starts doing cumulative damage under conditions of chronic stress), work time lost to illness, psychological well-being, and political efficacy. Finally, network data will provide evidence of how individuals at the center of the social support network in the affected communities may differentially carry the health burden of high-intensity policing (a crime prevention strategy that relies on heavy police surveillance of high-crime neighborhoods).

Bibliography

- Aamodt, Michael G. 2004. *Research in Law Enforcement Selection*. Boca Raton, FL: BrownWalker Press.
- Adorno, Theodor W., and Max Horkheimer. 2002. *Dialectic of Enlightenment: Philosophical Fragments*. Stanford: Stanford University Press.
- Aftergood, Steven. 2000. "Polygraph Testing and the DOE National Laboratories." *Science* 290 (5493) (November 3): 939–940.
- Alder, Ken. 2007. *The Lie Detectors: The History of an American Obsession*. New York: Free Press.
- Alexander, Michelle. 2012. *The New Jim Crow: Mass Incarceration in the Age of Colorblindness*. New York: New Press.
- Allport, Gordon W. 1979. *The Nature of Prejudice*. Unabridged, 25th anniversary ed. Reading, Mass: Addison-Wesley Pub. Co.
- Archambault, Joanne, and Kim Lonsway. 2008. "VAWA 2005 Restricts the Use of Polygraphs with Victims of Sexual Assault." Promising Practices: From the Desk of the Training Director. End Violence Against Women International. <http://www.ncdsv.org/images/VAWAPolygraphPromising%20Practices.pdf>.
- Archibold, Randal C. 2010. "U.S. Falters in Screening Border Patrol Near Mexico." *The New York Times*, March 11.
- Ayanian, John Z., Paul D. Cleary, Joel S. Weissman, and Arnold M. Epstein. 1999. "The Effect of Patients' Preferences on Racial Differences in Access to Renal Transplantation." *New England Journal of Medicine* 341 (22) (November 25): 1661–1669.
- Bach, Amy. 2009. *Ordinary Injustice: How America Holds Court*. 1st ed. New York: Metropolitan Books.
- Balkin, Jack M. 2008. "The Constitution in the National Surveillance State." *Minnesota Law Review* 93 (1): 1–25.
- Balkin, Jack M., and Sanford Levinson. 2006. "The Processes of Constitutional Change: From Partisan Entrenchment to the National Surveillance State." *Fordham Law Review* 75 (489).
- Bargh, John A. 1994. "The Four Horsemen of Automaticity: Awareness, Efficiency, Intention, and Control in Social Cognition." In *Handbook of Social Cognition*, Vol. 2, 1:1–31. Hillsdale, NJ: Erlbaum.
- Barland, Gordon H. 1995. "Foreign Use of the Polygraph – Summary v. 5.1" presented at the Annual Meeting of the American Polygraph Association, May 25, Las Vegas, NE.
- Baumann, Nick. 2013. "You'll Never Guess Where This FBI Agent Left a Secret Interrogation Manual." *Mother Jones*, December 20.
- Bechara, Antoine. 2004. "The Role of Emotion in Decision-Making: Evidence from Neurological Patients with Orbitofrontal Damage." *Brain and Cognition* 55 (1) (June): 30–40.
- Becker, Andrew. 2013. "During Polygraphs, Border Agency Applicants Admit to Rape, Kidnapping". *Center for Investigative Reporting*. Available online at <http://cironline.org/reports/during-polygraphs-border-agency-applicants-admit-rape-kidnapping-4325>.

- Becker, Judith V., and William D. Murphy. 1998. "What We Know and Do Not Know about Assessing and Treating Sex Offenders." *Psychology, Public Policy, and Law* 4 (1-2): 116–137.
- Bentham, Jeremy. 1838. *The Works of Jeremy Bentham*. Published under the Superintendence of his Executor, John Bowring. Vol. 7 of 11. Edinburgh: William Tait.
- Berghuis v. Thompkins, 130 S. Ct. 2250, 560 U.S. 370, 176 L. Ed. 2d 1098 (2010).
- Berinsky, A. J., G. A. Huber, and G. S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20 (3) (March 2): 351–368.
- Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94 (4) (September): 991–1013.
- Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-In-Differences Estimates?" *The Quarterly Journal of Economics* 119 (1) (February 1): 249–275.
- Bisson, JI, PL Jenkins, J Alexander, and C Bannister. 1997. "Randomised Controlled Trial of Psychological Debriefing for Victims of Acute Burn Trauma." *British Journal of Psychiatry* 171 (July): 78–81.
- Blair, Graeme, and Kosuke Imai. 2012. "Statistical Analysis of List Experiments." *Political Analysis* 20 (1) (January 17): 47–77.
- Blascovich, Jim, Steven J. Spencer, Diane Quinn, and Claude Steele. 2001. "African Americans and High Blood Pressure: The Role of Stereotype Threat." *Psychological Science* 12 (3) (May): 225–229.
- Bond, Charles F., and Bella M. DePaulo. 2008. "Individual Differences in Judging Deception: Accuracy and Bias." *Psychological Bulletin* 134 (4): 477–492.
- Bond, Charles F., A. Omar, U. Pitre, B.R. Lashley, L.M. Skaggs, and C.T. Kirk. 1992. "Fishy-Looking Liars: Deception Judgment from Expectancy Violation." *Journal of Personality and Social Psychology* 63 (6): 969–977.
- Bond, William F., Linda M. Schwartz, Kevin R. Weaver, Donald Levick, Michael Giuliano, and Mark L. Graber. 2010. "A Qualitative Review of Differential Diagnosis Generators". Conference poster presentation presented at the Diagnostic Errors in Medicine Conference, October, Toronto, Ontario, Canada. http://www.isabelhealthcare.com/pdf/Bond_A_Qualitative_Review_of_Differential_Diagnosis_Generators_DEM_2010.pdf.
- BORDERS National Center for Border Security and Immigration. 2013. "BORDERS' AVATAR on Duty in Bucharest Airport". Available online at <http://www.borders.arizona.edu/cms/news/borders-avatar-duty-bucharest-airport>.
- Bourdieu, Pierre. 1977. *Outline of a Theory of Practice*. trans. Richard Nice. Cambridge: Cambridge University Press.
- . 1990. *The Logic of Practice*. Stanford: Stanford University Press.
- Bourne, Randolph Silliman. 1992. *The Radical Will: Selected Writings, 1911-1918*. Berkeley: University of California Press.
- Brader, Ted. 2006. *Campaigning for Hearts and Minds: How Emotional Appeals in Political Ads Work*. Studies in Communication, Media, and Public Opinion. Chicago: University of Chicago Press.

- Branaman, Tim F., and Sheree N. Gallagher. 2005. "Polygraph Testing in Sex Offender Treatment: A Review of Limitations." *American Journal of Forensic Psychology* 23 (1): 45–64.
- Bresnahan, Michaeline, Melissa D Begg, Alan Brown, Catherine Schaefer, Nancy Sohler, Beverly Insel, Leah Vella, and Ezra Susser. 2007. "Race and Risk of Schizophrenia in a US Birth Cohort: Another Example of Health Disparity?" *International Journal of Epidemiology* 36 (4) (August): 751–758.
- Brewer v. Williams, 430 U.S. 387, 97 S. Ct. 1232, 51 L. Ed. 2d 424 (1977).
- Brownlee, Shannon. 2008. *Overtreated: Why Too Much Medicine Is Making Us Sicker and Poorer*. Pbk. ed. New York, NY: Bloomsbury.
- Buhrmester, M., T. Kwang, and S. D. Gosling. 2011. "Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?" *Perspectives on Psychological Science* 6 (1) (February 3): 3–5.
- Byrne, Margaret M., L. Jill Halman, Leonidas G. Koniaris, Peter A. Cassileth, Joseph D. Rosenblatt, and Michael C. Cheung. 2011. "Effects of Poverty and Race on Outcomes in Acute Myeloid Leukemia." *American Journal of Clinical Oncology* 34 (3) (June): 297–304.
- California v. Ciraolo, 476 U.S. 207, 106 S. Ct. 1809, 90 L. Ed. 2d 210 (1986).
- Callum, Robert. 2001. "The Case for Cultural Diversity in the Intelligence Community." *International Journal of Intelligence and CounterIntelligence* 14 (1): 25–48.
- Campbell, Donald T. 1998. "The Experimenting Society." In *The Experimenting Society: Essays in Honor of Donald T. Campbell*, edited by William N. Dunn. New Brunswick: Transaction Publishers.
- Carpenter, Daniel P. 2001. *The Forging of Bureaucratic Autonomy: Reputations, Networks, and Policy Innovation in Executive Agencies, 1862-1928*. Princeton, N.J: Princeton University Press.
- Cassell, Paul G., and Richard Fowles. 1998. "Handcuffing the Cops? A Thirty Year Perspective on Miranda's Effects on Law Enforcement." *Stanford Law Review* 50: 1055.
- Center on Budget and Policy Priorities. 2012. "SNAP Online: A Review of State Government SNAP Websites."
- Cohen, David B., and John Wilson Wells. 2004. *American National Security and Civil Liberties in an Era of Terrorism*. 1st ed. New York: Palgrave Macmillan.
- Cohen, G. L., J. Garcia, V. Purdie-Vaughns, N. Apfel, and P. Brzustoski. 2009. "Recursive Processes in Self-Affirmation: Intervening to Close the Minority Achievement Gap." *Science* 324 (5925): 400–403.
- Correll, Joshua, Bernadette Park, Charles M. Judd, Bernd Wittenbrink, Melody S. Sadler, and Tracie Keese. 2007. "Across the Thin Blue Line: Police Officers and Racial Bias in the Decision to Shoot." *Journal of Personality and Social Psychology* 92 (6): 1006–23.
- Cunningham, Scott. 2006. "The Florida Research." *The Police Chief: The Professional Voice of Law Enforcement* 73 (8).
- Damphousse, Kelly R., Laura Pointon, Deidra Upchurch, and Rebecca K. Moore. 2007. "Assessing the Validity of Voice Stress Analysis (VSA) Tools in a Jail Setting in Oklahoma City, Oklahoma". Norman, OK: University of Oklahoma.

- Danaher, Kelly, and Christian S. Crandall. 2008. "Stereotype Threat in Applied Settings Re-Examined." *Journal of Applied Social Psychology* 38 (6): 1639–55.
- Dao, James. 2005. "Lab's Errors Force Review of 150 DNA Cases." *The New York Times*, May 7.
- DeBell, Matthew, Catherine Wilson, Gary Segura, Simon Jackman, and Vincent Hutchings. 2011. "Methodology Report and User's Guide for the ANES 2010-2012 Evaluations of Government and Society Study". Palo Alto, CA, and Ann Arbor, MI: Stanford University and the University of Michigan.
- Department of Energy. 2006. "10 CFR Parts 709 and 710, Docket No. CN-03-RM-01, RIN 1992-AA33, Counterintelligence Evaluation Regulations." *Federal Register* 71 (189): 57386–97.
- Devine, Patricia G., E. Ashby Plant, David M. Amodio, Eddie Harmon-Jones, and Stephanie L. Vance. 2002. "The Regulation of Explicit and Implicit Race Bias: The Role of Motivations to Respond without Prejudice." *Journal of Personality and Social Psychology* 82 (5): 835–48.
- Dewey, John. 1954. *The Public & Its Problems*. 1st ed. Swallow Press.
- DiFonzo, J. Herbie. 2005. "The Crimes of Crime Labs." *Hofstra Law Review* 34 (1): 1–11.
- Dismore, David M. 2011. "Live-Blogging Women's History: March 27, 1922". *Ms. Magazine*. *Ms. Magazine Blog*.
<http://msmagazine.com/blog/blog/2011/03/27/live-blogging-womens-history-march-27-1922/>.
- Dovidio, John F., Peter Glick, and Laurie Rudman, eds. *On the Nature of Prejudice: Fifty Years after Allport*. 2005. Malden, MA: Blackwell Pub.
- Dow, Mark. 2004. *American Gulag: Inside U.S. Immigration Prisons*. Berkeley: University of California Press.
- Dworkin, Ronald. 2002. *Sovereign Virtue: The Theory and Practice of Equality*. Cambridge Mass.: Harvard University Press.
<http://www.amazon.com/Sovereign-Virtue-Theory-Practice-Equality/dp/0674008103>.
- Elaad, Eitan, Avital Ginton, and Gershon Ben-Shakhar. 2006. "The Role of Prior Expectations in Polygraph Examiners Decisions." *Psychology, Crime & Law* 7 (4): 1–16.
- Feldman, Stanley, and Karen Stenner. 1997. "Perceived Threat and Authoritarianism." *Political Psychology* 18 (4): 741–70.
- Foucault, Michel. 1980. *Power/Knowledge: Selected Interviews and Other Writings, 1972-1977*. ed. Colin Gordon, trans. Colin Gordon, Leo Marshall, John Mepham, Kate Soper. New York: Pantheon Books.
- . 1995. *Discipline and Punish : The Birth of the Prison*. 2nd Vintage Books ed. New York: Vintage Books.
- Freedman, David, and Persi Diaconis. 1981. "On the Histogram as a Density Estimator: L2 Theory." *Zeitschrift Für Wahrscheinlichkeitstheorie Und Verwandte Gebiete* 57 (4): 453–76.
- Frickel, Scott, and Kelly Moore. 2006. *The New Political Sociology of Science: Institutions, Networks, and Power. Science and Technology in Society*. Madison: University of Wisconsin Press.

- Furnas, Alexander. 2012. "Homeland Security's 'Pre-Crime' Screening Will Never Work." *The Atlantic*.
<http://www.theatlantic.com/technology/archive/2012/04/homeland-securitys-pre-crime-screening-will-never-work/255971/>.
- Furnham, Adrian. 2003. "Belief in a Just World: Research Progress over the Past Decade." *Personality and Individual Differences* 34 (5): 795–817.
- Gardner, Thomas. 2009. *Criminal Evidence: Principles and Cases*. 7th ed. Belmont CA: Wadsworth Cengage Learning.
- Geracimos, Ann. 2002. "A Special Kind of Education; FBI Trainees at Quantico Work Hard." *The Washington Times*, December 23. Available online at http://goliath.ecnext.com/coms2/gi_0199-999750/A-special-kind-of-education.html.
- Gerstein, Josh. 2013. "Next for Boston Marathon Suspect: 5 Legal Questions." *Politico*, April 19. <http://www.politico.com/story/2013/04/no-miranda-rights-for-now-for-bombing-suspect-90362.html>.
- Giannelli, Paul C. 2006. "Forensic Science." *The Journal of Law, Medicine & Ethics* 34 (2): 310–19.
- . 2011. "Daubert and Forensic Science: The Pitfalls of Law Enforcement Control of Scientific Research." *University of Illinois Law Review* 53 (1): 53–90.
- Gilens, Martin. 2000. *Why Americans Hate Welfare: Race, Media, and the Politics of Antipoverty Policy*. Studies in Communication, Media, and Public Opinion. Chicago: University Of Chicago Press.
- Gosling, Samuel D., Simine Vazire, Sanjay Srivastava, and Oliver P. John. 2004. "Should We Trust Web-Based Studies? A Comparative Analysis of Six Preconceptions About Internet Questionnaires." *American Psychologist* 59 (2): 93–104.
- Gottschalk, Marie. 2006. *The Prison and the Gallows: The Politics of Mass Incarceration in America*. Cambridge Studies in Criminology. Cambridge University Press.
- . 2008. "Hiding in Plain Sight: American Politics and the Carceral State." *Annual Review of Political Science* 11 (1): 235–60.
- Government Accountability Office. 1985. "DOD's Training Program for Polygraph Examiners". GAO/NSIAD-86-33BR. <http://gao.gov/assets/80/75023.pdf>.
- . 2013. "Border Security: U.S. Customs and Border Protection Provides Integrity-Related Training to Its Officers and Agents throughout Their Careers". GAO Integrity Training GAO-13-769R. <http://www.gao.gov/assets/660/657296.pdf>.
- Graber, Mark, David Tompkins, and Joanne J. Holland. 2009. "Resources Medical Students Use to Derive a Differential Diagnosis." *Medical Teacher: An International Journal of Education in the Health Sciences* 31 (6): 522–27.
- Green, Alexander R., Dana R. Carney, Daniel J. Pallin, Long H. Ngo, Kristal L. Raymond, Lisa I. Iezzoni, and Mahzarin R. Banaji. 2007. "Implicit Bias among Physicians and Its Prediction of Thrombolysis Decisions for Black and White Patients." *Journal of General Internal Medicine* 22 (9): 1231–38.
- Green, Amy S. 1996. "Girlz in Blue: Women Policing Violence in the NYPD." In *Police Violence: Understanding and Controlling Police Abuse of Force*, edited by William A. Geller and Hans Toch. New Haven, CT: Yale University Press.
- Greenemeier, Larry. 2012. "Avatar Officer Installed at Arizona-Mexico Border Station." *Scientific American*, August 6.

- Greenwald, Glenn. 2011. *With Liberty and Justice for Some: How the Law Is Used to Destroy Equality and Protect the Powerful*. New York: Picador.
- . 2013. “NSA Collecting Phone Records of Millions of Verizon Customers Daily.” *The Guardian*, June 5.
<http://www.theguardian.com/world/2013/jun/06/nsa-phone-records-verizon-court-order>.
- Guldi, E. Joanna. 2011. *The Road to Rule: The Expansion of the British Road Network, 1726-1848*. Cambridge, MA: Harvard University Press.
- Guston, David H. 2006. “On Consensus and Voting: From Asilomar to the National Toxicology Program.” In *The New Political Sociology of Science: Institutions, Networks, and Power*, 378–404. Science and Technology in Society. Madison, Wisconsin: University of Wisconsin Press.
- Hancock, Ange-Marie. 2007a. “When Multiplication Doesn’t Equal Quick Addition: Examining Intersectionality as a Research Paradigm.” *Perspectives on Politics* 5 (01): 63–79.
- . 2007b. “Intersectionality as a Normative and Empirical Paradigm.” *Politics & Gender* 3 (02): 248–54.
- Harris-Perry, Melissa V. 2011. *Sister Citizen: Shame, Stereotypes, and Black Women in America*. New Haven: Yale University Press.
- Healy, Gene, ed. 2004. *Go Directly to Jail: The Criminalization of Almost Everything*. Washington, D.C: Cato Institute.
- Hetherington, Marc J. 2009. *Authoritarianism and Polarization in American Politics*. New York: Cambridge University Press.
- Hibbing, John R. 1995. *Congress as Public Enemy: Public Attitudes toward American Political Institutions*. Cambridge Studies in Political Psychology and Public Opinion. Cambridge: Cambridge University Press.
- Hickman, Matthew J., and Alex R. Piquero. 2008. “Organizational, Administrative, and Environmental Correlates of Complaints About Police Use of Force: Does Minority Representation Matter?” *Crime & Delinquency* 55 (1): 3–27.
- Higginbotham, Adam. 2013. “Deception Is Futile When Big Brother’s Lie Detector Turns Its Eyes on You.” *Wired Threat Level*.
<http://www.wired.com/threatlevel/2013/01/ff-lie-detector/all/>.
- Hobbes, Thomas. 2009. *Leviathan: Or the Matter, Forme and Power of a Commonwealth Ecclesiasticall and Civil*. Translated by Noel Malcolm. Lexington, KY: Seven Treasures Publications.
- Hurwitz, Jon, and Mark Peffley. 1997. “Public Perceptions of Race and Crime: The Role of Racial Stereotypes.” *American Journal of Political Science* 41 (2): 375–401.
- . 2005. “Playing the Race Card in the Post-Willie Horton Era: The Impact of Racialized Code Words on Support for Punitive Crime Policy.” *Public Opinion Quarterly* 69 (1): 99–112.
- Iacono, William G. 2001. “Forensic ‘Lie Detection’: Procedures Without Scientific Basis.” *Journal of Forensic Psychology Practice* 1 (1): 75–86.
- Iacono, William G., and David T. Lykken. 1997. “The Validity of the Lie Detector: Two Surveys of Scientific Opinion.” *Journal of Applied Psychology* 82 (3): 426–33.
- Iacus, S. M., G. King, and G. Porro. 2011. “Causal Inference without Balance Checking: Coarsened Exact Matching.” *Political Analysis* 20 (1): 1–24.

- Imai, Kosuke. 2011. "Multivariate Regression Analysis for the Item Count Technique." *Journal of the American Statistical Association* 106 (494): 407–16.
- Institute of Medicine (U.S.). Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care. 2003. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, D.C: National Academies Press.
- Jasanoff, Sheila. 1995. *Science at the Bar: Law, Science, and Technology in America*. Twentieth Century Fund Books/Reports/Studies. Cambridge, Massachusetts: Harvard University Press.
- . 2003. "(No?) Accounting for Expertise." *Science and Public Policy* 30 (3): 157–62.
- Johnson, Kevin. 2006. "Police Agencies Find It Hard to Require Degrees." *USA Today*, September 18.
- Johnson, Lyndon B. 1965. "Equal Opportunity Is Not Enough". Commencement Address June 4, Howard University.
- Jost, John T., Mahzarin R. Banaji, and Brian A. Nosek. 2004. "A Decade of System Justification Theory: Accumulated Evidence of Conscious and Unconscious Bolstering of the Status Quo." *Political Psychology* 25 (6): 881–919.
- Jost, John T., and Orsolya Hunyady. 2003. "The Psychology of System Justification and the Palliative Function of Ideology." *European Review of Social Psychology* 13 (1): 111–53.
- Jost, John T., Ido Liviatan, Jojanneke van der Toorn, Alison Ledgerwood, Anesu Mandisodza, and Brian A. Nosek. "System Justification: How Do We Know It's Motivated?" In *The Psychology of Justice and Legitimacy: The Ontario Symposium*, edited by A.C. Kay et al. Vol. 11. Hillsdale, NJ: Erlbaum.
- Jost, John T., and Erik P. Thompson. 2000. "Group-Based Dominance and Opposition to Equality as Independent Predictors of Self-Esteem, Ethnocentrism, and Social Policy Attitudes among African Americans and European Americans." *Journal of Experimental Social Psychology* 36 (3): 209–32.
- Jussim, D. 1985. "The Privacy Invaders: Lies, Damn Lies - and Polygraphs." *The Nation*.
- Kalish, Brian. 2010. "Congress Passes Legislation Requiring Polygraphs for CBP Agents." *GovernmentExecutive.com*. December 22.
<http://www.govexec.com/mailbagDetails.cfm?aid=46774>.
- Kassin, Saul M, Christian A Meissner, and Rebecca J Norwick. 2005. "'I'd Know a False Confession If I Saw One': A Comparative Study of College Students and Police Investigators." *Law and Human Behavior* 29 (2): 211–27.
- Kassin, Saul M., Sara C. Appleby, and Jennifer Torkildson Perillo. 2010. "Interviewing Suspects: Practice, Science, and Future Directions." *Legal and Criminological Psychology* 15 (1): 39–55.
- Kassin, Saul M., and Christina T. Fong. 1999. "'I'm Innocent!': Effects of Training on Judgments of Truth and Deception in the Interrogation Room." *Law and Human Behavior* 23 (5): 499–516.
- Kaufman, Herbert. 1956. "Emerging Conflicts in the Doctrines of Public Administration." *American Political Science Review* 50 (4): 1057–73.
- . 1969. "Administrative Decentralization and Political Power." *Public Administration Review* 29 (1): 3–15.

- . 1981. "Fear of Bureaucracy: A Raging Pandemic." *Public Administration Review* 41 (1): 1–9.
- . 2006. *The Forest Ranger: A Study in Administrative Behavior*. Special reprint ed. Washington, DC: Resources for the Future.
- Kay, Aaron C., Martin V. Day, Mark P. Zanna, and A. David Nussbaum. 2013. "The Insidious (and Ironic) Effects of Positive Stereotypes." *Journal of Experimental Social Psychology* 49 (2): 287–91.
- Klosko, George, and Steven Wall. 2003. *Perfectionism and Neutrality: Essays in Liberal Theory*. Lanham, Md.: Rowman & Littlefield Publishers.
- Krane, D. E., V. Bahn, D. Balding, B. Barlow, H. Cash, B. L. Desportes, P. D'Eustachio, et al. 2009. "Time for DNA Disclosure." *Science* 326 (5960): 1631–32.
- Kuklinski, James H., Michael D. Cobb, and Martin Gilens. 1997. "Racial Attitudes and the 'New South.'" *The Journal of Politics* 59 (02): 323.
- Kuklinski, James H., Paul M. Sniderman, Kathleen Knight, Thomas Piazza, Philip E. Tetlock, Gordon R. Lawrence, and Barbara Mellers. 1997. "Racial Prejudice and Attitudes Toward Affirmative Action." *American Journal of Political Science* 41 (2): 402–19.
- Lamb, Marguerite. 2001. "Who Was Wonder Woman 1?" *Bostonia*.
<http://www.bu.edu/bostonia/fall01/woman/>.
- LaVeist, Thomas A., Darrel J. Gaskin, and Patrick Richard. 2009. "The Economic Burden of Health Inequalities in the United States". Joint Center for Political and Economic Studies.
- Lerman, Amy E., and Vesla Weaver. 2012. "Staying Out of Sight: Concentrated Policing and Local Political Action." In *Yale University*.
- Lerman, Amy E., and Vesla M. Weaver. 2013. "Staying out of Sight? Concentrated Policing and Local Political Action." *The ANNALS of the American Academy of Political and Social Science* 651 (1): 202–19.
- . 2014. *Policing Citizenship: America's Antidemocratic Institutions and the Creation of the New Civic Underclass*. University of Chicago Press.
- Lerner, Melvin J. 1980. *The Belief in a Just World: A Fundamental Delusion*. Perspectives in Social Psychology. New York: Plenum Press.
- Lerner, Melvin J., and Dale T. Miller. 1978. "Just World Research and the Attribution Process: Looking Back and Ahead." *Psychological Bulletin* 85 (5): 1030–51.
- Lerner, Melvin J., and Carolyn H. Simmons. 1966. "Observer's Reaction to the 'Innocent Victim': Compassion or Rejection?" *Journal of Personality and Social Psychology* 4 (2): 203–10.
- Levine, Timothy R., Lori N. Anders, John Banas, Karie Leigh Baum, Keriane Endo, Allison D. S. Hu, and Norman C. H. Wong. 2000. "Norms, Expectations, and Deception: A Norm Violation Model of Veracity Judgments." *Communication Monographs* 67 (2): 123–37.
- Levine, Timothy R., Rachel K. Kim, and J. Pete Blair. 2010. "(In)accuracy at Detecting True and False Confessions and Denials: An Initial Test of a Projected Motive Model of Veracity Judgments." *Human Communication Research* 36 (1): 82–102.

- Lichtblau, Eric. 2013. "In Secret, Court Vastly Broadens Powers of N.S.A." *The New York Times*, July 6. <http://www.nytimes.com/2013/07/07/us/in-secret-court-vastly-broadens-powers-of-nsa.html>.
- Lilienfeld, Scott O. 2007. "Psychological Treatments That Cause Harm." *Perspectives on Psychological Science* 2 (1): 53–70.
- Lipsky, Michael. 1980. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*. New York: Russell Sage Foundation.
- Littlewood, Roland, and Maurice Lipsedge. 1992. "Schizophrenia among Afro-Caribbeans." *The British Journal of Psychiatry* 160: 710–11.
- Locke, John. 1980. *Second Treatise of Government*. 1st ed. Indianapolis, Ind: Hackett Pub. Co.
- Locke, John. Ed. David Wootton. 2003. *The Political Writings of John Locke*. Indianapolis: Hackett Pub.
- Lowi, Theodore J. 2010. *The End of Liberalism: The Second Republic of the United States*. New York: W.W. Norton & Co.
- Lucas, Todd, Ludmila Zhdanova, and Sheldon Alexander. 2011. "Procedural and Distributive Justice Beliefs for Self and Others: Assessment of a Four-Factor Individual Differences Model." *Journal of Individual Differences* 32 (1): 14–25.
- Lykken, David. 1998. *A Tremor in the Blood: Uses and Abuses of the Lie Detector*. 2nd ed. New York: Plenum Trade.
- Lynch, Michael, Simon A. Cole, Ruth McNally, and Kathleen Jordan. 2008. *Truth Machine: The Contentious History of DNA Fingerprinting*. Chicago: University of Chicago Press.
- Mann, Michael. 1993. *The Sources of Social Power*. Reprint. Vol. 2. Cambridge: Cambridge University Press.
- . 1997. *The Sources of Social Power*. Reprint. Vol. 1. Cambridge: Cambridge University Press.
- March, James G., and Johan P. Olsen. 1989. *Rediscovering Institutions: The Organizational Basis of Politics*. New York: Free Press.
- Mayer, David C. 2000. "Reviewing National Security Clearance Decisions: The Clash Between Title VII and Bivens Claims." *Cornell Law Review* 85 (3): 786–821.
- McNally, Richard J., Richard A. Bryant, and Anke Ehlers. 2003. "Does Early Psychological Intervention Promote Recovery from Posttraumatic Stress?" *Psychological Science in the Public Interest* 4 (2): 45–79.
- Meijer, Ewout H., Bruno Verschuere, Harald L.G.J. Merckelbach, and Geert Crombez. 2008. "Sex Offender Management Using the Polygraph: A Critical Review." *International Journal of Law and Psychiatry* 31 (5): 423–29.
- Meissner, Christian A., and Saul M. Kassin. 2002. "'He's Guilty!': Investigator Bias in Judgments of Truth and Deception." *Law and Human Behavior* 26 (5): 423–80.
- Mendelberg, Tali. 2001. *The Race Card: Campaign Strategy, Implicit Messages, and the Norm of Equality*. Princeton N.J.: Princeton University Press.
- . 2008. "Racial Priming Revived." *Perspectives on Politics* 6 (1): 109–23.
- Metzl, Jonathan. 2009. *The Protest Psychosis: How Schizophrenia Became a Black Disease*. Boston: Beacon Press.
- Miller, Michael K. 2013. "The Uses and Abuses of Matching". George Washington University.

- National Research Council. 2003. *The Polygraph and Lie Detection*. Washington, DC: The National Academies Press.
- National Research Council. 2002. *Measuring Housing Discrimination in a National Study: Report of a Workshop*. Washington, DC: The National Academies Press.
- . 2010. *The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary*. Washington, DC: The National Academies Press.
- National Security Counselors. "Document Vault." *National Security Law*. http://nationalsecuritylaw.org/document_vault.html#storylink=relast.
- Nguyen, Hannah-Hanh D., and Ann Marie Ryan. 2008. "Does Stereotype Threat Affect Test Performance of Minorities and Women? A Meta-Analysis of Experimental Evidence." *Journal of Applied Psychology* 93 (6): 1314–34.
- Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.
- Nunamaker, Jay, and Elyse Golob. 2012. "National Center for Border Security and Immigration Work Plan - Year 5: July 2012 to June 2013." Available online at <http://www.borders.arizona.edu/cms/sites/default/files/BORDERS-YR5-Workplan-APPROVED%2008022012.pdf>.
- O'Brien, David M. 2008. *Constitutional Law and Politics*. New York: W.W. Norton & Co.
- . 1978. "The Fifth Amendment: Fox Hunters, Old Women, Hermits, and the Burger Court." *Notre Dame Lawyer* 54 (26): 26–72.
- O'Donnell, Sean Tath. 2007. "Courting Science, Binding Truth: A Social History of *Frye v. United States*". Dissertation, Harvard University. <http://gradworks.umi.com/32/85/3285520.html>.
- O'Reilly, Andrew. 2012. "Avatar Border Agent Screens Commuters At Arizona Post." *Fox News Latino*. <http://latino.foxnews.com/latino/news/2012/08/06/avatar-border-agent-screens-commuters-at-arizona-post/>.
- Office of Technology Assessment. 1983. "Scientific Validity of Polygraph Testing: A Research Review and Evaluation". A Technical Memorandum OTA-TM-H-15. Washington, D.C.: U.S. Congress.
- Osborne, Jason W. 2006. "Gender, Stereotype Threat, and Anxiety: Psychophysiological and Cognitive Evidence." *Electronic Journal of Research in Educational Psychology* 4 (8): 109–38.
- Pager, Devah. 2007. *Marked : Race, Crime, and Finding Work in an Era of Mass Incarceration*. Chicago: University of Chicago Press.
- Patriot Acts: Narratives of Post-9/11 Injustice. 2011. *Voice of Witness*. San Francisco: McSweeney's Books.
- Peffley, Mark, Jon Hurwitz, and Paul M. Sniderman. 1997. "Racial Stereotypes and Whites' Political Views of Blacks in the Context of Welfare and Crime." *American Journal of Political Science* 41 (1): 30–60.
- Pettigrew, Thomas F. 1979. "The Ultimate Attribution Error: Extending Allport's Cognitive Analysis of Prejudice." *Personality and Social Psychology Bulletin* 5 (4): 461–76.
- Porter, Theodore. 1996. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. 1st. pbk. ed. Princeton N.J.: Princeton University Press.

- Priest, Dana. 2007. "Detention and Interrogation." In *Crimes of War 2.0: What the Public Should Know*, edited by Anthony Dworkin, Roy Gutman, and David Rieff, Revised and Expanded. New York: W.W. Norton & Co.
- Priest, Dana, and William M. Arkin. 2010. "A Hidden World, Growing beyond Control." *Top Secret America: A Washington Post Investigation*. July 19. <http://projects.washingtonpost.com/top-secret-america/articles/a-hidden-world-growing-beyond-control/>.
- . 2011. *Top Secret America: The Rise of the New American Security State*. 1st ed. New York: Little, Brown and Co.
- Rabinowitz, Joshua L., David O. Sears, Jim Sidanius, and Jon A. Krosnick. 2009. "Why Do White Americans Oppose Race-Targeted Policies? Clarifying the Impact of Symbolic Racism." *Political Psychology* 30 (5): 805–28.
- Rampersad, Arnold. 1986. *The Life of Langston Hughes*. Vol. I: 1902–1941: I, Too, Sing America. New York: Oxford University Press.
- Rawls, John. 2005. *Political Liberalism*. Expanded ed. Columbia Classics in Philosophy. New York: Columbia University Press.
- Reaves, Brian A. 2012. "Hiring and Retention of State and Local Law Enforcement Officers, 2008 - Statistical Tables". NCJ 238251. Bureau of Justice Statistics - Statistical Tables. U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics.
- Rejali, Darius M. 2007. *Torture and Democracy*. Princeton: Princeton University Press.
- Rice, Paul R. 2009. *Evidence: Common Law and Federal Rules of Evidence*. 6th ed. Newark, NJ: LexisNexis.
- Richard, Orlando C. 2000. "Racial Diversity, Business Strategy, and Firm Performance: A Resource-Based View." *The Academy of Management Journal* 43 (2): 164–77.
- Roberts, Paul Craig, and Lawrence M. Stratton. 2008. *The Tyranny of Good Intentions: How Prosecutors and Law Enforcement Are Trampling the Constitution in the Name of Justice*. New York: Three Rivers Press.
- Rocke, D. J., S. Thomas, L. Puscas, and W. T. Lee. 2013. "Physician Knowledge of and Attitudes toward the Patient Protection and Affordable Care Act." *Otolaryngology -- Head and Neck Surgery*.
- Rosenbaum, D.P., and G.S. Hanson. 1998. "Assessing the Effects of School-Based Drug Education: A Six-Year Multi-Level Analysis of Project D.A.R.E." *Journal of Research in Crime and Delinquency* 35 (4): 381–412.
- Ross, Dorothy. 1992. *The Origins of American Social Science. Ideas in Context*. Cambridge: Cambridge University Press.
- Rothschild, Jeffrey M, Christopher P Landrigan, John W Cronin, Rainu Kaushal, Steven W Lockley, Elisabeth Burdick, Peter H Stone, et al. 2005. "The Critical Care Safety Study: The Incidence and Nature of Adverse Events and Serious Medical Errors in Intensive Care." *Critical Care Medicine* 33 (8): 1694–1700.
- Rourke, Francis E. 1987. "Bureaucracy in the American Constitutional Order." *Political Science Quarterly* 102 (2): 217–32.
- Rozelle, Susan D. 2007. "Daubert, Schaubert: Criminal Defendants and the Short End of the Science Stick." *Tulsa Law Review* 43: 597–607.

- Rubin, Zick, and Anne Peplau. 1973. "Belief in a Just World and Reactions to Another's Lot: A Study of Participants in the National Draft Lottery." *Journal of Social Issues* 29 (4): 73–93.
- Rubin, Zick, and Letitia Anne Peplau. 1975. "Who Believes in a Just World?" *Journal of Social Issues* 31 (3): 65–89.
- Saxe, Leonard. 1991. "Lying: Thoughts of an Applied Social Psychologist." *American Psychologist* 46 (4): 409–15.
- Saxe, Leonard, and Gershon Ben-Shakhar. 1999. "Admissibility of Polygraph Tests: The Application of Scientific Standards Post- Daubert." *Psychology, Public Policy, and Law* 5 (1): 203–23.
- Schmidt, Frank L., and John E. Hunter. 1998. "The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings." *Psychological Bulletin* 124 (2): 262–74.
- Scott, James C. 1998. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven: Yale University Press.
- Sears, David O., James Sidanius, and Lawrence Bobo. 2000. *Racialized Politics: The Debate about Racism in America*. Chicago: University of Chicago Press.
- Selten, Jean-Paul, and Elizabeth Cantor-Graae. 2005. "Social Defeat: Risk Factor for Schizophrenia?" *The British Journal of Psychiatry* 187: 101–2.
- Sening, Yvonne Koontz. 1989. "Heads or Tails: The Employee Polygraph Protection Act." *Catholic University Law Review* 39 (235).
- Seymour, Ben, and Ray Dolan. 2008. "Emotion, Decision Making, and the Amygdala." *Neuron* 58 (5): 662–71.
- Shojania, K. G. 2003. "Changes in Rates of Autopsy-Detected Diagnostic Errors Over Time: A Systematic Review." *JAMA: The Journal of the American Medical Association* 289 (21): 2849–56.
- Sidanius, Jim, and Felicia Pratto. 2001. *Social Dominance an Intergroup Theory of Social Hierarchy and Oppression*. Cambridge: Cambridge University Press.
- Siems, Larry. 2011. *Torture Report: What the Documents Say about America's Post-9/11 Torture Program*. New York: OR Books.
- Sijbrandij, Marit, Miranda Olff, Johannes B. Reitsma, Ingrid V.E. Carlier, and Berthold P.R. Gersons. 2006. "Emotional or Educational Debriefing after Psychological Trauma: Randomised Controlled Trial." *British Journal of Psychiatry* 189 (2): 150–55.
- Silvergate, Harvey A. 2009. *Three Felonies a Day: How the Feds Target the Innocent*. New York: Encounter Books.
- Sklansky, David Alan. 2006. "Not Your Father's Police Department: Making Sense of the New Demographics of Law Enforcement." *Journal of Criminal Law and Criminology* 96.
- Skowronek, Stephen. 1982. *Building a New American State: The Expansion of National Administrative Capacities, 1877-1920*. Cambridge: Cambridge University Press.
- Sloan, L.M. 1995. "Revictimization by Polygraph The Practice of Polygraphing Survivors of Sexual Assault." *Medicine and Law* 14: 225–67.
- Smith, Rogers M. 1999. *Civic Ideals Conflicting Visions of Citizenship in U.S. History*. New Haven: Yale University Press.

- Sniderman, Paul M., Philip E. Tetlock, and Thomas Piazza. 1992. "Codebook for the 1991 National Race and Politics Survey". Survey Research Center: Berkeley, CA.
- Snyder, Mark, Elizabeth D. Tanke, and Ellen Berscheid. 1977. "Social Perception and Interpersonal Behavior: On the Self-Fulfilling Nature of Social Stereotypes." *Journal of Personality and Social Psychology* 35 (9): 656–66.
- Sommers, Samuel R. 2006. "On Racial Diversity and Group Decision Making: Identifying Multiple Effects of Racial Composition on Jury Deliberations." *Journal of Personality and Social Psychology* 90 (4): 597–612.
- Sonneborn, Liz. 2004. *Miranda v. Arizona: The Rights of the Accused*. 1st ed. Supreme Court Cases through Primary Sources. New York: Rosen Pub. Group.
- Starr, Douglas. 2013. "The Interview: Do Police Interrogation Techniques Produce False Confessions?" *The New Yorker*, December 9.
- Steele, Claude M., and Joshua Aronson. 1995. "Stereotype Threat and the Intellectual Test Performance of African Americans." *Journal of Personality and Social Psychology* 69 (5): 797–811.
- Steele, Claude M., Steven J. Spencer, and Joshua Aronson. 2002. "Contending with Group Image: The Psychology of Stereotype and Social Identity Threat." *Advances in Experimental Social Psychology* 34: 379–440.
- Strakowski, S M, M Flaum, X Amador, H S Bracha, A K Pandurangi, D Robinson, and M Tohen. 1996. "Racial Differences in the Diagnosis of Psychosis." *Schizophrenia Research* 21 (2): 117–24.
- Sunshine, Jason, and Tom R. Tyler. 2003. "The Role of Procedural Justice and Legitimacy in Shaping Public Support for Policing." *Law & Society Review* 37 (3): 513–48.
- Taylor, Marisa. 2013. "National Reconnaissance Office Officials Accused of Contracting Crimes, Retaliation." *McClatchy Newspapers*, August 13.
- Tetlock, Philip E. 2006. *Expert Political Judgment*. Princeton, N.J.; Woodstock, UK: Princeton University Press.
- The Partnership for Civil Justice Fund. 2012. "FBI Documents Reveal Secret Nationwide Occupy Monitoring." <http://www.justiceonline.org/commentary/fbi-files-ows.html>.
- Thomas, Neal J., Padmanabhan Ramnarayan, Michael J. Bell, Prabhat Maheshwari, Shaun Wilson, Emily B. Nazarian, Lorri M. Phipps, et al. 2008. "An International Assessment of a Web-Based Diagnostic Tool in Critically Ill Children." *Technology and Health Care* 16: 103–10.
- Thompson, W. C. 2009. "Painting the Target around the Matching Profile: The Texas Sharpshooter Fallacy in Forensic DNA Interpretation." *Law, Probability and Risk* 8 (3): 257–76.
- Travis, Jeremy. 2013. "Assessing the Burden of Crime and the Criminal Sanction: A Public Health Perspective on Critical Issues in Criminal Justice". Speech September 19, John Jay College of Criminal Justice.
- Trierweiler, S J, H W Neighbors, C Munday, E E Thompson, V J Binion, and J P Gomez. 2000. "Clinician Attributions Associated with the Diagnosis of Schizophrenia in African American and Non-African American Patients." *Journal of Consulting and Clinical Psychology* 68 (1): 171–75.

- Tushnet, Mark. 2006. "The Supreme Court and the National Political Order." In *The Supreme Court and American Political Development*. Lawrence, Kan: University Press of Kansas.
- Tyler, Tom R. 1989. "The Psychology of Procedural Justice: A Test of the Group-Value Model." *Journal of Personality and Social Psychology* 57 (5): 830–38.
- . 2006. *Why People Obey the Law*. Princeton, N.J: Princeton University Press.
- U.S. Congress. 2010. *Anti-Border Corruption Act of 2010*. Pub.L. 111-376.
- . 1988. *Employee Polygraph Protection Act*. 102 Stat. 646, H.R. 1212.
- U.S. Department of Homeland Security. 2010. "FAST Privacy Threshold Analysis." http://epic.org/privacy/body_scanners/EPIC-DHS-FOIA-09-14-11.pdf.
- U.S. Department of Justice, Office of the Inspector General, Evaluation and Inspections Division. 2006. "Use of Polygraph Examinations in the Department of Justice". I-2006-008. <http://www.justice.gov/oig/reports/plus/e0608/final.pdf>.
- U.S. Department of Labor. 2009. "Other Workplace Standards: Lie Detector Tests". *elaws: Employment Law Guide*. <http://www.dol.gov/compliance/guide/eppa.htm>.
- . 2013. "Wage and Hour Compliance Action Data."
- U.S. Government Accountability Office. 2003. "Youth Illicit Drug Use Prevention". GAO-03-172R.
- . 2010. "Efforts to Validate TSA's Passenger Screening Behavior Detection Program Underway, but Opportunities Exist to Strengthen Validation and Address Operational Challenges". Report to the Ranking Member, Committee on Transportation and Infrastructure, House of Representatives GAO-10-763.
- . 2011. "TSA Has Taken Actions to Improve Security, but Additional Efforts Remain". Testimony Before the Subcommittee on National Security, Homeland Defense, and Foreign Operations, Committee on Oversight and Government Reform, House of Representatives GAO-11-807T.
- . 2013. "TSA Should Limit Future Funding for Behavior Detection Activities". Testimony Before the Subcommittee on Transportation Security, Committee on Homeland Security, House of Representatives GAO-14-1582.
- Unger, Roberto M. 1976. *Knowledge & Politics*. New York: Free Press.
- United States Department of State. 2010. "Program and Budget Guide". Bureau for International Narcotics and Law Enforcement Affairs. <http://www.state.gov/documents/organization/131027.pdf>.
- United States Government Accountability Office. 2010. "Mérida Initiative: The United States Has Provided Counternarcotics and Anticrime Support but Needs Better Performance Measures". Report to Congressional Requesters GAO-10-837.
- Van Meter, Larry A. 2007. *Miranda v. Arizona: The Rights of the Accused*. New York: Chelsea House.
- Vrij, Aldert, and Gün R. Semin. 1996. "Lie Experts' Beliefs about Nonverbal Indicators of Deception." *Journal of Nonverbal Behavior* 20 (1): 65–80.
- Waid, William M., Martin T. Orne, and Stuart K. Wilson. 1979. "Effects of Level of Socialization on Electrodermal Detection of Deception." *Psychophysiology* 16: 15–22.

- Walker, Samuel. 2007. *The Color of Justice: Race, Ethnicity, and Crime in America*. 4th ed. The Wadsworth Contemporary Issues in Crime and Justice Series. Belmont, CA: Wadsworth/Cengage Learning.
- Walker, Samuel, Randolph Boehm, and Kermit Hall, ed. 1997. *Records of the Wickersham Commission on Law Observance and Enforcement, Part 1: Records of the Committee on Official Lawlessness*. Microfilm edition. Research Collections in American Legal History. Bethesda, MD: University Publications of America.
- Weaver, Vesla M., and Amy E. Lerman. 2010. "Political Consequences of the Carceral State." *American Political Science Review* 104 (04): 817–33.
- Weber, Max. 2002. *The Protestant Ethic and the "Spirit" of Capitalism and Other Writings*. ed. and trans. Peter Baehr and Gordon C. Wells. New York: Penguin Books.
- . 2004a. *The Essential Weber: A Reader*. London: Routledge.
- . 2004b. *The Vocation Lectures*. Indianapolis: Hackett Pub.
- West, Heather C. 2010. "Prison Inmates at Midyear 2009 - Statistical Tables". NCJ 230113. Prison and Jail Inmates at Midyear Series. Bureau of Justice Statistics. <http://bjs.ojp.usdoj.gov/index.cfm?ty=pbdetail&iid=2200>.
- West, Steven L., and Keri K. O'Neal. 2004. "Project D.A.R.E. Outcome Effectiveness Revisited." *American Journal of Public Health* 94 (6): 1027–29.
- Whitman, G., and R. Koppl. 2010. "Rational Bias in Forensic Science." *Law, Probability and Risk* 9 (1): 69–90.
- Williams, David R., and Selina A. Mohammed. 2008. "Discrimination and Racial Disparities in Health: Evidence and Needed Research." *Journal of Behavioral Medicine* 32 (1): 20–47.
- Wilson, Timothy D. 2011. *Redirect: The Surprising New Science of Psychological Change*. 1st ed. New York, NY: Little, Brown and Company.
- Wilson, Woodrow. 1887. "The Study of Administration." *Political Science Quarterly* 2 (2): 197–222.
- Winter, Nicholas J.G. 2008. *Dangerous Frames: How Ideas about Race and Gender Shape Public Opinion*. Studies in Communication, Media, and Public Opinion. Chicago: University Of Chicago Press.
- Wolf, Naomi. 2012. "Revealed: How the FBI Coordinated the Crackdown on Occupy." *The Guardian*, December 29.
- Yoo, John. 2005. *The Powers of War and Peace: The Constitution and Foreign Affairs after 9/11*. Chicago: University of Chicago Press.
- . 2006. *War by Other Means: An Insider's Account of the War on Terror*. 1st ed. New York: Atlantic Monthly Press: distributed by Publishers Group West.
- . 2010. "... Or a Sensible Bow to Post-9/11 Reality: The Supreme Court May Mitigate the Harm of the President's Weak Antiterror Policies." *The Wall Street Journal*, June 8. Available online at <http://online.wsj.com/news/articles/SB10001424052748704764404575286394060691572>.
- Zelicoff, Alan. 2007. "Positive and Negative Predictive Values of Polygraphs: Results from Published 'Field' Studies." Manuscript.

Zuckerman, Miron, Richard Koestner, and Audrey O. Alton. 1984. "Learning to Detect Deception." *Journal of Personality and Social Psychology* 46 (3): 519–28.

Zuckerman, Miron, Richard Koestner, and Robert Driver. 1981. "Beliefs about Cues Associated with Deception." *Journal of Nonverbal Behavior* 6 (2): 105–14.

Zulawski, David E. 2002. *Practical Aspects of Interview and Interrogation*. 2nd ed. Boca Raton: CRC Press.

The following survey experimental instruments appear in the University Library's digital repository, LIBRA, as supplemental materials:

Chapter 5 – Behavioral decision-making survey experiments:

Appendix - Polygraph chart survey experiment (Experiment 1)

Appendix - Polygraph chart survey experiment (Experiment 2)

Appendix - Polygraph chart survey experiment (Experiment 3)

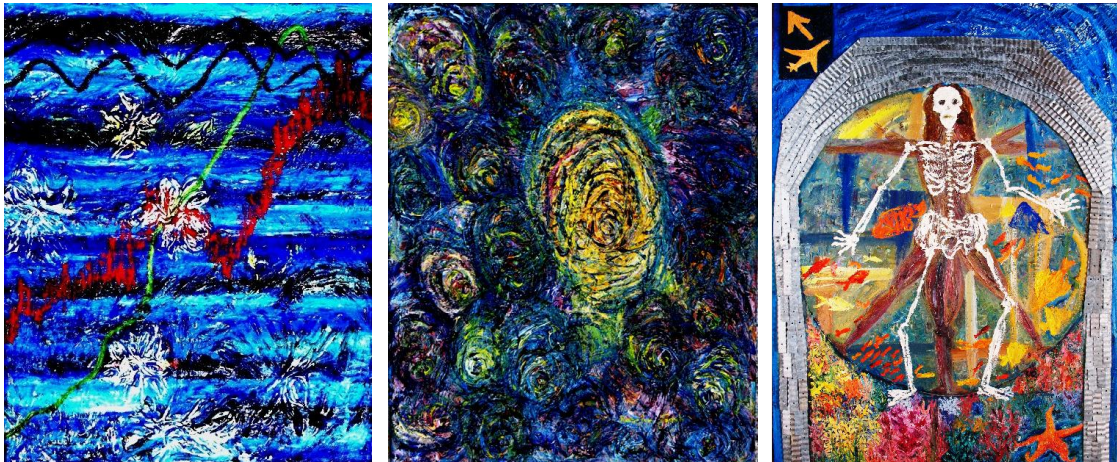
Appendix - Polygraph chart survey experiment (Experiment 4)

Appendix - Medical diagnosis survey experiment (Experiment 5)

Appendix - Welfare benefits survey experiment (Experiment 6)

Chapter 6 – Attitudinal survey experiment

Appendix - Attitudinal survey experiment



“Gasp” and “Fingerprint,” oils on 16” x 20” stretched canvasses; “Vitruvian Woman Floating through Airport Security,” oils and mixed media on 36” x 48” stretched canvas.