THE ROAD TO HELL . . . : THE *STATISTICS PROPOSAL* AS THE FINAL SOLUTION TO THE SOVEREIGN'S HUMAN RIGHTS QUESTION

SAMUEL R. LUCAS*

ABSTRACT

Some statistical analysts advocate using statistical analysis to certify the existence of human rights violations and to identify both victims and perpetrators. Their advocacy has been effective, as prosecutors at the International Criminal Tribunal for the former Yugoslavia used statistical analyses in their indictments, and the International Criminal Court now employs crime pattern analysis in its effort to identify human rights violations. The *statistics proposal* is scrutinized here and found to be exceedingly dangerous–although motivated by a desire to reinforce victims' justice-seeking, it will actually vitiate human rights enforcement. The case against the proposal has four key parts. First a sociological analysis elaborates professional statisticians' good-intentioned attention to human rights issues in sociological terms as a professionalization project, outlines the power-shift the project entails, and suggests the damage that can be done to victims' interests and human rights adjudication as a result of these aspects. Next, subtle epistemological changes in the definitions of crime against humanity and genocide that accompany the use of statistical analysis as a certifying tool are shown to insidiously weaken the foundation for human rights. Then, fundamental

* Professor of Sociology, University of California-Berkeley. A previous version of this paper was presented at the Philosophy and Social Science Conference in Prague on May 14, 2008. Many thanks to Bill Scheuerman, Emily Bauman, Osagie K. Obasogie, Sarah Matari, Camilo Romero and Miranda Massie for incisive comments on an earlier draft. All errors are those of the author. Please direct all correspondence to Samuel R. Lucas / Sociology Department / University of California-Berkeley / 410 Barrows Hall #1980 / Berkeley, CA 94720-1980 or via e-mail to Lucas@demog.berkeley.edu

> dissension among statisticians, masked in discussions of the proposal, actually belies any claim to objectivity; as defendants hire consulting statisticians to exploit these paradigmatic divisions, contradictory analyses will proliferate, rendering statistical research ultimately unable to sustain any charge of a crime against humanity or genocide, *after* the adoption of statistics proposal will have implicitly delegitimated other forms of evidence. The dire nature of this prediction may seem extreme, yet a fourth stage of the analysis considers the adjudication of allegations of discrimination in the United States as a historical cautionary tale, foreshadowing a future where the statistics proposal has rendered statistics *the* basis for assessing human rights issues. If present evidence on discrimination litigation in the United States is any guide, in that future it will be, at best, nearly impossible to substantiate any charge of a crime against humanity or genocide.

I. INTRODUCTION

Both allegations and incidents of crimes against humanity and genocide show no signs of abating.[1] And, efforts to bring alleged

[1] Helen Fein, *Accounting for Genocide After 1945: Theories and Some Findings,*1 INT'L J. ON GROUP RTS. 79, 79 (1993); Barbara Harff & Ted Robert Gurr, *Toward Empirical Theory of Genocides and Politicides: Identification and Measurement of Cases Since 1945*, 32 INT'L STUD. Q. 359–70 (1988); Mark Levene, *Why is the Twentieth Century the Century of Genocide?*, 11 J. WORLD HISTY. 305, 305 (2000).

perpetrators to justice appear to be increasing.[2] Nation-states, their leadership, and their citizens are the alleged perpetrators of these crimes. At present both states and trans-state actors can enforce human rights protections. Thus, states occupy a dual role, a circumstance that produces multiple challenges. One response has been to develop multinational organizations to prosecute such serious charges, but this development has been resisted by some important states.[3]

In this complex context of conflict and multi-sided national and human interest, proposals to institutionalize and standardize evaluative processes and requirements concerning human rights find articulation.[4] The present investigation concerns a standardization proposal that is partially a response to the challenging structure noted above.[5] Under the *statistics proposal*, advocates propose the use of statistical analyses to certify the existence of some types of human rights violation, identify perpetrators of those violations, and aid prosecution of those alleged to have perpetrated those violations.[6] The hope is that use of the standardized instrument of allegedly objective statistics[7] will force investigations and prosecutions of cases that exceed the thresholds identified by statistical analysis. Prosecutors have already begun implementing the proposal, using statistics in multiple indictments, most

---

2 *E.g.*, James Crawford, *The UN Human Rights Treaty System: A System in Crisis?*, *in* THE FUTURE OF UN HUMAN RIGHTS TREATY MONITORING 1, 1 (Philip Alston & James Crawford eds., 2000); Carol M. Glen & Richard C. Murgo, *United Nations Human Rights Conventions: Obligations and Compliance*, 31 POL. & POL'Y 596, 597, 611 (2003).

3 *See, e.g.*, Monroe Leigh, *The United States and the Statute of Rome*, 95 AM. J. INTL. L. 124, 124 (2001).

4 *E.g.*, Gates Garrity-Rokous & Raymond H. Brescia , *Procedural Justice and International Human Rights: Towards a Procedural Jurisprudence for Human Rights Tribunals*, 18 YALE J. INT'L. L. 559, 603 (1993); Allison Marston Danner, *Enhancing the Legitimacy and Accountability of Prosecutorial Discretion at the International Criminal Court*, 97 AM. J. INT'L L. 510, 552 (2003); Evan J. Wallach, *The Procedural and Evidentiary Rules of the Post-World War II War Crimes Trials: Did They Provide an Outline for International Legal Procedure?*, 37 COLUM. J. TRANSNAT'L L. 851, 882 (1999).

5 Herbert F. Spirer & William Seltzer, *Obtaining Evidence for the International Criminal Court Using Data and Quantitative Analysis*, *in* STATISTICAL METHODS FOR HUMAN RIGHTS 195, 195 (Jana Asher et al. eds., 2008).

6 *E.g.*, STATISTICAL METHODS FOR HUMAN RIGHTS (Jana Asher et al. eds., 2008).

7 See Russel Lawrence Barsh, *Measuring Human Rights: Problems of Methodology and Purpose*, 15 HUM. RTS. Q. 87, 121 (1993), for an example of criticizing current standards of statistical objectivity while still accepting the possibility of objective statistics for human rights. Many statistical works claim to produce objective results. *See, e.g.*, Yongtao Guan, Michael Sherman & James A. Calvin, *A Nonparametric Test for Spatial Isotropy Using Subsampling*, 99 J. AM. STAT. ASS'N 810, 810 (2004); Zhiqiang Tan, *A Distributional Approach for Causal Inference Using Propensity Scores*, 101 J. AM. STAT. ASS'N 1619, 1630 (2006).

notably in the indictment of Slobodan Milošević.[8] I contend, however, that the statistics proposal is inherently flawed and exceedingly destructive. Despite the gain it may appear to offer, not only will it likely fail in the vast majority of specific instances, but its cumulative impact will likely subvert the developing transnational effort to protect human rights, substituting an impotent procedural formalism for a muscular commitment to substantive justice. If this occurs the statistics proposal will have satisfied sovereigns' desire to resist evaluation by rendering evaluations ritually thorough, yet almost certain to fail to confirm either the existence of a crime or the identity of the perpetrators.

The argument draws on sociology, history, statistics, and epistemology. Part I first relates key implications of the relation between the nation-state and transnational human rights, positioning the nation-state and, specifically, the head of state, as structurally predisposed to resist the legitimation of extra-state evaluation. It further conveys working definitions of crimes against humanity and genocide, highlighting key aspects of those definitions. After describing specific problems of evidence collection and adjudication that accompany consideration of crimes against humanity and genocide, Part I then relates two responses to the challenge: 1) the historic solution and 2) the statistics proposal.[9]

Parts II through V demonstrate why the statistics proposal should not be implemented. Part II reinterprets professional statisticians' good-intentioned attention to human rights issues in sociological terms as part of a professionalization project, outlines the power-shift the project entails, and suggests the diminution of human rights that follow. Part III outlines important epistemological changes in the definition of crime against humanity and genocide occasioned by using statistics as an evaluative tool, changes that insidiously weaken the foundation for human rights. Part IV illuminates fundamental dissensus among statisticians, contestation that runs far deeper than that involved in the usual expert versus expert conflict common in courtrooms around the world. And, Part V, owing to historical similarity between discrimination and human rights violation,[10] takes the adjudication of accusations of

---

[8] *E.g.*, Prosecutor v. Milošević, Milutinović, Šainović, Ojdanić & Stojiljković, Case No. IT-99-37-5, Indictment, ¶ 4 (May 23, 1999).

[9] *E.g.*, STATISTICAL METHODS FOR HUMAN RIGHTS, *supra* note 6.

[10] GORDON W. ALLPORT, THE NATURE OF PREJUDICE (Addison-Wesley Publishing Company, Inc. 1958); SAMUEL ROUNDFIELD LUCAS, THEORIZING DISCRIMINATION IN AN ERA OF CONTESTED PREJUDICE: DISCRIMINATION IN THE UNITED STATES, VOLUME 1 (2008).

discrimination in the United States as a historical cautionary tale, foreshadowing a future where the statistics proposal has rendered statistics indispensable for investigating human rights allegations and indicting and prosecuting human rights violators. However, in that future, it may be exceedingly difficult and perhaps impossible to substantiate any contested accusation of a crime against humanity or genocide, even though no architect of the statistics proposal may have that outcome as their ultimate goal.

The case begins by first considering the structure of state interest in relation to human rights enforcement.

II. STATE SOVEREIGNTY AND HUMAN RIGHTS: CONTEXT, DEFINITIONS, AND RESPONSES

A. STATE SOVEREIGNTY AND THE THREAT OF HUMAN RIGHTS

States' are both key enforcers and possible transgressors of human rights, and one problem that follows from states' dual role is that sometimes sovereigns attend to alleged human rights violations as part of a search for advantage rather than justice.[11] Thus, even setting aside the possibility of unreflective cultural imperialism,[12] we can recognize that sovereign nations may allege crimes against humanity or genocide in pursuit of other goals, perhaps as part of a strategic propaganda offensive.[13]

Two key results follow. First, states that resist dominant states can be saturated with attention while horrendous violations perpetrated by allies of dominant states, or the dominant states themselves, are not the subject of human rights enforcement because attending to them will either not further, or perhaps may even endanger, dominant nation-states' perceived geo-political interests.[14] Indeed, some analysts note that

---

[11] See David Chandler, *Rhetoric Without Responsibility: The Attraction of 'Ethical' Foreign Policy*, 5 BRIT. J. POL. & INT'L REL. 295, 295 (2003), for additional complexities that may lead the sovereign to articulate a commitment to human rights in foreign policy while failing to follow the words with action.

[12] Jack Donnelly, *The Relative Universality of Human Rights*, 29 HUM. RTS. Q. 281, 303 (2007).

[13] *E.g.*, Ken Roth, *War in Iraq: Not a Humanitarian Intervention*, *in* HUMAN RIGHTS WATCH WORLD REPORT 2004 13, 13 (2004), *available at* http://www.hrw.org/wr2k4/3.htm.

[14] *E.g.*, Roberta Cohen, *People's Republic of China: The Human Rights Exception*, 9 HUM. RTS. Q. 447, 474 (1987); Edward S. Herman, *The United States Versus Human Rights in the Third World*, 4 HARV. HUM. RTS. J. 85, 104 (1991); Michael G. Palmer, *Compensation for Vietnam's Agent Orange Victims*, 8 INT'L J. HUM. RTS. 1, 2 (2004).

leaders of dominant states were not charged with crimes against humanity despite executing individual juveniles or killing thousands of non-combatants with carpet-bombing, while leaders of poorer nations have been subject to human rights violation indictments.[15]

Second, because of past false accusations of human rights violations, state authorities and the wider public may be especially skeptical of any human rights allegations.[16] As Power maintains, one reason reports of Nazi atrocities were slow to galvanize authorities and wider publics was that both had learned that allegations of German excesses in World War I (e.g., the Belgian atrocities) had been grossly over-stated.[17]

The importance of these implications of states' dual role vis à vis transnational human rights, which we might term the *rhetorical resource* implications, is difficult to overstate. Yet, a second string of implications from the dual role is also relevant. In this second line of implication the sovereign state resists being bound by the judgment or authority of others.[18] This second line of implication reflects the sovereign's general interest to remain unfettered. As such, this line of implication, which we might term the *sovereign immunity* line, challenges the establishment of logics or extra-state entities that might render judgment of the behavior of the state, especially with respect to its own subjects or citizens.[19] Thus, this second line directly implicates the question of the limits of sovereignty.

Many theorists have debated the basis and implications of state sovereignty.[20] Although there are few statements about sovereignty to

---

[15] *E.g.*, Amy C. Harfeld, Article, *Oh Righteous Delinquent One: The United States' International Human Rights Double Standard – Explanation, Example and Avenues for Change*, 4 N.Y. CITY L. REV. 59, 60 (2001); Richard Gwyn, *International Law Should Not be Victor's Justice*, TORONTO STAR, July 4, 2001, at A19.

[16] SAMANTHA POWER, "A PROBLEM FROM HELL": AMERICA AND THE AGE OF GENOCIDE 36, 505 (Perennial ed., 2003) (2002).

[17] *Id.* at 36 & 524 n.21 (explaining that at the outbreak of World War II analysts had debunked the claim of Belgian atrocities, but post-World War II research later confirmed the Belgian atrocities).

[18] *E.g.*, A. Lawrence Lowell, *The Limits of Sovereignty*, 2 HARV. L. REV. 70, 84 (1888).

[19] *See id.* at 84–85.

[20] *See* EDMUND BURKE, REFLECTIONS ON THE REVOLUTION IN FRANCE 247 (L.G. Mitchell ed., Oxford Univ. Press 1993) (1790); GEORG WILHELM FRIEDRICH HEGEL, ELEMENTS OF THE PHILOSOPHY OF RIGHT 275 (H. B. Nisbet trans., Allen W. Wood ed., Cambridge Univ. Press 1992) (1820); THOMAS HOBBES, LEVIATHAN, at I (W.G. Pogson Smith ed., Clarendon Press, 5th prtg. 1958) (1651); JOHN LOCKE, TWO TREATISES OF GOVERNMENT 445–46 (Peter Laslett ed., Cambridge Univ. Press 1964) (1689); JOHN RAWLS, THE LAW OF PEOPLES (1999); JEAN-JACQUES ROUSSEAU, THE SOCIAL CONTRACT 5 (Charles Frankel ed., Hafner Publ'g Co. 9th prtg.

which all interlocutors would agree, it is clear that many states have relinquished some of their sovereignty by delegating responsibility for trying crimes against humanity and genocide to transnational authorities.[21] However, some key states have yet to fully embrace that effort.[22] Notably, although the policy of the United States toward the developing transnational human rights institutional structure vacillated during the Clinton administration, since 2002 the United States has actively opposed it.[23]

In resisting the establishment of extra-state institutions and standards concerning human rights, the United States remains consistent with its long-standing reluctance to cede its sovereignty.[24] Undoubtedly not alone in this stance, it remains noteworthy because the United States often articulates foreign policy positions in universalistic terms using a language of liberty and freedom.[25] Despite its appeals to universalism, at the end of World War I, at the inception of the international effort to add force to the protection of human rights, the United States demurred insofar as concrete efforts threatened to subject heads of states to the jurisdiction of authorities of other nations, an act that US negotiators claimed would violate the core logic of sovereignty.[26] American representatives, arguing that heads of states, by *representing* rather than *holding* the people's sovereignty, can only be held accountable by their citizens, concluded that "the essence of sovereignty consists in the fact that it is not responsible to any foreign sovereignty."[27]

1961) (1762). The relevant literature is massive, and these canonical works only brush the surface.

21 WILLIAM A. SCHABAS, AN INTRODUCTION TO THE INTERNATIONAL CRIMINAL COURT 19 (2001).

22 Stefanie Grant, *The United States and The International Human Rights Treaty System: For Export Only?*, *in* THE FUTURE OF UN HUMAN RIGHTS TREATY MONITORING 317, 317 (Philip Alston & James Crawford eds., 2000).

23 Jean Galbraith, *The Bush Administration's Response to the International Criminal Court*, 21 BERKELEY J. INT'L L. 683, 683–85 (2003); Devyani Kacker, *Coming Full Circle: The Rome Statute and the Crime of Aggression*, 33 SUFFOLK TRANSNAT'L L. REV. 285, 286–87 (2010).

24 Louis Henkin, *U.S. Ratification of Human Rights Conventions: The Ghost of Senator Bricker*, 89 AM. J. INT'L L. 341, 348 (1995).

25 See Melvyn P. Leffler, *Bush's Foreign Policy*, 144 FOREIGN POLICY, Sept. – Oct. 2004, at 22 (2004) for discussion of the consistency between Bush administration articulation of U.S. foreign policy in universal terms and other U.S. administrations' articulation of same in universal terms.

26 Report, *Commission on the Responsibility of the Authors of the War and on Enforcement of Penalties*, 14 AM. J. INT'L L. 95, 135–36 (1920).

27 *Id.* at 148.

Debate concerning the status and character of the sovereign continues.[28] Although the inviolability of human rights has been unevenly accepted, the view that human rights are inviolable has become more widely-held.[29] The American view in 1919, that sovereignty trumped human rights, has not prevailed. Thus, as the twentieth century drew to a close, the Clinton administration, aware that using certain terms might obligate the sovereign to act, is reported to have prohibited US officials from referring to events in Rwanda as a genocide.[30] Thus, at one end of the century and at the other, US officials recoiled from protecting targets of genocide.[31]

One may take genocide as a paradigmatic example of human rights violation. Samantha Power indicates that US reluctance to combat genocide has been largely consistent over several decades, and many other states have often adopted a similar reluctance to become involved.[32] Power identifies important structural factors of governing that perpetuate that consistency, concluding correctly that states are unwilling to expend political capital, material resources, and their soldiers' lives on issues they regard as tangential to their own interests and on issues about which domestic publics appear indifferent.[33] Although this conclusion is partly accurate, it does not fully reflect the two important lines of analysis noted above. The rhetorical resource line suggests that the main value of human rights law to the executive authority is as a rhetorical resource in struggles between nation-states on the one hand and against domestic rivals on the other.[34] The sovereign immunity line further asserts that states are reluctant to subjugate themselves to extra-state evaluation, suggesting that any effort to attach obligations of enforcement or minimal standards of behavior to human rights rhetoric will be resisted by at least some noteworthy states.

---

[28] *Compare* JOHN YOO, WAR BY OTHER MEANS: AN INSIDER'S ACCOUNT OF THE WAR ON TERROR (2006), *with* JOSEPH MARGULIES, GUANTÁNAMO AND THE ABUSE OF PRESIDENTIAL POWER (2006).

[29] *See* LYNN HUNT, INVENTING HUMAN RIGHTS: A HISTORY (2007)

[30] David Aronson, *Congo Games*, NEW REPUBLIC, Jan. 5–Jan. 12 1998, at 13.

[31] See POWER, *supra* note 16, at 14, for analysis showing that at the end of World War I US negotiators opposed the holding of war crimes trials to address, among other issues, the Armenian genocide at the hands of Turkish authorities, and explicitly refused to take part should such trials go forward.

[32] *Id.* at 503.

[33] *Id.* at 508–10.

[34] Aaron Wildavsky, *The Two Presidencies*, 4 TRANS-ACTION, Dec. 1966, at 7–11, 14.

The sovereign immunity line of reasoning is crucial, for if states were only reluctant to act in areas they regarded as tangential, there would be little reason for them to resist the construction of transnational structures that release them from at least some of the obligations of enforcement. However, notable resistance to transnational structures remains, and one reason resistance may remain is that endowing any supra-national entity with enforcement authority around such a weighty matter constitutes a potentially serious diminution of state sovereignty.

The relevance of this resistance for the statistics proposal becomes clear in the following sequence of observations. As human rights has become a widely-available frame in the dialogue between states, non-state actors, citizens, and others, sovereigns have found it difficult to publicly repudiate human rights obligations.[36] Yet, sovereigns have often been able to delay intervening to stop on-going transgressions,[37] to avoid investigating allegations,[38] to decline to prosecute accused parties,[39] and to dodge indictment for even their admitted violations.[40] Although often successful, the ad hoc nature of the effort to avoid responsibility renders each new atrocity a risk-laden occasion in which the moral authority and legitimacy of the alleged transgressors, the structures through which they acted (e.g., democratic governments, theocratic governments), and any governments that refuse to intervene all may face critical scrutiny.

The view that human rights law imposes a unique limit on sovereignty is over-stated,[41] such that the emerging human rights regime

---

[35] *E.g.*, Aaron Fichtelberg, *Democratic Legitimacy and the International Criminal Court: A Liberal Defence*, 4 J. INT'L CRIM. JUST.765 (2006); John R. Worth, Note, *Globalization and the Myth of Absolute National Sovereignty: Reconsidering the "Un-signing" of the Rome Statute and the Legacy of Senator Bricker*, 79 IND. L.J. 245 (2004). *See* Michalel P. Scharf, *The Politics Behind U.S. Opposition to the International Criminal Court*, BROWN J. WORLD AFF., Winter/Spring 1999, at 97 (noting that seven nations did not sign the treaty that produced the International Criminal Court: the United States, China, Libya, Qatar, Israel, Iraq, and Yemen).

[36] Lee A. Steven, Note, *Genocide and the Duty to Extradite or Prosecute: Why the United States is in Breach of Its International Obligations*, 39 VA. J. INT'L. L. 425, 447 (1999).

[37] *E.g.*, HERBERT DRUKS, THE FAILURE TO RESCUE (1977); Roger J. Moran, *Confronting Genocide in Africa*, NATIONAL WAR COLLEGE 1–3 (Mar. 28, 2003), http://www.dtic.mil/dtic/tr/fulltext/u2/a442597.pdf.

[38] *E.g.*, POWER, *supra* note 16, at 36, 504–06.

[39] *E.g.*, *id.* at 14–16, 489–90.

[40] GEORGE W. BUSH, DECISION POINTS 169 (2010).

[41] *See* Stephen D. Krasner, *International Political Economy: Abiding Discord*, 1 REV. INT'L POL. ECON. 13 (1994) (providing notable historical examples of other limits on sovereignty).

is not the first threat to sovereignty nation-states have perceived.[42] However, it is a serious high stakes threat. We can historicize the seriousness of the stakes involved here by considering states' response to an earlier perceived threat.[43]

As European principalities engaged in nation-building, citizenship in these developing nation-states was based, at least in part, on ethnic, cultural, or religious categorization.[44] Emergent state authorities interpreted ethnic, cultural, or religious minorities as posing a problem–a threat–and especially regarded Jews as unreliable minorities within the emerging nation-state.[45] Various solutions to "the Jewish Question" were offered by those hostile to and amicable toward European Jewry. As Dawidowicz notes, "The classic illustration is the 'solution' offered by Constantine Pobyedonostsev, chief adviser to Czar Alexander III, in 1881: one-third of the Jews were to emigrate, one-third to convert, and one-third to die of hunger."[46] This particular proposal has explicitly troubling elements, yet, all proposals, by presupposing the lesser status of Jews and other minorities are implicitly troubling. Thus, the very articulation of "the Jewish Question," marks states' failure to constitute themselves in supportive relation to human rights.

Seen in this way, the holocaust was a straightforward extension of the logic that had long-buttressed European states, a logic not only of anti-Semitism but, more fundamentally, a logic that denied religious, cultural, or other minorities equal status. In other words, the logic upon which the nation-state was founded denied *human* rights. Escalating Pobyedonostsev's proposal, the Nazis dubbed their response the Final Solution to the Jewish Question because no Jews would survive, no progeny would follow, and thus no bearers of a culture and ethos that empowered persons to resist full submission to the dictates of the state would likely remain.[47]

---

[42] *E.g.*, Robert Jackson, *Sovereignty in World Politics: A Glance at the Conceptual and Historical Landscape*, 47 POL. STUD. 441, 453 (1999).

[43] I reference *perceived* threat because the *view* that a threat exists is the issue, not the actual material existence of a threat. In the case discussed here the state is not, actually, threatened, only a logic of states' possession of unfettered discretion is.

[44] *E.g.*, Jennifer Jackson Preece, *Minority Rights in Europe: from Westphalia to Helsinki*, 23 REV. INT'L STUD. 75, 78, 82 (1997).

[45] Jews were one of several categories so-regarded. Most notably, the Roma were regarded in a similar way. Just as notably, the Nazis targeted the Roma in the Porajmos.

[46] LUCY S. DAWIDOWICZ, THE WAR AGAINST THE JEWS 1933–1945, at xiv (1975).

[47] *See id*.

Seen in this historical perspective, use of the term "Final Solution" here is not hyperbole, for it references the way in which extra-state adjudication of human rights allegations challenges state power, a challenge akin to that posed by disenfranchised minorities in the emerging nation-states.[48] In the present case the sovereign faces a "Human Rights Question": how may the sovereign maintain legitimate unfettered discretion over those persons subject to their control? While those who regard human rights as inviolable may see the question as an implied shield of protection in that there really is no way to legitimate unfettered discretion over persons while maintaining human rights,[49] at least some sovereigns see this question as articulating a problem they need navigate to buttress unfettered discretionary authority over those subject to their power.[50] Any solution that would forever foreclose the possibility of bringing actionable accusations of human rights violations would, from the point of view of such states, provide a welcomed final solution to the sovereign's "Human Rights Question."[51]

From the sovereign's perspective, such a solution will maintain states' ability to use human rights allegations as a rhetorical resource yet render futile any claims as to observer states' responsibility to prevent atrocities, investigate alleged human rights violations, and prosecute alleged perpetrators. Such a solution might make processes and

---

[48] A voluminous literature has explored the causes of the holocaust. Analysts' explanations have ranged from macro-structural, to the cultural, to the psychological. *See generally* ZYGMUNT BAUMAN, MODERNITY AND THE HOLOCAUST (1989) (highlighting a brutalizing logic inherent in modern bureaucratic states); DANIEL JONAH GOLDHAGEN, HITLER'S WILLING EXECUTIONERS (1996) (arguing that an anti-Semitic culture predisposed citizens to willingly participate); Arthur G. Miller, *What Can the Milgram Obedience Experiments Tell Us About the Holocaust?: Generalizing from the Social Psychology Laboratory*, *in* THE SOCIAL PSYCHOLOGY OF GOOD AND EVIL (Arthur G. Miller ed., 2004) (demonstrating a predisposition on the part of many to accept authority). All provide a banal foundation for evil. *See* HANNAH ARENDT, EICHMANN IN JERUSALEM: A REPORT ON THE BANALITY OF EVIL (1963). Without delving into that literature, it is clear that the holocaust had many causes. Without evaluating other possible causes, I have highlighted one relevant for our concerns.

[49] I distinguish *power*, which is the ability to accomplish some outcome, and *authority*, which is the *legitimated* ability to accomplish some outcome.

[50] YOO, *supra* note 28, at 241; Elizabeth M. Iglesias, *Article II: The Uses and Abuses of Executive Power*, 26 U. MIAMI L, REV. 181, 181 (2008).

[51] Notably, "the Jewish Question" pre-supposed the second-class status of minorities. The Nazi Final Solution to the question murdered millions of Jews, Roma, homosexuals, disabled persons, and other alleged enemies of the state. The post-war trials of Nazi officials provided key precedents for the development of transnational human rights institutions. Yet, the "Human Rights Question" re-states sovereigns' view of the requirements of sovereignty as unfettered discretion over all subjects. Thus, "the Jewish Question"–which denied the humanity of minorities–and its Final Solution is inextricably linked to "the Human Rights Question"–which denies the implications of the humanity of all–and any threatened Final Solution of it.

investigations ritually thorough, yet would also produce few (if any) certified violations, resulting in few (if any) obligatory interventions and few (if any) prosecutions. Such a solution could even reinforce the moral legitimacy of the sovereign, maintaining their ability to eloquently decry alleged crimes against humanity and genocide in pursuit of any other aims they may have, while allowing them to defer to the results of a technical investigation that is unlikely to necessitate intervention.[52] Thus, such a solution would practically limit the chance that states would be obligated to follow any denunciation with consequential or reparative action or, more important, ultimately be required to answer in court for any human rights violations they themselves might be accused of committing. If such a process could be found or devised, it would resolve the predicament in which human rights logic places states and heads of states in favor of sovereign immunity. Of course, such a resolution would destroy the emerging[53] transnational effort to protect the rights and bodily integrity of all human beings. In this sense, such a resolution would be a Final Solution.

The question addressed here is whether the statistics proposal constitutes a Final Solution. I hasten to add, positing the question does not imply that any statistician has conceived of the statistics proposal in such terms. Multiple lines of social science research recognize that, owing to the complexity of any decision-making opportunity, limits on human information-processing (either of an individual or of a collective body), and the predispositions persons bring to the decision-making moment, actors act in a complex environment such that the result of any given proposed course of action is often not only *not* the basis of its promulgation or adoption but, also, often is not even visible to many advocates of the course of action.[54] Thus, the question does not imply an accusation of intent.

Although it is unlikely that sovereigns have seen the statistics proposal as an answer to the sovereign's Human Rights Question, research indicates that executives sometimes resist fundamental change not by direct confrontation, but instead, by adopting symbolic changes

---

[52] See Chandler, *supra* note 11, for a general discussion of how states may seek legitimacy through using human rights rhetoric.

[53] Hunt, *supra* note 29.

[54] *E.g.*, Michael D. Cohen et al., *A Garbage Can Model of Organizational Choice*, 17 ADMIN. SCI. Q. 1, 19 (1972); Karl E. Weick, 21 *Educational Organizations as Loosely Coupled Systems*, ADMIN. SCI. Q. 1, 1 (1976); Charles R. Schwenk, *Cognitive Simplification Processes in Strategic Decision-making*. 5 STRATEGIC MGMT. J. 111, 123 (1984); Ann Swidler, *Culture in Action*, 51 AM. SOC. REV. 273, 284 (1986).

that substitute for material change, thus blunting the movement for social change while solidifying their power.[55] However, executives' ability to implement this strategy of symbolism requires the presence of alternatives that would have that effect. They may embrace such an alternative once its effects are visible. I contend that the realities of statistics, as opposed to the stylized facts of statistics, make the statistics proposal just such an alternative. Thus, my thesis, posed as a question: Does the statistics proposal constitute a Final Solution to the emerging effort to enforce transnational human rights? To address this question we need first provide a working definition of the crimes with which we are concerned, crimes against humanity and genocide.

B. WORKING DEFINITIONS OF CRIMES AGAINST HUMANITY AND GENOCIDE: THE ROME STATUTE

Definitions of human rights in general, and crimes against humanity and genocide in particular, are contested. For example, some analysts have contested the absence of economic rights from the list of official human rights.[56] To proceed, however, we need not resolve such disagreements; because our focus concerns the official adjudication of allegations of crimes against humanity and genocide, we only require working definitions consonant with definitions used to adjudicate such allegations either in the past, the present, or both.

Articles 6 and 7 of Part 2 of the Rome Statute define genocide[57] and crimes against humanity,[58] two classes of crimes for which the

---

[55] *See, e.g.*, MARY L. DUDZIAK, COLD WAR CIVIL RIGHTS: RACE AND THE IMAGE OF AMERICAN DEMOCRACY (2000).

[56] Analysts have long-noted that some dominant powers and their allies emphasize civil and political rights as human rights, whereas other dominant powers and *their* allies emphasize socioeconomic rights as human rights. *See, e.g.*, David P. Forsythe, *Socioeconomic Human Rights: The United Nations, the United States, and Beyond*, 4 HUM. RTS. Q. 433, 434, 440–41 (1982). Analysts also disagree on whether human rights are individual rights *per se* or, instead, whether group rights, as such, exist. *See, e.g.*, Peter Jones, *Human Rights, Group Rights, and Peoples' Rights*, 21 HUM. RTS. Q. 80, 80 (1999). Notably, the Western industrial nations have tended to adopt the former position, while developing nations have tended to adopt the latter. Yet, complexities exist–the concept of genocide, for example, makes no sense without reference to groups, and thus Western governments have not consistently denied the relevance of group membership, even as definitional inconsistencies may undermine the efficacy and application of the law. See e.g., Beth Van Shaack, *The Definition of Crimes Against Humanity: Resolving the Incoherence*, 37 COLUM. J. TRANSNAT'L L. 787, 834–35 (1999).

[57] Part 2, Article 6 of the Rome Statute reads:

For the purpose of this Statute, "genocide" means any of the following acts committed with intent to destroy, in whole or in part, a national, ethnical, racial or religious group, as such: (a) Killing members of the group; (b) Causing serious bodily or mental harm to members of the

International Criminal Court (ICC) has jurisdiction.[59] ICC jurisdiction transcends state boundaries[60] and covers crimes of international concern,[61] rather than cases that constituted local and national-level authorities are able to investigate and prosecute. The number of persons who are directly harmed in the case of genocide or crime against humanity need not be large; however, by definition, both crimes involve large numbers of persons targeted on the basis of membership in large groups such as race or religion.[62] These provisions effectively limit ICC attention to cases in which the state is actively targeting large numbers of persons or, at least, acquiescing to such activity.

Under the law a mass murderer might not be regarded as a perpetrator of genocide, and a serial rapist or even a gang of serial rapists might not be treated as perpetrators of crimes against humanity, for at least two reasons. First, if the states in which they act pursue their apprehension (more or less), the ICC lacks jurisdiction.[63] Operationally this has meant that only states low in capacity to resist external intervention have been subject to ICC attention. Second, observers may regard victims as unsystematically targeted.[64] For example, the targeting

---

group; (c) Deliberately inflicting on the group conditions of life calculated to bring about its physical destruction in whole or in part; (d) Imposing measures intended to prevent births within the group; (e) Forcibly transferring children of the group to another group.

Rome Statute of the International Criminal Court, art. 6, July 17, 1998, 2187 U.N.T.S. 38544.

58 Part 2, Article 7 of the Rome Statute reads:

For the purpose of this Statute, "crime against humanity" means any of the following acts when committed as part of a widespread or systematic attack directed against any civilian population, with knowledge of the attack: (a) Murder; (b) Extermination; (c) Enslavement; (d) Deportation or forcible transfer of population; (e) Imprisonment or other severe deprivation of physical liberty in violation of fundamental rules of international law; (f) Torture; (g) Rape, sexual slavery, enforced prostitution, forced pregnancy, enforced sterilization, or any other form of sexual violence of comparable gravity; (h) Persecution against any identifiable group or collectivity on political, racial, national, ethnic, cultural, religious, gender as defined in paragraph 3, or other grounds that are universally recognized as impermissible under international law, in connection with any act referred to in this paragraph or any crime within the jurisdiction of the Court; (i) Enforced disappearance of persons; (j) The crime of apartheid; (k) Other inhumane acts of a similar character intentionally causing great suffering, or serious injury to body or to mental or physical health. *Id.* at art. 7.

59 *Id.* at art. 5.

60 *Id.* at art. 4.

61 *Id.* at art. 5.

62 The Rome Statute references intent and incomplete eradication of a group. As intent and/or incomplete eradication may trigger a charge of genocide, the number of material victims needed to indict on charges of genocide is far less than the total population of a group. *Id.* at art. 6.

63 *See* Danner, *supra* note 4, at 517.

64 Intriguingly, this second explanation resembles the official government denials often issued in the wake of charges of human rights violation. *See* Stanley Cohen, *Government Responses to*

of persons by sex is not necessarily seen as evidence of systematicity, a feature some have criticized.[65]

The treatment of women, the contested status of socioeconomic rights, and more render the Rome Statute arguably incomplete. Thus, as some critics imply, the Rome Statute is not the exclusive source of coherent definitions of genocide and crimes against humanity, nor does the treaty enforce every potentially serious international crime. Notably, the crime of aggression was originally included, but was only defined in Kampala in June 2010, twelve years after the original statute, with enforcement to begin no earlier than January 1, 2017.[66] However, without embracing the Rome Statute as the conclusive articulation of human rights definitions, it has the value of offering an institutionalized statute, covering a small set of crimes, a subset of which we consider, during a time in which the statistics proposal is being advanced. Hence, the Rome Statute appears to be a useful definition.

C. THE PROBLEMS OF EVIDENCE COLLECTION AND ADJUDICATION OF ALLEGATIONS OF CRIMES AGAINST HUMANITY AND/OR GENOCIDE

In light of the Rome Statute definitions, what may be lost in the extremity of considering an alleged crime against humanity or genocide is that in some respects crimes against humanity and genocide are not unlike other crimes. As in most other crimes, alleged perpetrators may seek to hide their behavior, the crime, or both from detection. Further, witnesses may be wholly unavailable, or, if available, may be intimidated from coming forward. Witnesses who are available may have ulterior motives to any testimony they provide. Ostensible third parties who

---

Human Rights Reports: Claims, Denials, and Counterclaims, 18 HUM. RTS. Q. 517, 536–37 (1996).

[65] For example, the human rights regime has reified a distinction between public and private, effectively shunting support away from women as a class (often) behind closed doors. *See, e.g.*, Celina Romany, *Women as* Aliens*: A Feminist Critique of the Public/Private Distinction in International Human Rights Law*, 6 HARV. HUM. RTS. J. 87 (1993). Indeed, traditionally human rights statutes fail to recognize womens' human rights directly. *See, e.g.*, Catharine A. MacKinnon, *Rape, Genocide, and Women's Human Rights*, 17 HARV. WOMEN'S L.J. 5, 5 (1994).

[66] *Review Conference of the Rome Statute of the International Criminal Court*, INT'L CRIMINAL COURT 6, 17 (June 11, 2010), *available at* http://www.icc-cpi.int/iccdocs/asp_docs/ASP9/OR/RC-11-ENG.pdf.

investigate the allegation may have ulterior motives as well, or even innocent biases that push them toward one or the other conclusion.[67]

All efforts to investigate and prosecute crimes must respond to these conditions. However, crimes against humanity and/or genocides also often have additional conditions that complicate evidence generation and allegation adjudication processes. Most notably, although exceptions arguably exist,[68] research indicates that external war and internal socio-political conflict or civil war predict the existence and timing of crimes against humanity and/or genocide.[69] We need not chronicle the specific results of such upheavals in any detail; it is clear that war, civil war, and civil strife create chaotic conditions, uprooting communities, dispersing individual members, separating families, and disrupting normal activities in ways that may leave those who remain at the mercy of hunger, disease, and ancillary violence. Such conditions can greatly hinder efforts of investigators to determine what happened or is happening.

These conditions exacerbate the problem of evidence and adjudication. Witnesses, both accusatory and exculpatory, may fail to be identified because the chaos may have obscured or even obliterated evidence of their existence or possible value to investigators. Even if investigators know of potential witnesses, chaotic conditions may have temporarily or permanently rendered those potential witnesses inaccessible.[70] Further, records on which prosecutions might be based, even ostensibly benign ones such as requisition orders, receipts, and transportation manifests, may also be unavailable–misfiled, lost, or inadvertently or intentionally destroyed.[71] Finally, unrelated violence (e.g., induced by material scarcity) may mask the existence of state- or non-state-sponsored human rights violations.

---

[67] *See, e.g.*, Christian A. Meissner & Saul M. Kassin, *"He's guilty!": Investigator Bias in Judgments of Truth and Deception*, 26 LAW & HUM. BEHAV. 469, 469 (2002) (discussing investigators' bias toward an assumption of guilt).

[68] For example, in the post-Reconstruction United States (1877–1965), numerous human rights violations occurred. Some victimized individuals only, whereas others, such as at Rosewood, Florida, and Tulsa, Oklahoma, assaulted whole communities. Thomas Dye, *Rosewood, Florida: The Destruction of an African American Community*, 58 THE HISTORIAN 605 (1996); Walter F. White, *The Eruption of Tulsa*, 112 NATION, 909, 909–10, (1921).

[69] *E.g.*, Ted Robert Gurr, *The Political Origins of State Violence and Terror: A Theoretical Analysis*, *in* GOVERNMENT VIOLENCE AND REPRESSION 45 (Michael Stohl & George H. Lopez eds., 1986); Matthew Krain, *State-Sponsored Mass Murder: The Onset and Severity of Genocides and Politicides*, 41 J. CONFLICT RESOL. 331 (1997).

[70] *See* Megan A. Fairlie, *Due Process Erosion: The Diminution of Live Testimony at the ICTY*, 34 CAL. W. INT'L L.J. 47, 65–66, 68 (2003).

[71] *See generally*, Linda Barnickel, *Spoils of War: The Fate of European Records During World War II*, 24 ARCHIVAL ISSUES 7 (1999).

These observations cumulate to one basic point: it would be difficult to overestimate the difficulties investigators and prosecutors encounter in identifying either the existence of a crime against humanity or genocide on the one hand, or the perpetrator of such a crime on the other. Even so, multiple prosecutions of crimes against humanity and genocide have occurred, usually shortly after war or extreme civil strife.[72] These prosecutions indicate not only the way in which human rights investigators proceed in response to the problem of evidence, but also that there *are* ways to proceed that can bring alleged perpetrators to justice other than by using the methods provided by the statistics proposal. The section below briefly describes the historic response that has made possible the formal indictment of accused perpetrators despite the difficult environment in which such actions commonly occur.

D. THE HISTORIC SOLUTION

Multiple cases have followed in the wake of the landmark trials held in the immediate aftermath of World War II in Nuremberg, Tokyo, and Yokohama. Researchers have critically appraised the full set of cases in an effort to discern patterns of change and stability.[73] Their efforts establish the existence of a historically-evolving solution to the challenge of evidence collection and adjudication in cases of crimes against humanity and genocide.

The Nuremberg, Tokyo, and Yokohama trials were arranged and administered by the dominant nation-states of the Allied coalition at the end of World War II, leading many members of the elite at the time, including in the United States, to see the trials as possibly "victor's justice," punitive acts of hypocritical Allied propaganda.[74] One way to possibly rebut this charge was to develop and follow formal trial procedures evoking due process.[75]

An important implication of the multinational jurisdiction of the trials is that no nation's judicial system governed the trials.[76] Although

---

[72] *See* Richard May & Marieke Wierda, *Trends in International Criminal Evidence: Nuremberg, Tokyo, The Hague, and Arusha*, 37 COLUM. J. TRANSNAT'L L. 725, 725 (1999).

[73] *E.g.*, Theodor Meron, *Reflections on the Prosecution of War Crimes by International Tribunals*, 100 AM. J. INT'L L. 551 (2006); POWER, supra note 16.

[74] *E.g.*, Jeremy Rabkin, *Nuremberg Misremembered*, SAIS REV., Summer–Fall 1999, at 81, 86.

[75] Leslie Scheuermann, *Victor's Justice? The Lessons of Nuremberg Applied to the Trial of Saddam Hussein*, 15 TUL. J. INT'L & COMP. L. 291, 291–92 (2006).

[76] May & Wierda, *supra* note 72, at 728.

the trials reflected the prior experiences of the different powers,[77] the resulting procedures reflected conflicting actors' efforts to address multiple conflicting aspects of the situation, including the constellation of power amongst the victors, the political realities in the sites, victors' racial prejudice toward some defeated powers but not others and, notably, the chaotic conditions at the conclusion of hostilities in the theatre of war. Although the result was likely not a self-conscious response in every particular to the chaos of the post-war context, the existence and character of the trials does establish a precedent of how one might proceed in such an environment. That precedent drew on tendencies in different national legal systems as well as both criminal and civil procedures,[78] setting the stage for later development of the institutional form.

May and Wierda contend that the result, and cases that followed, reflected a hybrid form, an *adversarial* model with affinities with the US system when it came to the presentation of evidence, but an *inquisitorial* model closer to European courts when it came to questions of admissibility.[79] Evidence suggests that US prosecutors in the Nuremberg, Tokyo, and Yokohama trials sought to prevent defendants' from receiving rights that would be provided in a domestic trial, and they had some success in that effort.[80] For example, at Nuremberg prosecutors extensively searched German archives for evidence, but allowed the defense access to only the materials introduced at trial.[81] However, over time transnational human rights officials added and implemented protections for the rights of defendants, such that, at present, several principles commonly observed in many domestic legal systems also pertain in transnational human rights adjudications. For example, even though the prosecution has historically borne the burden of proof, the protection against self-incrimination was not as developed in the early trials as it has become in later ones.[82] Further, trials immediately after World War II did not require disclosure of exculpatory materials, but later trials have required prosecutors to disclose materials to defendants

---

77 Wallach, *supra* note 4, at 854.

78 *See* May & Wierda, *supra* note 72, at 727–28.

79 *Id.* at 727.

80 Wallach, *supra* note 4, at 854–59.

81 *E.g.*, Otto Kranzbuhler, *Nuremberg Eighteen Years Afterwards*, 14 DEPAUL L. REV. 333, 336 (1965).

82 May & Wierda, *supra* note 72, at 761–64.

and vice versa.[83] And, as for the right to confront one's accuser, defendants have that right, though owing to the possibility that the trial may be occurring amidst continued crimes against humanity or on-going genocide, the defendant's right is not absolute; it is balanced against the desire to protect witnesses from persecution.[84]

With respect to evidence, in Nuremberg, Tokyo, and Yokohama affidavits were ruled admissible rather than inadmissible as hearsay, and thus substituted for witnesses where witnesses were not available.[85] However, in those trials affidavits from witnesses not subject to the submission of defendants' interrogatories were often stricken from the record.[86] Further, even when affidavits were admitted, the weight given an affidavit varied according to judges' assessment of the authenticity of the document and the extent to which it was corroborated by other sources and emerging understandings.[87]

In sum, the research suggests that the historic solution to the problem of evidence in allegations of genocide and crimes against humanity has been to rely on eyewitness testimony, administrative documents, and physical evidence which, taken together, supply information that coalesces into a coherent gestalt in light of which defendants' actions, guilt, or innocence can be ascertained. Historically, this approach has led to the conviction of some defendants and the exoneration of others,[88] suggesting the procedures are not necessarily simple mechanisms for visiting victors' vengeance upon the defeated. More recent trials have led to the indictment of members of each party to the conflict, further suggesting that victor's justice need not follow from the historic solution.[89] Finally, these most egregious of crimes have historically had no statute of limitations, allowing prosecution whenever incriminating evidence is developed or surfaces, making any escape insecure.[90] Thus, it appears that despite the complex, chaotic, and

---

83 *Id.* at 757–59.

84 *Id.* at 733.

85 Otto Pannenbecker, *The Nuremberg War-Crimes Trial*, 14 DEPAUL L. REV. 348, 352 (1965).

86 May & Wierda, *supra* note 72, at 751.

87 *Id.* at 751–52.

88 Henry T. King, Jr, *Universal Jurisdiction: Myths, Realities, Prospects, War Crimes and Crimes Against Humanity: The Nuremberg Precedent*, 35 NEW ENG. L. REV. 281, 281 (2001).

89 Kelly D. Askin, *Reflections on Some of the Most Significant Achievements of the ICTY*, 37 NEW ENG. L. REV. 903, 905–06 (2003).

90 Christopher C. Joyner, *Strengthening Enforcement of Humanitarian Law: Reflections on the International Criminal Tribunal for the Former Yugoslavia*, 6 DUKE J. COMP. & INT'L L. 79, 85 (1995).

conflictual environment in which atrocities often occur, the historic solution has demonstrated that due process can prevail in such cases.

E. THE STATISTICS PROPOSAL

Spirer and Seltzer clearly articulate the statistics proposal:

> The judicial systems of many countries recognize the value of scientific evidence, including statistical analysis of data in many civil and criminal cases. Quantitative analyses of human rights violations can also have evidentiary value comparable to DNA testing, forensics, and chemical analysis. . . . [W]e describe and discuss evidentiary issues that arise in International Criminal Tribunals that try alleged perpetrators for serious human rights crimes. Our goal is to promote the effective use of statistical and demographic data and methods in those settings.[91]

Spirer and Seltzer contend that statisticians may offer into evidence descriptive statistics and causal analysis,[92] and specifically highlight the prospect of offering data and analysis on overall and group-specific patterns of victimization as well as the identification of specific victims of specific crimes.[93] They argue that statisticians may answer the question of how many persons were harmed, contending that "[I]n the jurisprudence of the International Tribunal widely adopted in the ICTY [International Criminal Tribunal for the former Yugoslavia], the phrase 'widespread' refers to the large-scale nature of the attack and the number of targeted persons, while the phrase 'systematic' refers to the organized nature of the acts of violence and the improbability of their random occurrence."[94] They regard the numbers of victims as politically important, and thus claim analysts must provide numerical estimates.[95] Fortunately, they contend, if good conditions pertain (e.g., no effort to hide bodies, maintenance of the site of the transgression), then accurate

---

[91] Spirer & Seltzer, *supra* note 5, at 195.
[92] *Id.* at 198.
[93] *Id.* at 201.
[94] *Id.* at 200.
[95] *Id.*

counts can be obtained.[96] Further, Spirer and Seltzer claim that the data can be used to interrogate itself, i.e., the data can be used to ascertain the quality of the data, to determine whether the data have been manipulated.[97] Some researchers have provided illustrative analyses that assess whether human rights-relevant data show signs of manipulation.[98]

Notably, not only victims may be identified, but also perpetrators, by noting a rise in deaths associated with the presence of specific persons or organizations in an area.[99]

Given these important roles for statistical analysis, Spirer and Seltzer conclude that "There is a clear need for at least some minimum training in quantitative methods for investigators and others working at the ICC . . . . Of course, . . . the Court will also require access to those with advanced training and specialized experience in statistics and demography."[100]

The statistics proposal as articulated suggests that anyone with advanced graduate level statistical training–including economists, sociologists, demographers, and perhaps others–would be able to advise the court. However, in order to avoid the cumbersome listing of all relevant fields, henceforth the paper will refer usually to statisticians or sometimes to statistical analysts. Further warrant for this label is drawn from recognition that it is the key professional association of statisticians, the American Statistical Association (ASA), which has spearheaded advocacy for the statistics proposal, providing an institutional home for the effort.[101]

The above articulates the statistics proposal as an idea. Part of the statistics proposal has already been implemented. Notably, ICTY prosecutors used statistical evidence in many investigations and prosecutions.[102] During the ICTY, multiple statistical experts from

---

[96] *Id.* at 203.

[97] *Id.*

[98] *See* Douglas A. Samuelson & Herbert F. Spirer, *Use of Incomplete and Distorted Data in Inference About Human Rights Violations, in* HUMAN RIGHTS AND STATISTICS: GETTING THE RECORD STRAIGHT 62, 62–63 (Thomas B. Jabine & Richard Pierre Claude eds., 1992) (demonstrating several examples of data manipulation).

[99] Spirer & Seltzer, *supra* note 5, at 203.

[100] *Id.* at 214–15.

[101] Thomas B. Jabine & Douglas A. Samuelson, *Human Rights of Statisticians and Statistics of Human Rights: Early History of the American Statistical Association's Committee on Scientific Freedom and Human Rights, in* STATISTICAL METHODS FOR HUMAN RIGHTS, 181 (Jana Asher et al. eds., 2008).

[102] *E.g.*, Prosecutor v. Slobodan Milošević, Case No. IT-99-37, Indictment, ¶¶ 97–98 (May 23, 1999),

multiple fields submitted reports on the counts of persons displaced[103] or killed[104] and, in addition, used statistical analysis to assign responsibility for those acts.[105]

Further, a developing institutional structure increasingly supports such use of statistics. The Investigative and Strategies Analysis Unit of the Investigative Division of the Office of the Prosecutor of the ICC now employs statisticians for the purpose of crime pattern analysis.[106] As a further resource, the American Statistical Association Committee on Scientific Freedom and Human Rights is available to aid with statistical questions on human rights data.[107]

These developments indicate that some aspects of the statistics proposal are no longer merely ideas. Indeed, prosecutors are now using statistics to certify crimes against humanity and genocide and to assign culpability. But, defendants have rarely used statistical evidence in their defense.[108] If defendants' routine use of statistical evidence in their

---

http://www.icty.org/x/cases/slobodan_milosevic/ind/en/mil-ii990524e.htm; Prosecutor v. Momčilo Krajišnik, Case No. IT-00-39-A, Judgment, ¶¶ 606, 608 (Mar. 17, 2009), http://www.icty.org/x/cases/krajisnik/acjug/en/090317.pdf; Patrick Ball, et al., *Killings and Refugee Flow in Kosovo March-June 1999: A Report to the International Criminal Tribunal for the Former Yugoslavia* (Jan. 3, 2002), http://www.icty.org/x/file/About/OTP/War_Demographics/en/s_milosevic_kosovo_020103.pdf.

[103] *See* Prosecutor v. Biljana Plavšić, Case No. IT-00-39 & 40/1-S, Sentencing Judgment, ¶ 36 (Feb. 27, 2003), http://www.icty.org/x/cases/plavsic/tjug/en/pla-tj030227e.pdf; EWA TABEAU ET AL., ETHNIC COMPOSITION AND DISPLACED PERSONS AND REFUGEES IN 47 MUNICIPALITIES OF BOSNIA AND HERZEGOVINA, 1991 TO 1997–98: EXPERT REPORT FOR THE CASE OF SLOBODAN MILOŠEVIĆ (IT-02-54) (2003), *available at* http://hague.bard.edu/reports/Ewa_Tabeaupdf.pdf.

[104] *E.g.*, Prosecutor v. Stanislav Galić, Case No. IT-98-29-T, Decision Concerning the Expert Witnesses Ewa Tabeau & Richard Phillips (July 3, 2002), http://www.icty.org/x/cases/galic/tdec/en/020703.pdf; EWA TABEAU ET AL., POPULATION LOSSES IN THE "SIEGE" OF SARAJEVO 10 SEPTEMBER 1992 TO AUGUST 1994: RESEARCH REPORT PREPARED FOR THE CASE OF STANISLAV GALIĆ (IT-98-29-I) (May 10, 2002), *available at* http://www.icty.org/x/file/About/OTP/War_Demographics/en/galic_sarajevo_020510.pdf.

[105] *See* Ball et al., *supra* note 102; Prosecutor v. Blagoje Simić, Case No. IT-95-9-PT, Decision on Prosecutor's Request to Have One Additional Witness Added to her Witness List (Int'l Crim. Trib. for the Former Yugoslavia Sept. 4, 2001), http://www.icty.org/x/cases/simic/tdec/en/10904WGU16297.htm; EWA TABEAU & JAKUB BIJAK, CHANGES IN THE ETHNIC COMPOSITION IN BOSANSKI ŠAMAC AND ODŽAK, 1991 AND 1997 (2001), http://www.icty.org/x/file/About/OTP/War_Demographics/en/simic_samac_and_odzak_010809.pdf.

[106] *E.g.*, *Vacancy: Associate Analyst, The Hague*, UNJOBS.ORG, http://unjobs.org/vacancies/1305608470129 (last visited May 10, 2012).

[107] *Committee on Scientific Freedom and Human Rights*, AM. STATISTICAL ASS'N, http://www.amstat.org/committees/commdetails.cfm?txtComm=CCNPRO05 (last visited May 10, 2012).

[108] I could find no example of a defendant using statistics in their defense.

defense signals full institutionalization of the statistics proposal, then full implementation has not yet occurred.

F. CLARIFICATION AND SCOPE CONDITIONS OF THE EVALUATION OF THE STATISTICS

Four conditions clarify and bound the scope of the analysis. First, statistics proposal advocates claim that statistics provides a means for answering two questions human rights investigators raise: 1) Was a crime against humanity or a genocide committed?, and, if so, 2) Who committed the crime? Though the statistical methods for answering these questions can differ in some ways, the statistics proposal logic concerning them is unitary. Thus, the analysis will address the statistics proposal's limitations for both simultaneously.

Second, the statistics proposal references two kinds of statistics. Statistics is often divided into *descriptive* and *inferential* branches. Descriptive statistics describe or summarize conditions–e.g., 2,996 of the 3,414 townspeople, approximately 87.8 percent of the town population, were found dead. Inferential statistics explain the conditions–e.g., 711 of the villager deaths are attributed to famine, 314 are attributed to disease, and so forth. Statistics proposal advocates indicate that statisticians can offer both descriptive and causal analysis;[109] under the statistics proposal causal analysis is necessary to identify that a human rights violation occurred and to identify victims and perpetrators. Causal analysis entails inferential statistics. Inferential statistics for causal inference is far more complex than descriptive statistics, for the former can require elaborate data manipulation to eliminate alternative possible causes. Such efforts can require multiple, often untestable assumptions. Descriptive statistics, in contrast, simply summarizes the data. One cannot make causal inferences with descriptive statistics, such that the statistics proposal cannot be implemented using descriptive statistics. Thus, in what follows "the statistics proposal" references the inferential mode of statistical work.

The distinction between descriptive and inferential statistics is important, because if we fail to make that distinction we risk crediting inferential statistics for contributing to the proceedings when only descriptive statistics have been used. If we credit inferential statistics for non-inferential statistical tasks, we are more likely to see inferential

---

[109] Spirer & Seltzer, *supra* note 5, at 198.

statistics as a useful tool for *other* tasks that are much more challenging. Yet, if descriptive statistics is all that has been provided, then the act has provided no information as to the value of inferential statistics, i.e., causal analysis, for human rights adjudication.

Third, the statistics proposal concerns the use of statistics in human rights adjudication, not in the academicians' analysis of such issues. Earlier analysts proposed the use of statistics for academic human rights research, and outlined a more complex pathway through which statistical research on human rights might influence policy; notably, these analysts did not suggest statistics could be used to identify specific individual perpetrators.[110] Others responded, raising serious questions about the value of statistics for the academic analysis of human rights violations, owing principally to the non-comparability of data based in the realities of the conditions of data collection and contestation about the concept of interest.[111] However, I accept use of statistics and many other tools for basic social science research. In social science statistics is useful for understanding general phenomena, *not* for establishing the etiology of specific events. For example, inferential statistics may help us understand the phenomenon of war between nations (e.g., what macro-structural conditions may reduce the likelihood of war), but for technical reasons statistics cannot reveal who caused a *particular* war between nations. The distinction between these kinds of questions is of paramount importance.

I accept statistics in social science research but not necessarily in the international courts for two main reasons. First, so rarely does scholarly social scientific research quickly affect policy[112] that it seems unnecessary to rule a common analytic approach out of bounds in an effort to avoid damaging policy. The glacial pace by which social science is translated into law or policy provides ample time for investigators to reassess the findings. Of course, specific tools within statistics and non-quantitative research are not so protected, which leads to the second reason–the way the scholarly process works is that every claim, method, data, or other element of an analysis can become the subject of critical but relatively leisurely attention. Lacking the need to make an immediate

---

[110] Thomas B. Jabine & Richard P. Claude, *Preface* to HUMAN RIGHTS AND STATISTICS: GETTING THE RECORD STRAIGHT, at xii–xiii (Thomas B. Jabine & Richard P. Claude eds., 1992).

[111] *E.g.*, Robert Justin Goldstein, *The Limitations of Using Quantitative Data in Studying Human Rights Abuses*, 8 HUM. RTS. Q. 607, 612–13, 622 (1986).

[112] Carol H. Weiss, *The Haphazard Connection: Social Science and Public Policy*. 23 INT'L J. EDUC. RES. 137, 140–41 (1995).

decision on the freedom or imprisonment of a particular person, scholars in every discipline have time to draw on a plethora of techniques and tools to dig into their chosen field, seeking to excavate evidence relevant to their particular concerns. Within the academy these actors engage in sustained debate about the substance, theories, and methods that constitute the field. Notably, the aim of such research is not to cast individual persons into prison but, instead, to understand natural, social, and behavioral phenomena and, perhaps, to eventually inform policy development and implementation.

Although scholarship falls short of meeting criteria of the Habermasian ideal situation of practical discourse,[113] in scholarly discourse analysts routinely bring a variety of methods to bear, some work and some do not, a vigorous debate can ensue to assess which is which and, even if such debate occurs, either way analysts can continue their research having learned more about the phenomena of their field. In this context, statistics is not a problem.[114] However, proposing to bring statistics to bear in the international criminal courts to determine whether a particular act was committed by particular individuals involving particular others on a particular day is quite another matter. As such, the proposal elicits critical appraisal.

Finally, the statistics proposal not only concerns the application of statistics to the complex process of human rights adjudication, but also calls for statistics to refashion the definition of human rights violation, as noted by Jean-Louis Bodin, former President of the International Statistical Institute, who asserts that "it is possible to use statistical

---

[113] JÜRGEN HABERMAS, MORAL CONSCIOUSNESS AND COMMUNICATIVE ACTION 104–05 (Christian Lenhardt & Shierry Weber Nicholsen trans., MIT Press 2d ed. 1990) (1983).

[114] My acceptance of statistics for social science work is no mere strategic gesture of tolerance toward others' analytic tastes on the way to a more general rejection of statistical research approaches; quite the contrary, the vast majority of my own research is statistical, *e.g.*, Samuel R. Lucas, *Selective Attrition in a Newly Hostile Regime: The Case of 1980 Sophomores*, 75 SOC. FORCES 511 (1996); SAMUEL ROUNDFIELD LUCAS, TRACKING INEQUALITY: STRATIFICATION AND MOBILITY IN AMERICAN HIGH SCHOOLS 18–19 (1999); Samuel R. Lucas, *Effectively Maintained Inequality: Education Transitions, Track Mobility, and Social Background Effects*, 106 AM. J. SOC. 1642 (2001), or self-consciously attempts to construct an edifice for using statistics to study social phenomena social scientifically (*e.g.*,. Samuel R. Lucas, *Hope, Anguish, and the Problem of Our Time: An Essay on Publication of* The Black-White Test Score Gap, 102 TCHRS. C. REC. 463–75 (2000); LUCAS, *supra* note 10; Samuel R. Lucas, Phillip N. Fucella, & Mark Berends, *A Neo-Classical Education Transitions Approach: A Corrected Tale for Three Cohorts*, 29 RES. SOC. STRATIFICATION & MOBILITY 263–85 (2011)). Thus, my published work and my public posture is not hostile to statistics; in fact, evidence is more consistent with my having a preference for statistics. However, I would say that I support statistics for what statistics can do, and accept that there is much statistics–as any human endeavor–cannot at present do.

methods to measure human rights violations, included [sic] in the most extreme situations as genocides or large-scale conflicts (and moreover to contribute to update the definition of genocide from a legal point of view)."[115] The promise of the proposal is, therefore, to enter into the adjudication process for crimes against humanity and genocide and, what is more, to update, i.e., *to redefine the very concept of human rights violation*.

Is this promise cause for hope, or is this promise cause for dread? Viewing the statistics proposal through the lenses of sociology, epistemology, statistics, and history will provide a cumulating critical response to the query.

III. SOCIOLOGICAL ASSESSMENT OF THE STATISTICS PROPOSAL

The statistics proposal is articulated as an offering of statisticians to the wider public, an act of service to the greater good.[116] Some statisticians' point to the case of Argentine government economist Graciela Mellibovsky Saidler as motivating interest in human rights.[117] In 1976, Dr. Saidler documented the dire situation in Buenos Aires slums. Her analysis angered the military government, and she was kidnapped later that year. Eight years later her father asked the American Statistical Association to assist in her case and, shortly thereafter, their efforts produced a response from a former Death Squad member. Apparently she had been tortured and killed. Her remains were never recovered.

Such a tragic, real-life story tugs at the heart, understandably. Unfortunately, it is not the only such story; similarly tragic narratives of the brutal treatment and horrid death of scholars and others working to document atrocities and repair communities accompany statisticians' growing involvement in human rights work, either as individuals or in crafting the developing institutional structure for statisticians' collective human rights efforts.[118] Some statisticians analyzing human rights issues were themselves targeted by repressive regimes, and moved elsewhere

---

[115] Jean-Louis Bodin, *Foreward* to STATISTICAL METHODS FOR HUMAN RIGHTS, at vii (Jana Asher et al. eds., 2008).

[116] *E.g.*, Spirer & Seltzer, *supra* note 5, at 195–226; Jorge L. Romeu, *Statistical Thinking and Data Analysis: Enhancing Human Rights Work*, *in* STATISTICAL METHODS FOR HUMAN RIGHTS 65, 65–85 (Jana Asher et al. eds., 2008).

[117] STATISTICAL METHODS FOR HUMAN RIGHTS, *supra* note 6, at v–vi.

[118] Jabine & Samuelson, *supra* note 101.

while continuing their efforts.[119] Thus, it is likely that many statisticians are motivated by personal experience of repression, by the specific example of Dr. Saidler, by other similar cases, or by some combination of the above. However well-meaning, conscious motivations and social implications are not the same, and, alas, the purity of one does not purify the other. Consequently, it is necessary to consider the statistics proposal sociologically.

A full sociological analysis would involve multiple dimensions of inquiry.[120] A main dimension of analysis, however, would consider the proposal in the context of the process of the professionalization of statistics as a discipline. The analysis asks: does the statistics proposal constitute a moment in a larger professionalization project and, if so, what implications follow?

A. SERVICE OR POWER? TWO DISTINCT VIEWS OF PROFESSIONALS AND PROFESSIONALIZATION

There are two different approaches to defining the concept of a profession: 1) a taxonomic approach, and, 2) a process approach.[121] The taxonomic approach, the more widely known approach, asserts that professionals apply a systematic theory to specific cases based on their judgment as to whether the theory applies, independent of clients' expressed desires (but not antagonistic to clients "real" needs).[122] Professionals' training is long and technically involved, conveying both techniques and principles needed to assess situations and develop

---

[119] *E.g.*, Romeu, *supra* note 116, at 67.

[120] Dimensions of study would include whether and how key actors' citizenship biases their collective work, perhaps thereby exploring the extent to which the statistics proposal does or does not constitute another moment in Western imperialism, as well as the class, gender, sexuality, racio-ethnic category, or other social location of statistical researchers, and how these impact their interaction with, and sympathies for, contexts and subjects of study, research that might excavate hidden assumptions and cultural factors that, on-the-ground, may subtly shape human rights work. In addition, one would analyze existing institutional structures and locations which are supportive of the statistics proposal, an effort that might discern foundational elements of the proposal, unacknowledged blind spots, and possibly unconscious additional motives. Other dimensions of study would also be included for a full sociological analysis. Although such an analysis would be more academically comprehensive it would not alter the findings reported here.

[121] Douglas Klegon, *The Sociology of Professions: An Emerging Perspective*, 5 SOC. WORK & OCCUPATIONS 259, 259–60 (1978).

[122] Douglas E. Mitchell & Charles T. Kerchner, *Labor Relations and Teacher Policy*, *in* HANDBOOK OF TEACHING AND POLICY (Lee S. Shulman & Gary Sykes eds., Longman Publishing Group 1983).

responses.[123] Further, admission is controlled by the field, not outsiders. Most important, owing to the specialized training, only peers are competent to evaluate each other's work. Thus, the profession is free of lay control, as professionals police themselves with little outside oversight.[124] To support solid practice, professionals subscribe to a code of ethics that highlights service to clients and the greater good, de-emphasizing personal motives, including profit.[125]

By this taxonomic definition statisticians are, arguably, a profession. Statistics is an arcane body of knowledge, and statisticians exercise a great deal of judgment over its applicability, control admission to the discipline via graduate school admissions and examinations, police themselves via mechanisms such as peer review, have a code of ethics,[126] and use complex technical non-formulaic procedures. Seen in this way, the statistics proposal can be interpreted as service professionals offer to the world community in support of the greater good of human rights adjudication, just as some maintain.[127]

Sociologists, however, have also offered critical perspectives on professions based on considering the process – professionalization – by which a field becomes regarded as a profession, rather than considering the set of attributes purported to characterize a professionalized field.[128] An internal dynamic, specific to the field, and an external dynamic, concerning the context in which practitioners operate, constitutes the operation of the professionalization process.[129] Under the internal dynamic members and collectivities of would-be professions seek to construct markets[130] and identify services members of the field can claim to offer.[131] Success in this endeavor secures employment and, perhaps, prestige. Resources in this effort include but are not limited to a code of

---

[123] Harold L. Wilensky, *The Professionalization of Everyone?*, 70 AM. J. SOC. 137, 142–45 (1964).

[124] William J. Goode, *Encroachment, Charlatanism, and the Emerging Profession: Psychology, Sociology, and Medicine*, 25 AM. SOC. REV. 903 (1960).

[125] Ernest Greenwood, *Attributes of a Profession*, SOC. WORK, Jul. 1957, at 45, 49–50.

[126] Committee on Professional Ethics, *Ethical Guidelines for Statistical Practice*, AM. STAT. ASS'N (Aug. 7, 1999), http://www.amstat.org/about/ethicalguidelines.cfm.

[127] Jabine & Samuelson, *supra* note 101, at 192.

[128] Julius A. Roth, *Professionalism: The Sociologist's Decoy*, 1 SOC. WORK & OCCUPATIONS 6 (1974).

[129] Klegon, *supra* note 121.

[130] See Neil Fligstein, *Markets as Politics: A Political-Cultural Approach to Market Institutions*, 61 AM. SOC. REV. 656 (1996), for a discussion of the political and cultural foundations of markets and processes of market construction.

[131] Klegon, *supra* note 121.

ethics,[132] a professional association, and the ability to control admission and thus limit supply and/or maintain standards deemed relevant.[133]

The external aspect of the professionalization process draws the analyst to consider the socio-historical context within which a particular claim to professional expertise is articulated. Many claims may be made, but only some may be viable, and still fewer may succeed. The socio-historical context is sufficiently complex that hard-and-fast rules are difficult to identify but, suffice it to say, the implication of a role for context is that professions are not made simply by virtue of the internal characteristics of the field.[134]

Seeing professions as the result of a process of professionalization reinterprets many of the attributes the taxonomic definition highlights. Thus, where a taxonomic perspective sees control over entrance into the field as a key attribute of professions, a process perspective sees control over entrance to the field as a resource for creating and/or maintaining scarcity or even the socio-demographic composition of the field (e.g., white males, Parisians, the celibate), either of which may be used to maintain or increase fees and prestige. Where taxonomists see self-policing as necessary owing to the technical nature of the work and as essential to the independence required for professionals to dispassionately serve clients and the wider society, process-oriented analysts see an effort to concentrate power in the discipline and to reduce public accountability. And, where taxonomists see control over assessing the relevance of their expertise as a bulwark against the power of self-interested parties, process oriented analysts see an indispensable resource in any effort to enter or construct lucrative markets members of the professionalized field might eventually monopolize.

Notably, none of these reinterpretations impugns the conscious motivations of the actors involved; a basic sociological observation is that acts often have meanings and implications beyond those to which

---

[132] Policing the code of ethics is complex owing to the nature of professional knowledge and skill. The knowledge base of professions cannot be so broad that it lacks a technical language, nor so narrow that the application of mere formulas would furnish required answers (*see* Wilensky, *supra* note 123, at 148–49.). If a field lacks a technical language, there is no opportunity for exclusion and thus neither expertise nor policing can be monopolized. Yet, if the work is so technically precise that predictable results follow from formulaic responses, then policing cannot be monopolized, because results provide a sufficient basis for evaluating the competence and integrity of the practitioner. *See* RANDALL COLLINS, THE CREDENTIAL SOCIETY, 132–33 (1979).

[133] COLLINS, *supra* note 132, at 133.

[134] Klegon, *supra* note 121, at 270–74.

the central actors may orient or subscribe. Only an asociological read of the analysis would transform these observations into an accusation. But, the observations stand worthy of attention either way.

Statistics is a long-standing discipline, dating back centuries,[135] and thus some may see statisticians' professionalization project as having been successfully completed decades ago. Yet, American Statistical Association presidents have noted incomplete professionalization or claimed insufficient status for statisticians' contributions,[136] suggesting professionalization is on-going. In general, professionalization is a project whose work is never done; professional associations survive because even ostensibly secure professions can be threatened, if not with immediate extinction, then certainly with a diminution of power and resources. Professional associations are many things, but one thing they are is maintained; two aims of maintaining a professional association are: 1) to *prevent* the need for mobilization by proactive engagement in the environment in a preemptive posture against possible threats, and, 2) to *mobilize* the members of the profession if a perceived threat materializes.[137] Accordingly, the age of a discipline does not determine whether a discipline is engaging in a professionalization project.

As conditions are ever-changing, professions attempt to protect existing markets for their services while simultaneously seeking out opportunities where a plausible case for the relevance of their claimed special expertise can be made.[138] In this connection we may interpret the advent and continued development of the institutional structure for enforcing human rights law as offering one such emerging opportunity. Certainly, the prestige and power of a profession compared to that of other key individual, collective, and institutional actors partly determines how far the claimed expertise can be stretched and still appear relevant in the eyes of important constituencies. Yet, that organized fields make such claims as part of a professionalization project is clear and, thus, it is plausible to read the statistics proposal as one element in the professionalization project of contemporary statisticians.

---

[135] ALAIN DESROSIÈRES, THE POLITICS OF LARGE NUMBERS 16–18 (Camille Naish trans., Harvard Univ. Press 1998) (1993).

[136] *See, e.g.*, J. Stuart Hunter, *Statistics as a Profession*, 89 J. AM. STAT. ASS'N 1, 1 (1994); Jonas H. Ellenberg, *Statisticians' Significance*, 95 J. AM. STAT. ASS'N 1, 1 (2000).

[137] David Knoke, *Associations and Interest Groups*,12 ANN. REV. SOC. 1, 9 (1986).

[138] Ronald L. Akers, *The Professional Association and the Legal Regulations of Practice*, 2 L. & SOC. REV. 463, 477 (1968).

Professionalization projects are not sinister[139] or rare,[140] and thus their existence may not generally be worthy of comment. However, the case of focus here concerns an issue of grave importance. Thus, critical appraisal is necessitated.

B. EMPLOYMENT IMPLICATIONS

Because the statistics proposal is to use statistics to verify human rights allegations, full implementation requires trained statisticians to help administer human rights law. This requirement will multiply employment opportunities for statistical analysts (as consultants, analysts, teachers, expert witnesses, and more).

Further, adoption of the statistics proposal may subtly alter recruitment into the field of human rights adjudication. As statistics, based on abstruse technical processes, becomes the measuring rod for determining whether a crime against humanity or genocide occurred, actors with statistical skills and training may become increasingly central, and actors without those skills and training may become less central. This shift in recruitment can, by itself, produce a shift in focus. The full implications of that shift cannot be sketched until Part III, but, suffice it to say here, individuals hold intellectual commitments and postures toward phenomena. As the recruitment patterns change, one should expect a shift in the commitments and postures of the staff employed to enforce human rights edicts.

C. THE TRANSFER OF POWER TO AN INTERESTED CLASS

The sociological analysis suggests that another key way in which the proposal would likely alter matters is that, should the proposal become fully adopted, any effort to discern whether genocide or other human rights violations have occurred may lack legitimacy if the imprimatur of statisticians is lacking. If so, then volumes of eyewitness testimony will be simultaneously rendered insufficient.

---

[139] Sociologists debate this point. Some sociologists argue that all professionalization *is* exploitive. For example, see, Aage B. Sørenson, *Toward a Sounder Basis for Class Analysis*, 105AM. J. SOC. 1544 (2000), which attempts to solidly define the concept of exploitation and argues that the definition implies that professionals exploit non-professionals. In response, see, Erik Olin Wright, *Class, Exploitation, and Economic Rents: Reflections on Sørensen's "Sounder Basis"*, 105 AM. J. SOC. 1559 (2001).

[140] ANDREW ABBOTT, THE SYSTEM OF PROFESSIONS 1–3 (1988).

Such a result will constitute a transfer of power from victims, witnesses, and prosecutors to statisticians who determine whether the accusations reflect a transgression statisticians can document as bearing statistical marks of a crime against humanity or genocide. If the statistical analysis cannot uncover a pattern to the categorization of the victims, there would allegedly be no genocide or crime against humanity case to bring. Thus, adjudicating authorities lose power to the ostensibly disinterested professional cadre of statisticians under the statistics proposal.

Yet, once we see the statistics proposal, sociologically, as part of a professionalization project, we also see that statisticians – collectively and individually – are interested parties, not disinterested ones. Statisticians' interest is in rendering their discipline indispensable, an interest that requires they undermine the sufficiency of the remaining kinds of evidence, at least implicitly. Thus, most individual statisticians likely have no interest in the outcome of a particular proceeding, but statisticians as a class have an interest in the *course* of the proceedings, an interest served if space is made for statistics in the business of the court. Consequently, they are not disinterested parties.

An important implication of this observation is that the statistics proposal, as part of the professionalization project of statisticians, is narrowly focused on attaining court access. Certainly, proposal advocates probably see statistics as a powerful tool and thus may seek to make it available to aid the court. These two observations have two ramifications. First, it means that statisticians seeking access for their discipline may be less likely to ascertain the problems that accompany the statistics proposal. Thus, oversights may not be owing to malfeasance, corruption, or incompetence, for it is simply the case that every intellectual and institutional commitment comes with lenses that help scholars see some issues while necessarily reducing their chance to see *other* sets of issues.[141] Once we see the statistics proposal as part of a process of professionalization, we recognize that we are unlikely to obtain an exhaustive, critical evaluation of the proposal if we rely solely on statisticians embedded in the professional associations of the discipline, the very associations that provide the institutional base for advocacy for the proposal. A much more diverse set of voices must be

---

[141] Ilana Ritov & Jonathan Baron, *Status Quo and Omission Biases*, 5 J. RISK & UNCERTAINTY 49, 60 (1992); William Samuelson & Richard Zeckhauser, *Status Quo Bias in Decision Making*, 1 J. RISK & UNCERTAINTY 7, 47 (1988).

brought into the dialogue to have any chance of an exhaustive evaluation of the proposal.[142]

Second, the observations also indicate that the proposal was probably not conceived to support the very different goal of sovereign immunity. Thus, it would be a gross error to impute the aim of buttressing sovereign immunity to statisticians, for statisticians only intend to bring to bear a tool they regard as of incredible illuminating power. Still, the good intentions of advocates should not shield the proposal from systematic analysis of its likely implications.

D. REDUCED SUPPORT FOR VICTIMS, NEW SUPPORT FOR DEFENDANTS

Thus, the statisticians' interest as a class, though carried in a well-meaning proposal, is not, to the sociological analysis, fundamentally about improving the proceedings, which becomes clear as soon as we realize that adopting the statistics proposal necessarily establishes previously unavailable grounds for contesting eyewitness reports. Rather than a process in which victims and eyewitnesses confront the accused in the presence of judges prepared to weigh each presentation for coherence and possibly physical forensic and document-based corroboration, a third actor, likely absent from the environment in which atrocities may have occurred, enters the process bearing the crest of objectivity. The very presence of such allegedly authoritative actors in the proceedings at issue will lessen the power of victims' claims, for victims will usually be in no position during the alleged atrocities to determine the degree to which the brutalities they are experiencing are *systematically* perpetrated on a broad, identifiable target population. Running from a possible perpetrator, hiding out of fear, they may have access only to increasingly distressing rumors, just as one might expect in the chaotic environments that nurture crimes against humanity and genocide. Historically, the response to this situation has been to afford witnesses wide latitude; for example, prosecutors have altered

[142] It is certainly the case that "insiders" can oppose a state of affairs. The observation is simply that insiders are less likely to raise critical questions than outsiders.

requirements concerning corroboration in cases of genocidal rape,[143] and made other adjustments owing to the extremity of the conditions.[144]

The statistics proposal, however, promises a different approach; rather than take the extremity of the situation into account, the statistics proposal promises to shore up the grounds for victims' testimony while preserving formal procedures that secure defendants' rights in many domestic court systems. But, because victims' testimony was already secure owing to the historic solution to the evidence problem, the proposal can offer little in that regard. Instead, the major impact of the proposal is in raising the level of support available for the accused. This effect may be desirable up to a point; indeed, many of the impediments to which the defense was subjected in Nuremberg, Tokyo, and Yokohama have been repudiated by the rules of more recent trials.[145] The question, however, is whether adoption of the statistics proposal tips the balance so much in favor of defendants that victims and prosecutors might never prevail.

E. INCREASED OPPORTUNITIES FOR PERPETRATOR MANIPULATION

Serious implications follow from the statistics proposal's opening of another route to rebutting human rights allegations, for some may act to prepare that route even as they prepare to engage in genocide or crimes against humanity. Sovereigns may use state resources to prepare their statistical defense prior to the violent phase of action. History indicates that sovereigns avail themselves of state resources to

---

[143] Kate Fitzgerald, *Problems of Prosecution and Adjudication of Rape and Other Sexual Assaults Under International Law*, 8 EUR. J. INT'L L. 638, 646 (1997).

[144] May & Wierda, *supra* note 72, at 733. These adjustments to the exigencies of the situation are arguably consistent with the partial suspension of some aspects of due process with the imposition of emergency powers, for both can be argued to be required in extreme situations in order to preserve the nation, in the case of emergency powers or a contemporary function of the nation-state system, the provision of justice (in the case of human rights violations). *See*, William E. Scheuerman, *Emergency Powers and The Rule of Law after 9/11*, 14 J. POL. PHIL. 61, 71 (2006); John Ferejohn & Pasquale Pasquino, *The Law of the Exception: A Typology of Emergency Powers*, 2 INT'L J. CONST. L. 210, 210–39 (2004). See Scheuerman, *supra*, at 62, for a legal debate as to whether emergency powers necessarily violate civil liberties and on whether emergency powers violate the rights of accused parties. Without embracing any specific emergency powers regime, we note that just as history indicates that emergency powers often entail changes in the processes deemed sufficient to constitute due process, historical precedent indicates adjustments in due process procedures have been acceptable to constituted legal authority in adjudicating allegations of crimes against humanity and genocide. *E.g.*, Fitzgerald, *supra* note 143.

[145] May & Wierda, *supra* note 72, at 757–63.

violate their citizens' and subjects' human rights. As one example, note that the United States Census Bureau–the primary statistical agency of the US federal government–aided the seizure of those of Japanese ancestry during World War II. The Census Bureau provided the military with information on neighborhoods' demographic composition[146] to aid the targeting of efforts to seize those of Japanese ancestry, while Congress legalized Census Bureau release of data on individual persons, data that had been collected with a promise of confidentiality.[147] Those seized had most of their property confiscated, and they were moved to concentration camps for the duration of the war. Thus, the United States Census Bureau aided in this human rights violation, indicating that even ostensibly democratic nations have used their statistical agencies to aid their efforts to violate human rights.[148] Thus, it is not far-fetched to expect a general or executive to call on government statisticians to aid or advise on how to proceed, including how to proceed so as to mask their planned human rights violations.[149] Indeed, with sufficient lead time before the violent phase of the atrocities, even censuses can be manipulated to understate the very presence of the targeted population, a preemptive manipulation that would make statistical documentation of later victimization, even murder (of people said never to have existed), extremely difficult. Thus, should statistical evidence take precedence in determining the existence of and culpability for crimes against humanity or genocide, government statisticians will be empowered thereby to aid the state in violating persons' rights by limiting the risk of discovery.

Although some statisticians claim data fabrication or manipulation (henceforth fabrication) is discoverable,[150] it should be obvious that no matter how many discoveries of fabrication occur, one cannot prove that all data fabrications are discovered, much less that they

---

[146] William Seltzer & Margo Anderson, *Using Population Data Systems to Target Vulnerable Population Subgroups and Individuals and Incidents*, *in* STATISTICAL METHODS FOR HUMAN RIGHTS, *supra* note 5, at 296.

[147] Margo Anderson & William Seltzer, *Challenges to the Confidentiality of U.S. Federal Statistics, 1910-1965*, 23 J. OFFICIAL STAT. 1, 21 (2007).

[148] See Seltzer & Anderson, *supra* note 146, at 292–93, for at least 17 examples of state consideration or use of statistical agencies to violate human rights, and see RICHARD PIERRE CLAUDE, SCIENCE IN THE SERVICE OF HUMAN RIGHTS, 113 (2002), for an additional example of state manipulation of statistical data that might have become relevant for subsequent investigation of human rights violations.

[149] All that would be needed to hinder discovery is advice on how to keep offenses low enough to fall within sampling error of a feasible analysis. As I show below, this can still entail large numbers of offenses.

[150] Theodore P. Hill, *The Difficulty of Faking Data*, CHANCE, Summer 1999, at 27, 28.

are discovered in time to matter for relevant proceedings. Indeed, identifying noteworthy discoveries of fabrication or counting the number of discovered data fabrications may produce a stronger belief in discoverability than is warranted. The appropriate evaluation requires a comparison of the total number of discovered fabrications and the total number of *un*discovered fabrications. A solid evaluation would calculate the proportion of all fabrications that are ever discovered and assess the discovery rate over time (e.g., how long after the fabrication is it discovered?). However, because one cannot know the number of undiscovered fabrications, one cannot obtain the denominator needed to estimate either the proportion of fabrications discovered or how long it takes to discover them, and thus the required statistics are unidentified. Hence, confidence in discoverability is unwarranted.

The chance of discovery is probably further reduced in that the same tools used to discover fabrication can be used to construct better data fabrications. For example, Benford's law states that the first digits of naturally occurring data (e.g., lengths of rivers in the world, accounting transactions in a firm) follow a specific logarithmic function.[151] A statistician aware of this law can assure that digits in their fabricated data, and various reasonable subsets of that data, conform to it. Indeed, every tool used to identify fabricated data can be used by the data fabricator to assure the fabricated data does not trip the alarm of the assessment tool. A primary reason the tools often work is that many data fabricators are unaware of the tools. However, a state attempting to hide its data fabrication may allocate the resources needed to discover and use those tools.[152]

Pulling together the threads of the sociological analysis suggests that full implementation of the statistics proposal will provide powerful support to *de facto* sovereign immunity. Hence, despite statisticians'

---

[151] Kuldeep Kumar & Sukanto Bhattacharya, *Detecting the Dubious Digits: Benford's Law in Forensic Accounting*, 4 SIGNIFICANCE 81, 81 (2007).

[152] One might maintain that atrocities are carried out by mobs, and it will be difficult to hide such behavior. Yet, some crimes against humanity and genocides have not principally involved mob violence. *See*, *e.g.*, DAWIDOWICZ, *supra* note 46. Further, a great deal of mob violence is instigated by elites, either directly at the time or indirectly by articulating grievances in divisive ways. *See, e.g.*, Peter Uvin, *Reading the Rwandan Genocide*, INT'L STUD. REV. Fall 2001, at 75, 79–81. If elites can facilitate or instigate mob violence, it is at least plausible to presume that elites may successfully limit mob violence if such violence undermines other elite aims (*e.g.*, avoiding detection of a crime against humanity or genocide). Thus, it seems unwise to rely on the existence of uncontrollable mob violence to reassure us that adoption of the statistics proposal will not, ultimately, facilitate data fabrication that will undermine efforts to convict those who commit crimes against humanity or genocide.

explicit good intentions, the sociological analysis suggests that there are noteworthy costs of statisticians' professionalization project in the area of human rights.

IV. EPISTEMOLOGICAL IMPLICATIONS OF THE STATISTICS PROPOSAL

Seeing the statistics proposal as part of a professionalization project is to accept, but regard as insufficient, the stated objectives of proposal advocates. Motive, of course, is not the central issue; the key issue is whether the proposal will reinforce or undermine human rights adjudication. Advocates describe the statistics proposal as a way to support human rights, but I submit that its subtle alteration of the definitions of crime against humanity and genocide will reduce support for human rights beyond the effects already described.

A. EPISTEMOLOGICAL IMPLICATIONS OF SHIFTS IN STAFF RECRUITMENT

Proposal implementation is partly reflected in the staffing of the Office of the Prosecutor of the ICC. The sociological analysis suggests that with the change in composition will come changes in the interests and assumptions of ICC staff. These changes have implications for what has to happen to move a hypothesis or belief into the category of widely-accepted fact.

Observe that statistics emphasizes the systematic (though non-formulaic) application of explicit, pre-designated procedures. The emphasis on formal procedure follows from the constraints of the field, typified in that statistical logic can guarantee that a sampling procedure produces unbiased samples *on average*, but cannot guarantee that a *specific* sample is unbiased.[153] In many analytic situations this procedural focus poses no problem.[154]

---

[153] T. M. F. Smith, *The Foundations of Survey Sampling: A Review*, 139 J. ROYAL STAT. SOC. 183, 183–84 (1976).

[154] For example, in academic research on the causes of college entry, even if a specific sample is biased, many other analysts will use different, independent samples. Thus, the effect of bias in one sample can be reduced or even eliminated by analyses of multiple other datasets. Further, none of the academic analyses will be used to admit or reject a specific college applicant, such that persons' opportunities are not directly implicated.

However, under the statistics proposal one uses statistics to identify specific perpetrators. The nature of the question–did the general commit crimes against humanity on specific days in specific villages– likely precludes the statistician from obtaining multiple independent samples for analysis. Conclusions from such samples are justified by the logic of probability sampling, but it will be impossible to verify that the specific sample is unbiased.

The important implication extends beyond sampling. Individuals tend to use similar approaches and logics across multiple domains.[155] Statistics as a discipline highlights formal, systematic processes, *not* the results of those processes. In a dominant logic of the discipline of statistics knowledge is produced by solid, replicable, ideally transferable procedures, such that one has warrant to accept results only when they follow from such procedures. There are alternative logics. For example, another logic might highlight the exhaustiveness of the inquiry, even if producing the exhaustive inquiry requires use of non-standard or even idiosyncratic investigatory means. Given the disciplinary preference for formal, replicable procedures, to the extent statisticians displace persons from other fields in human rights adjudication, fields striking a different relation between procedures and results, one should expect an increasing commitment to procedural instead of substantive justice.[156] It is possible that recruiting more analysts versed in statistics and fewer versed in other fields will increase the push for standardization of human rights processes and spur religious adherence to those processes such that those processes come to represent, and thus replace, the goal of providing substantive justice.[157] Such a dynamic could result in a symbolically powerful but substantively weak human rights process, an outcome consistent with the interests of sovereign immunity.

B. THE ASCENDANCE OF STATISTICS AS EVIDENTIARY SCREEN

Statistics will ascend in importance for human rights adjudication over time if the statistics proposal is fully adopted. At

---

[155] *E.g.*, D. J. French et al., *Decision-Making Style, Driving Style, and Self-Reported involvement in Road Traffic Accidents*, 36 ERGONOMICS 627, 628 (1993).

[156] This is a long-standing distinction in the law. See Tom R. Tyler, *Procedural Justice Research*, 1 SOC. JUST. RES. 41, 42 (1987) for discussion of social psychological factors involved in the relation between procedural and substantive justice.

[157] Lauren B. Edelman, *Legal Ambiguity and Symbolic Structures: Organizational Mediation of Civil Rights Law*, 97 AM. J. SOC. 1531, 1541 (1992).

present analysts have myriad ways to conclude that crimes against humanity or genocide have occurred using working definitions of genocide and crime against humanity. The statistics proposal, if adopted, will rework those working definitions, and this re-working will spur the ascendance of statistics.

We see outlines of the subtle re-working already. For example, as noted earlier, statistics proposal advocates contend that "the phrase 'widespread' refers to the large-scale nature of the attack and the number of targeted persons, while the phrase 'systematic' refers to the organized nature of the acts of violence and the improbability of their random occurrence."[158] However, an organized act–a systematic act–can be implemented such that its outcomes are random. For example, a commander may order soldiers to enter a village and, after detaining every villager, to take each villager one-by-one and, for each, to roll a six-sided die to determine whether or not to kill the villager, killing only those for whom the dice-roll produces a number one. Thus, the murders would be organized, systematic, with a long-term murder rate of approximately 16.7 per 100. Despite their systematicity, the murders would appear random. A statistician perusing counts and characteristics of dead villagers would have to conclude that no groups were targeted, because deaths would appear random in relation to, for example, sex, age, hair color, and more. The commander would be absolved because statistics proposal advocates will have fused two distinct concepts–"systematicity" and "improbably random"–into one. This fusion actually undermines prosecutors' ability to identify human rights violations, because simple responses by actors on the ground can easily frustrate the statistical investigation.[159]

The re-working of the definitions of crime against humanity and genocide is important. Surely, advocates of the statistics proposal claim that in order to draw causal inferences from statistics one needs corroboration from other evidence.[160] Note that in the historic solution to the evidence problem testimony, documents, and physical evidence do

---

[158] Spirer & Seltzer, *supra* note 5, at 200.

[159] If a higher murder rate were desired, multiple sides of the die could require the villager be killed, or a flip of a coin could be used instead, with one side indicating death. The randomization process could be implemented at the village level instead or in addition, rendering the operation of the military unit across villages random. However these randomization processes might be implemented, it should be clear that any villagers killed for that reason were victims of a systematic crime against humanity. Alas, the statistical analysis will not identify them as such.

[160] Spirer & Seltzer, *supra* note 5, at 203.

not require statistical corroboration.[161] However, if causal inferences from statistical analysis require corroboration by other evidence, one can expect other evidence to soon require statistical corroboration as well. If the corroboration requirement applies to statistical analyses alone, then statistical analysis will be unnecessary, for it will have been rendered either an extraneous echo of the existing, traditionally-available evidence, or an errant and ignorable tone amidst the resonance of the remaining evidence. One-way corroboration of statistical results by traditional evidence, thus, seems both unlikely and at variance with the logic of the statistics proposal.

However, once statistical corroboration is required for other sources of evidence the statistical tool will become the determinative one in comparison to victim's testimony, for basic reasons lodged in the structure of the situation. The structuring of roles will produce power for statistics and statisticians and impotence for victims, because patterning is key to the definition of crime against humanity and genocide,[162] and victims tend to be unable to testify concerning patterns, at least directly.[163] Victims can report their victimization and the victimization of others of which they are aware, but they usually cannot survey a large geographic area to determine the outcomes of persons in multiple other places, nor may they ascertain the treatment of persons in other categories. The historical solution has not required such a survey to document crimes against humanity or genocide, yet tell-tale patterns have been discernible.

Once statistical corroboration of eyewitness evidence becomes necessary, such eyewitness evidence will be insufficient to seat a tribunal or indict or convict a perpetrator. Further, because statisticians may be able to discern patterns, they will become the validator of victims' accounts, as witnesses become secondary to statisticians. In such a process witnesses may allege, but statisticians certify.

C. COUNTING THE COUNTABLE, IGNORING THE REMAINDER

Epistemologically speaking, however, statistical analysis may only certify offenses that are feasible to count. Consequently, the

---

[161] May & Wierda, *supra* note 72, at 755–56.

[162] Xabier Agirre Aranburu, *Sexual Violence Beyond Reasonable Doubt: Using Pattern Evidence and Analysis for International Cases*, 35 LAW & SOC. INQUIRY 855, 855 (2010).

[163] See *id.* at 874–75, for a list of eyewitness categories commonly able to report on patterns of atrocities, as in the Tokyo trial; notably absent from the list are victim categories.

methods of genocide or crime against humanity most amenable to counting and thus to statistical analysis will become the determinative aspects of whether a crime against humanity or genocide is deemed to have occurred.

One might expect genocide to be easy to discern statistically–if enough people are missing, then the statistical aspect of the case may appear to be made. Above I indicated that with sufficient lead time the very census counts of groups can be manipulated to lower preexisting estimates of the size of the targeted group, thus lowering the chance that a check of the population count after the atrocities have allegedly occurred will detect that persons are missing. Alternatively, state statistical bureaus can fail to classify persons along lines relevant to the allegation–if population counts of vulnerable persons are unknown, documenting their victimization statistically will require more assumptions and, thus, become more contestable.

Even so, murder is not the only genocidal offense; Part 2, Article 6 of the Rome Statute contains five examples of genocidal actions.[164] They are not equally amenable to quantitative investigation. For example, item (d) concerns "imposing measures intended to prevent births within the group," and item (b) concerns "causing serious bodily or mental harm to members of the group."[165] Compared to measuring deaths or a decline in births, measuring a decline in mental health will be extremely difficult. Mental health indicators may be culturally-specific,[166] and thus measurement of mental health may be difficult.

Even a measured decline in births can be difficult to use as evidence of genocide, because birth declines are common. For example, US total fertility rates were 2.95 for women born between 1936-1940, 2.47 for women born 1941-1945, 2.108 for women born 1946-1950, and 1.92 for women born 1951-1955.[167] Many competing explanations for the decline exist, and such would be the case in many human rights cases. For example, war has been shown to reduce fertility.[168] In fact,

---

[164] Rome Statute of the International Criminal Court art. 6, July 17, 1998, 2187 U.N.T.S. 38544.

[165] *Id.*

[166] Horacio Fabrega, Jr., *Cultural Challenges to the Psychiatric Enterprise*, 36 COMPREHENSIVE PSYCHIATRY 380–82 (1995).

[167] Norman B. Ryder, *Observations on the History of Cohort Fertility in the United States*, 12 POPULATION AND DEV. REV. 617, 622 (1986).

[168] *See* Patrick Festy, *Effets et Répercussions de la Première Guerre Mondiale sur la Fécondité Française* [*Effects and Repercussions of the First World War on French Fertility*], 39 POPULATION (FRENCH EDITION) 977, 1010 (1984) (discussing French fertility decline in World War I); John C. Caldwell, *Social Upheaval and Fertility Decline*, 29 J. FAM. HIST. 382, 383–85

even non-violent social upheaval, such as the Czech transition from communism to capitalism, is associated with major fertility declines.[169] Thus, a statistician using a birth dearth to certify the existence of a genocide would need to eliminate such alternative explanations of the decline. This task can be extremely difficult to accomplish.

Because of such difficulties, the parts of the definition of genocide that analysts can study using approaches that meet the standards needed for appropriate statistical analysis (e.g., probability samples, administration of instruments validated in the population under scrutiny[170]) will come to the fore in establishing the existence of such crimes, and other elements of genocide, despite investigators' best efforts, will likely be unable to sustain such a claim.

A similar case can be made concerning crimes against humanity. It is far easier to count forced or deported migrants than to count victims of torture or disappeared persons, because migrants arrive somewhere, whereas the disappeared may remain buried inside the offending regime unless and until the regime topples and, as stated above, the population size of their demographic group can be problematized prior to the physical phase of the crime, increasing the challenge of identifying groups as specifically targeted.

Relatively difficult to discover aspects of the crime will likely become more common, as actors may concentrate their actions on those aspects less amenable to quantitative analysis, such as harming mental health through, for example, waterboarding, or reducing the number of births, and thus pursue their goal with less chance of discovery. That investigators' focus will change the way actors behave is not an idle concern or mere fantasy; in fact, the pattern of reaction is so common that it has the metaphorical status of a social scientific law, penned over thirty years ago, to wit, "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption

---

(discussing multiple European wars and nations); Magali Barbieri et al., *La Situation Démographique du Viêt Nam* [*Demographic Trends in Vietnam*], 50 POPULATION (FRENCH EDITION) 621, 651 (1995) (discussing fertility decline in Vietnam).

169 Caldwell, *supra* note 168, at 390, 398.

170 For a discussion of the significant difficulties that attend establishing the validity of scales on populations beyond those for which the scale was developed, see Walter J. Lonner, *Issues in Testing and Assessment in Cross-Cultural Counseling*, 13 COUNSELING PSYCHOLOGIST 599, 599–600 (1985); Fritz Drasgow & Charles L. Hulin, *Cross-cultural Measurement*, 21 REVISTA INTERAMERICANA DE PSICOLOGÍA [INTERAMERICAN J. PSYCHOL.] 1, 1–2 (1987). For an example of non-equivalence for two mental health issues in one society see James P. Hambrick et al., *Cross-Ethnic Measurement Equivalence of Measures of Depression, Social Anxiety, and Worry*, 17 ASSESSMENT 155 (2010).

pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor."[171] Accordingly, if the countable is measured and, of course, the uncountable is not, as statisticians' measurements become essential to offering an indictment we can expect motivated actors to engage much more intensively in uncountable acts.

As the countable becomes more important than the uncountable in providing a basis for indictments, subtle changes in the working definitions of crimes against humanity and genocide that truncate the definition of human rights will occur. *De facto* truncation will follow because statistical analysis will drive the working definitions that way, because statistical analysis works more effectively the more explicit the list of the physical objects or conceptual phenomena that constitute the class of interest is. Once one has such a list, data collection can proceed. For example, a statistical analyst interested in discovering whether genocide has occurred would be most able to make that determination if they had in hand an explicit, exhaustive list of crimes that constitute genocide, such as rape, forced abortion, castration, and so on until the full set of possible genocidal acts is enumerated. Given that list, the analyst could then collect data on such events.[172]

In contrast, the relevant statutes are written in an open-ended manner.[173] As one lesson of human behavior is that there may be no end to the creativity that can be brought to bear to torture, maim, otherwise harm, and kill others, the open-ended nature of the statutes is a positive. Any exhaustive list provided on the basis of historic acts may easily fall out of date by the time of the next set of alleged atrocities, but the very existence of a formerly exhaustive list could call forth all sorts of complications among international bodies and their member states were amendments to the list proposed. Writing the statutes in open-ended terms avoids such challenges.[174]

What this means, however, is that the statutes are not written in terms conducive to statistical analysis. If one were simply interested in

---

[171] Donald T. Campbell, *Assessing the Impact of Planned Social Change*, 2 EVALUATION & PROGRAM PLAN. 67, 85 (1979).

[172] Further analysis would be needed, but this data collection would furnish the raw material for subsequent analysis. Thus, absent an exhaustive list, the raw data will be incomplete, likely reducing the chance of finding evidence of genocide.

[173] Jelena Pejic, *The International Criminal Court Statute: An Appraisal of the Rome Package*, 34 INT'L LAW. 65, 74 n.57 (2000).

[174] Article 6c of the Rome Statute is an open-ended definition of genocide, and Articles 7g and 7k of the statute provide open-ended definitions of crime against humanity. Rome Statute of the International Criminal Court art. 6, 7, July 17, 1998, 2187 U.N.T.S. 38544.

counting where possible, and leaving other evidence intact, no problems might follow. But, as statistics proposal advocates seek to refashion the very definition of crime against humanity and genocide,[175] and as counts are the raw data for statistical work, as the statistics proposal is fully implemented it is likely that what cannot be counted will eventually become that which cannot be considered.

In that context statisticians will proceed to construct an operational definition of the phenomenon of crime against humanity or genocide. Operationalization is a process whereby the analyst translates an abstract or theoretical concept (e.g., passionate romantic love) into observable acts or phenomena (e.g., open-mouth kissing) such that the analyst can interpret the latter as signs of the former.[176] This effort to transform concepts so as to facilitate statistical work will necessarily highlight those elements of the definition that are most amenable to counting. Over time, the very elaboration of the concepts of crime against humanity and genocide that the statutes currently facilitate through their open-ended articulation will head toward only those phenomena amenable to quantitative measurement. Yet perpetrators, seeking to hide their crime, will likely elaborate their techniques in ways to *reduce* their susceptibility to counting (and thus to later statistical analysis) to the extent perpetrators are forward-looking and able to do so. These two broad dynamics–statisticians constructing increasingly precise operational definitions that facilitate counting, perpetrators seeking ways to hinder detection–mean that, over time, the total effect of the statistics proposal will be to push the more and more elaborated structure of human rights adjudication toward more and more precise estimates of what occurred, but those more and more precise estimates will increasingly concern less and less relevant aspects of any transgressions that are actually occurring.

D. THE ASCENDANCE OF NUMBER, THE DECLINE OF HORROR

Indeed, we see this already in the claims of statistics proposal advocates. The question of "how many" receives great attention, while qualitative aspects of treatment recede from view. But qualitative aspects matter, and may matter more than the numerical aspects. Put bluntly, we

---

[175] *E.g.*, Bodin, *supra* note 115, at vii. See Spirer & Seltzer, *supra* 5, at 200 for an example of such re-fashioning concerning the definition of systematic.

[176] EARL BABBIE, THE PRACTICE OF SOCIAL RESEARCH, 46–47 (12th ed. 2010).

may ask advocates of the statistics proposal: Is our collective horror at the Nazi holocaust, the Rwandan genocide, the killing fields, or even the Tuskegee Syphilis[177] and Northfield Labs[178] experiments dependent upon the number of victims or, more directly, on victims' experience of these events in all their brutality and disregard? The quantity answer ignores the quality answer, as if the threat crimes against humanity and genocide present to persons' safety, empowerment, bodily integrity, and human dignity depend primarily on how many rather than what.

In short, epistemologically, the statistics proposal reflects and manifests the ascendance of the *number* dimension and the decline of the *horror* dimension in relevance. But, key questions the statisticians cannot answer include: How many people must one dismember alive to intimidate others? How many men, women, or children need be raped to subdue a population? Absent an answer to such questions there is no way to calibrate the quantities that form the focus of statistical work in a way that includes the dimension of horror.[179]

E. STATISTICAL CALIBRATION AS LEGITIMATING HUMAN RIGHTS VIOLATIONS

Of course, statisticians calibrate routinely. But, what counts as calibration is summarized in the term *statistical significance*. Generally, when a finding is statistically significant it means that the researcher has compared two groups and has concluded that the gap between them is probably more than zero, i.e., the two probably differ.[180] Setting aside the

---

[177] *See generally* JAMES H. JONES, BAD BLOOD: THE TUSKEGEE SYPHILIS EXPERIMENT (Free Press 1993)

[178] In the Northfield Labs experiment, which ended in 2007, trauma and accident victims in 27 US cities were given artificial blood without their consent. Adverse reaction rates and mortality rates were statistically significantly higher for those given artificial blood than for those given standard treatment. The FDA-approved experiment is described in Thomas M. Burton, *Red Flags: Amid Alarm Bells, A Blood Substitute Keeps Pumping*, WALL ST. J., Feb. 22, 2006, at A1.

[179] Is use of the term *horror* perceived as too passionate for scholarly legal consideration? Would sanitized labels for the phenomena be more palatable? Alas, the crimes with which we are concerned can involve physically painful, psychologically traumatic, often fatal assaults on persons, often in moments of their extreme vulnerability. To sanitize the reality would seem to take a step toward the very distancing that some agents take to dehumanize their victims and thus facilitate the agent's involvement in such assaults. Integrity, and full awareness of the seriousness of the matter at hand, would seem to demand we avoid sanitizing the brutality necessarily implicated in these issues. Thus, I have referenced some acts that may fall near the high end of the dimension of horror.

[180] THOMAS H. WONNACOTT & RONALD J. WONNACOTT, INTRODUCTORY STATISTICS 291–92 (5th ed. 1990).

calculations, basically the term "statistically significant" means that once we account for the uncertainty occasioned by the use of samples, we still believe we can discern a difference between these two groups. Consequently, we would state that the difference is statistically significant. Key to this calculation is the *standard error*, the indicator of estimate precision. The lower the standard error, the more precise the estimate and, thus, the easier it is to discern differences.

The problem, however, is that the standard error, the statistical measure of the precision of the estimation algorithm, is a function not of the importance of the question but, instead, of the amount of information the statistician is able to bring to bear to answer it, and in statistics this essentially means the standard error is a function of the sample size.[181] Two points should be noted. First, criminal investigators and statistical analysts will often have little to no control over the sample sizes used for the cases they consider. Second, in the best case scenario in which criminal investigators and statisticians are able to control the sample size, they could base the sample size on the importance of the question, drawing larger samples for more important questions. However, note the strictures of a perspective that sees each person as inviolable. In such a view, to take the horror dimension seriously in an investigation of an allegation of a crime against humanity one must set the sample size at a high enough level to be able to discern a difference between one incident and zero incidents.[182] Such a level will be very high and very expensive. However, failing to set the sample size high enough to discern the existence of one incident essentially tolerates some non-zero number of human rights violations. More pointedly, if one does not set the sample size that high, the logic in use will ultimately institutionalize that tolerance at the very center of the effort to bring perpetrators to justice, thus institutionalizing acceptance of one or more human rights violations. The actual number institutionally accepted will depend on the precision of the statistical estimate, and will likely substantially exceed one.

Hence, absent the use of extremely large high quality samples, the statistics proposal obliterates the dimension of horror. In the statistics

---

[181] *See id.* at 197. In addition to sample size, the amount of variation in the phenomenon of interest matters. However, variation in the phenomenon is not under the analysts' control.

[182] One might use the full population for study; for example, one might count the total number of deaths in the population in order to compare death rates for different regions, sexes, races, ethnicities, religions, or other groups of interest. Yet, in the chaotic conditions that often prevail one can expect errors in population counts, such that some means of accounting for the imprecision in the approach would still be needed. That need reintroduces the problem noted here, such that population data does not resolve the problem raised here.

proposal the numeric dimension is paramount. Just as clearly, the horror dimension appears to vanish. This occurs even though even elementary statistics texts note that statistical significance is not substantive significance.[183] Thus, the calibrations routinely available within statistics, useful for many goals, will not help one who does not want to forget the dimension of horror when adjudicating accusations of crimes against humanity or genocide.

F. THE ASCENDANCE OF EFFECTS OVER OCCURRENCES

The final epistemological transformation is that the statistics proposal essentially adopts the following logic–find an effect of an alleged event, and this effect documents that the alleged event occurred. The opposite is then seen to follow–if we can find no effect of an alleged event, then we cannot infer that the event occurred. This logic, however, erroneously equates an effort to find effects with an investigation of the occurrence of events. Adoption of this logic implicitly changes the jurists' task, producing a bias toward indeterminance. And, in a legal system in which defendants are innocent until proven guilty, a bias toward indeterminance is a bias toward not guilty verdicts. This bias toward indeterminance follows from the logic the statistics proposal invokes, a logic that denies that many events occur without measurable effect owing to limitations in our measurement and analytic ability.[184]

For example, imagine that a California resident takes a week-long tropical vacation. Thankfully isolated from world news during the trip, they do not learn that a small earthquake shook their house and neighborhood midway through their vacation. They return home and find everything *just* as they left it (within their senses' sensitivity). Seeing no visible effect of the earthquake their neighbors claim to have experienced, the vacationer concludes there has been no earthquake. Clearly, the neighbors are pulling a fast one.

The vacationer's inability to discern an effect of the quake does not mean the quake was of no consequence. For example, the small quake might have shifted deep underground streams such that, a decade

---

[183] WONNACOTT & WONNACOTT, *supra* note 180, at 291.

[184] Analysts note that all efforts to establish the occurrence of events require effects on our sense apparatus. However, beyond that basic condition, there is a difference between documenting occurrences and documenting the effect of occurrences (on entities other than the observer's sense apparatus). *See* LUCAS, *supra* note 10, at 209.

later, the foundation of their house eventually becomes unsettled, requiring costly repairs. However, tracing their unsettled foundation back to the specific earthquake that occurred during their tropical vacation, or any quake at all, may prove impossible. But note the depth of the problem; because they did not discern an effect, they not only denied an effect of the quake, they denied the very existence of the quake. Denying an effect is defensible, for the vacationer discerned no effect. However, denying the existence of the quake is not defensible, for doing so conflates phenomena and their measurable effects at the same time as it brands all those who experienced the quake liars simply because no effects were visible to a later observer. To deny the quake's existence is to commit the logical fallacy of conflation, for the denial depends on treating phenomena and their measurable effects as one and the same such that if effects are not visible, alleged phenomena are not real. Thus, the statistics proposal also commits the logical fallacy of conflation.

This conflation transforms the task of a jurist, even if one were to have perfect measurement of effects. So, for example, imagine an officer who subjects villagers to painful, torturous treatment. However, the officer employs a method that does not leave a mark.[185] Thus, later examination of the villagers will discern no broken bones or other physical effects of the torture. Under the logic of the statistics proposal an analyst, upon finding no physical effects of the torture (because, in fact, there are no lingering physical effects), will infer the torture never occurred, even though every villager may testify otherwise. Again, because no effects were visible to a later observer, the violation is denied, despite voluminous testimony attesting to the violation. It is just such logic the statistics proposal invokes.

One may document the existence of earthquakes and crimes against humanity without recourse to their damaging effects–lines on a seismograph document the earthquake occurred, and testimony, textual records, and more document human rights violations occurred. However, the statistics proposal threatens all other evidence one might bring to bear in a proceeding concerning crimes against humanity and genocide, while implicitly proposing to institutionalize a logic that conflates the existence and measurable effects of phenomena. Yet, researchers have shown that estimating effects is far more difficult than ascertaining the

---

[185] *E.g.*, Hans D. Petersen & Peter Jacobsen, *Life-Threatening Torture Without Visible Marks*, 13 SCANDINAVIAN J. SOC. MED. 87, 87–88 (1985); DARIUS M. REJALI, TORTURE AND DEMOCRACY 250–51 (2007).

existence of an event.[186] The point follows directly: on this basis alone the statistics proposal constitutes a substantial increase in the difficulty of demonstrating the existence of a crime against humanity or genocide. This is an unacknowledged implication of the statistics proposal.

Taken together, these observations suggest that full implementation of the statistics proposal will make guilt extremely difficult to prove, providing powerful support to *de facto* sovereign immunity.

V. STATISTICAL COMPLEXITIES OF THE STATISTICS PROPOSAL

All these difficulties might be navigable, were statistics truly formulaic. Alas, professionalization is only possible if uncertainty in the domain is high enough that actors must apply non-formulaic judgment in specific cases.[187] Thus, ironically, the very uncertainties in the world that make statistics ripe for professionalization undermine the possibility of statistically objective adjudication of accusations of crimes against humanity or genocide.

Indeed, fundamental disagreements amongst statisticians exist, and they undermine the utility of statistics for the court. There are many fundamental disagreements, but I will consider only two by way of illustration. First, and concerning what appropriate statistics can do, while the statistics proposal purports to offer causal analyses to identify victims and perpetrators, statisticians disagree about whether statistics allows causal inference using the kind of data that human rights investigators will be able to access (i.e., *observational data*, that is, data based on observations of persons' experiences in the world, rather than data drawn from designed experiments). Second, and concerning what constitutes an appropriate statistical analysis, statisticians fundamentally disagree about the very nature of statistics.

---

[186] LUCAS, *supra* note 10, at 206–16.

[187] *E.g.*, Wilensky, *supra* note 123, at 148–49; Mitchell & Kerchner, *supra* note 122, at 215, 236.

A. THE CONTESTED POSSIBILITY OF STATISTICAL CAUSAL INFERENCE

Statisticians debate the minimal conditions that must be satisfied to reach a causal conclusion. One framework, known as Rubin's Model, developed in part in the pages of the *Journal of the American Statistical Association* in the 1980s, has gained increasing acceptance in the social sciences as a way to consider questions of causality.[188] Many statisticians and social scientists who use statistics are now skeptical about whether statistical analysis of observational data can allow causal inference absent heroic assumptions.[189] Legal scholars have also cast a skeptical eye on the way in which statistics is used to establish causation in US courts.[190] To understand the skepticism we need consider the structure of the causal inference problem.[191]

Rubin's Model maintains that there are two instances to which one needs access in order to draw causal inferences–the case in which some unit is exposed to the treatment, and the case where *that same unit* is *not* exposed to the treatment. The difference between the two outcomes is the causal effect of the treatment in comparison to the alternative.

The Frank Capra classic, *It's a Wonderful Life*, nicely illustrates the inferential challenge analysts confront.[192] In the film the protagonist, George Bailey, feels stuck in a small town despite his desire to travel the world. Reaching his wits' end after yet another problem manifests itself, he stands on a bridge on Christmas eve, staring at the river below as a

---

[188] *See, e.g.*, Christopher Winship & Stephen L. Morgan, *The Estimation of Causal Effects from Observational Data*, 25 ANN. REV. SOC. 659, 660, 678–79 (1999); Michael E. Sobel, *Discussion: 'The Scientific Model of Causality'*, 35 SOC. METHODOLOGY 99, 100–01, 105 (2005).

[189] See DAVID A. FREEDMAN, STATISTICAL MODELS: THEORY AND PRACTICE 195–200 (2005) for several examples of such skepticism.

[190] *See, e.g.*, D. James Greiner, *Causal Inference in Civil Rights Litigation*, 122 HARV. L. REV. 533. 539 (2008).

[191] n describing this perspective, I draw heavily on a classic statement of the position by Paul Holland. Paul W. Holland, *Statistics and Causal Inference*, 81 J. AM. STAT. ASS'N 945 (1986).

[192] IT'S A WONDERFUL LIFE (Liberty Films (II) 1946). This inferential puzzle is a movie staple, appearing in films such as the Polish PRZYPADEK (P. P. Film Polski 1981), directed by Krzysztof Kiéslowski, SLIDING DOORS (Paramount Pictures 1998) with Gwyneth Paltrow, and many other film and television productions.

belief that his life has been of no consequence gnaws at his resolve. An angel arrives and conveys to him the effect he has had on many townspeople, and on others far beyond the town, by revealing what would have happened had he never lived. For example, the angel reveals that the lives of the dozens of soldiers his brother saved during the war would have been lost had George Bailey not saved his brother's life years ago. From the perspective of Rubin's Model the angel's comparison of the two worlds, one with and one without the character, identifies the causal effect of George Bailey.

The problem for causal inference in the real world is that we generally have no angel to bail us out of our causal dilemma. Trapped in one world, we observe only one instance and thus remain ignorant of the outcome that would have followed some alternative act. Thus, we cannot compare the two outcomes to determine the causal effect, if any.

In an effort to resolve this problem, analysts invoke assumptions to try to leverage the one world they observe in order to make causal inferences. In other words, they try to compare what they observe (a *factual* observation) with what they believe they would have observed had the same unit been exposed to the alternative condition (a *counter-factual* "observation").

So, for example, a chemist might take a piece of iron from a shelf and pour a solution onto it. If the iron upon which the solution has been poured changes its properties (perhaps becoming weaker), the chemist might hypothesize that the solution poured onto the iron caused the change. There are at least two ways to secure this inference. The chemist might assume *unit invariance* or *temporal invariance*. The unit invariance assumption maintains that all the other pieces of iron on the shelf are substitutable for each other and thus would respond the same way. The temporal invariance assumption maintains that the piece of iron is the same before and after the pouring of the solution, except for the pouring of the solution. Thus, comparing bars with and without exposure to the solution, or the same bar before and after exposure to the solution, estimates the causal effect of the solution on iron.[193]

---

[193] We see again that it is far easier to discern that an event occurred (the pouring of the solution) than to discern a *causal effect* of the event. Showing that the bar is weaker after the pouring of the solution does not, by itself, establish the poured solution as the cause of the weakening. Before we even consider the implications of Rubin's model, alternative causal explanations (e.g., something on the gloves the chemist wore while moving the iron bar caused the bar to weaken; something in the air around the table upon which the experimental iron bar was placed caused the bar to weaken) are easy to propose and can be difficult or even impossible to eliminate. Acceptance of Rubin's model only intensifies the challenge of identifying causal effects.

In social science research the closest approximation to this chemical investigation is an experiment in which analysts randomly assign persons to different groups and then administer a treatment to one group and not the other. Random assignment justifies assuming the groups are substitutable on average before treatment administration, allowing one to compare the groups' outcomes after treatment to estimate the causal effect of treatment.

For this approach to work analysts must randomly assign persons to treatment and control groups. Clearly, observational studies of human rights violations do not satisfy this condition. When this condition is violated causal inference becomes incredibly challenging, for in order to sustain the inference one must identify groups to compare and one must defend the implicit assumption that the groups are substitutable. If one cannot defend the substitutability assumption successfully, causal inference cannot succeed.[194]

In the case of whether a soldier caused the death of civilians of an allegedly targeted ethnic group, Rubin's model implies that if one lacks a counterfactual observation of what would have happened had the civilians never encountered the soldier, one is unable to provide any evidence on the matter of the soldier's guilt or innocence. One might attempt to estimate the counterfactual observation by leveraging other regions through which the soldier's unit did not pass, to compare the pattern of deaths in places the unit did and did not traverse or occupy. However, in order to make this approach work one must invoke a unit invariance assumption, claiming that regions the soldier occupied are the same as regions the soldier did not occupy, people the soldier encountered are the same as people the soldier did not encounter, people the soldier encountered behaved the same as people the soldier did not encounter would have behaved had they encountered the soldier, and so on *ad nauseum*. Yet, in the conditions that commonly will concern investigators attempting to ascertain whether human rights violations were committed, unit invariance is no easy assumption to sustain. For one, the heterogeneity of social, strategic, demographic, and geographic terrain provides the reason troops are deployed in different ways and may use different routes in their travels (e.g., air, sea, various land

[194] George Maldonado & Sander Greenland, *Estimating Causal Effects*, 31 INT'L J. EPIDEMIOLOGY 422, 427–28 (2002).

pathways). In short, troops are not deployed randomly. This suggests unit invariance is an incorrect or, at least, easily contestable assumption.[195]

Alternatively, one might invoke a temporal invariance assumption: life would have been undisturbed, but the soldier's unit arrived. Again, such a contention can be difficult to sustain as wartime routinely involves upheaval for those inside and outside the theatre of operations.[196]

In sum, for the statistical logic of causal inference to estimate the causal effect the analysis requires substitutability. However, even in normal circumstances humans (and human groups) are not necessarily as substitutable as are iron bars. Consequently, one must justify any substitutability assumption. Yet, substitutability is so difficult to justify in general, and so much more difficult in the conditions likely to concern human rights analysts, that Rubin's model might imply the impossibility of drawing causal inferences concerning genocide and crimes against humanity through statistical analysis of observational data.

Of course, the aim here is not to argue that the statistician's framework for causal inference is applicable beyond statistics–other fields (e.g., the law) have other frameworks that serve their aims. However, a significant proportion of statisticians *are* persuaded by the argument that statistics cannot allow one to reach a causal inference from observational data unless one invokes strong assumptions.[197]

Thus, despite the advocacy of some statisticians who favor the statistics proposal, the implication for using statistics to establish causality in cases concerning a charge of crimes against humanity or genocide is dire. Any investigation justified by a statistical inference of causation will be vulnerable to losing any legitimacy it might have had once other analysts express doubt concerning the possibility of reaching a causal conclusion as to the existence of actionable offenses on the basis of statistical analysis of observational data. Any prosecutor who indicts a defendant for genocide or crimes against humanity on the basis of a statistical inference of causal responsibility is simply begging that defendant to enlist the aid of any one of the many other statistical analysts to testify on their behalf and, in that expert witness' eyes, against the inappropriate use of statistics for questions they, as statisticians, doubt statistics can address.

---

[195] I have framed the matter in terms of characteristics one might observe. Causal inferences become even more difficult if we posit characteristics one cannot observe.

[196] *E.g.*, Krain, *supra* note 69.

[197] *See* Freedman, *supra* note 189, at 187.

Unleashing a statistician versus statistician dynamic threatens the entire effort to adjudicate allegations of crimes against humanity and genocide. The threat is deepened as we consider an even more fundamental divide amongst statisticians, a divide that raises–but fails to answer–the fundamental question of what actually constitutes an appropriate statistical analysis.

B. PARADIGMATIC CONTESTATION

Statisticians fall into at least two large, distinct groups: 1) frequentists and, 2) bayesians. These camps fundamentally disagree about what counts as evidence, what formulas signify, and what an analysis should entail. Frequentist analysis dominate many scientific literatures;[198] indeed, the frequentist perspective is so presently dominant that the statistics proposal often appears unreflectively articulated in frequentist terms.[199] We can discern the serious implications of the frequentist-bayesian disagreement by comparing how a frequentist and a bayesian would investigate the same question.

Assume the best case scenario with respect to data collection–researchers will be able to study all entities sampled, there will be no refusals to take part in the study, the sampling frame will accurately identify truly eligible entities, and only such entities, and other design issues will be unproblematic. As shown above, such advantageous conditions, always unlikely, will be even rarer after adoption of the statistics proposal, because forward-looking actors may undermine those conditions before and during the commission of the crime. Yet, to isolate the issues dividing frequentists and bayesians, issues that will interact with and further complicate the use of statistics in such cases, here those difficulties will be assumed away.

A frequentist who wanted to use statistics to determine whether women of one group were forcibly sterilized by persons of another group hostile to the first group–a genocidal act–would begin by determining how large a sample of entities to draw. To make this determination the frequentist would identify a degree of certainty they would like to reach, i.e., the frequentist might assert that they want to be able to discern a

---

[198] HUGH G. GAUCH, JR., SCIENTIFIC METHOD IN PRACTICE 364–65 (2003); Donald A. Berry, *Teaching Elementary Bayesian Statistics with Real Applications in Science*, 51 AM. STATISTICIAN 241 (1997).

[199] But see Spirer & Seltzer, *supra* note 5, at 203, for one sentence more consistent with a bayesian than a frequentist perspective.

difference of size *Y* between the jurisdiction of interest and some other, comparable place.[200] The idea is that if the analysis results in an estimated difference larger than the minimal discernible difference, *Y*, then the finding will be regarded as consistent with differential treatment in the two different jurisdictions. However, any difference smaller than that minimal discernible difference will lead analysts to infer there is insufficient evidence of differential treatment. The size of *Y* translates into a sample size, N, large enough to detect a difference at least as large as *Y* (given other assumptions the analyst will make about variation in the two locales).[201] Thus, the analyst will draw a sample of size N, which will imply a size for the standard error, our afore-mentioned measure of precision.

For our substantive case one might conduct medical examinations of a sample of women of child-bearing age from two jurisdictions to ascertain the percentage of women in each locale who are sterile.[202] The frequentist will calculate the difference in those percentages and then assess whether the percentage point difference is larger than Y; if so, it is likely that the contexts actually do differ. To make the assessment the frequentist might estimate upper and lower bounds on the difference between the two contexts; if zero falls within this (usually 95 percent confidence) interval it means that the difference between the two jurisdictions could easily be zero. If so, then the frequentist cannot conclude that the contexts are different.[203] If zero does not fall in the interval and the suspect jurisdiction has higher sterility incidence, the frequentist will infer there is likely a difference, and this is consistent with the occurrence of genocidal sterilization.

Bayesians proceed differently. The first act a bayesian will take to address the question will be to guess how many more or fewer sterilizations occurred in the suspect jurisdiction. The guess constitutes a *prior*, and in fact will be a set of guesses reflecting the probability of different answers. To construct this prior, the bayesian uses their own tastes and proclivities plus whatever information–prior information–to

---

[200] We have already noted that setting Y>(1 victim) obliterates the horror dimension from concern. If Y is a proportion one may simply rescale the value, such that an analyst that sets Y>(1/Population) can tolerate the infliction of a crime against humanity on at least one person.

[201] Russell V. Lenth, *Some Practical Guidelines for Effective Sample Size Determination*, 55 AM. STATISTICIAN 187, 187 (2001).

[202] One could conduct the analysis using comparative data for the same jurisdiction from prior to the period of alleged atrocities, if such data exists and other assumptions are satisfied.

[203] Martin J. Gardner & Douglas G Altman, *Statistics in Medicine: Confidence Intervals Rather Than P Values: Estimation Rather Than Hypothesis Testing*, 292 BRIT. MED. J. 746, 747 (1986).

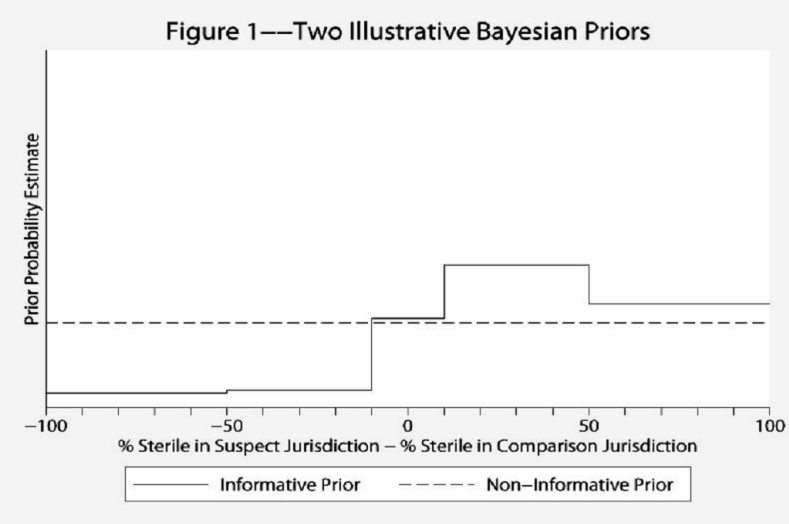
which they have access. Prior information could come from anywhere–knowledge about similar conflicts in the past between different peoples, knowledge of the history of the two groups, reports of other atrocities occurring at the same time involving or not involving these groups, information about available technological resources that might have facilitated certain actions, information about the claimed aims of various political actors in the environment, or other sources. The bayesian analyst would summarize their belief by first identifying every single difference between the two contexts (e.g., 10 percentage points, 11 percentage points) that the analyst believes is possible based on existing information.[204] Then, for every single difference the analyst believes is possible the analyst would assign a subjective probability to each difference, a probability stating the chance that *each* difference is *the* difference between the two contexts.[205]

The full set of possibilities and their respective probabilities form a distribution of probabilities of possible differences between the two jurisdictions, termed a *prior* distribution in this context; this prior distribution summarizes the analyst's subjective belief of the possibilities based on existing information before conduct of the statistical analysis. As these are subjective beliefs, different analysts may have different prior distributions (i.e., priors). Figure 1 sketches the priors two different bayesians might have. The horizontal axis ranges from negative one hundred to one hundred. Negative values indicate the suspect context has less sterilization, whereas positive values indicate the suspect context has more sterilization. The higher on the vertical axis a line rises above a value on the horizontal axis, the greater the subjective probability that *that* value on the horizontal axis is the difference between the suspect context and the comparison context.

---

[204] *See* WILLIAM M. BOLSTAD, INTRODUCTION TO BAYESIAN STATISTICS 6 (2nd ed. 2007).

[205] *Id.*



As Figure 1 shows, one analyst doubts that the difference will be more negative than -50 percentage points. This analyst believes that the difference has a better chance of falling between -50 and -10 percentage points; this belief is reflected in the slightly raised probability for this set of outcomes in comparison to the estimate for the values falling between -100 and -50. While the analyst also believes an estimate between -10 and +10 is more possible than the more negative values, this analyst's prior belief is much more weighted toward expecting a much higher proportion of sterilization in the suspect context, as reflected in the much higher probability for those values. Perhaps this analyst knows of historical treatment in the context of focus which leads her to suspect that, once again, some atrocity has occurred.

In contrast, another bayesian analyst believes that each difference is equally likely. Thus, this bayesian proposes a *non-informative prior*, in which no result is any more expected than any

other.[206] This belief is reflected in the horizontal dashed line in Figure 1, a line which shows the analyst does not see any difference between contexts is any more or less likely than any other.

Either bayesian analyst could then collect data just as the frequentist did or, alternatively, use the exact same data the frequentist collected. Either way, the bayesian would combine their prior distribution with estimates obtained from the collected data to develop an updated, *posterior* distribution of estimates of the difference between the contexts. That updated distribution, reflecting an appropriate alteration in the analyst's prior beliefs based on additional information, would be the bayesian conclusion.

To reveal the implications of the paradigmatic difference, assume a frequentist and a bayesian use the same data, and the frequentist finds that the difference between the suspect context and the comparison context is 20, with a standard error of 13. This implies a (95 percent) confidence interval ranging from approximately -5.5 to +45.5, such that the frequentist estimates that the suspect context has anywhere from 5.5 percentage points *fewer* sterile women to 45.5 percentage points more sterile women than the comparison context. Because the interval includes zero, the frequentist cannot reject the supposition that the two contexts have equal sterilization incidences. Thus, the frequentist would conclude that there is insufficient statistical support for the claim that genocidal sterilization occurred.

Parenthetically, note that this standard error implies that if the estimated percentage difference between the jurisdictions is less than approximately 25.48 percentage points, the frequentist will be unable to infer the occurrence of genocidal sterilization. This means that for every 1,000 women of childbearing age in the suspect jurisdiction, this application of the statistical criterion accepts the genocidal sterilization of up to 254 of them without raising an alarm and without validating the alarm of others. This is what I meant when I noted that adoption of the statistics proposal institutionalizes acceptance of one or more crimes against humanity at the center of the human rights enforcement process.

Figure 2 traces the bayesian update of the informative prior of Figure 1. This updated conclusion is called a bayesian posterior distribution.[207] I have overlaid the informative prior in a solid line so that

---

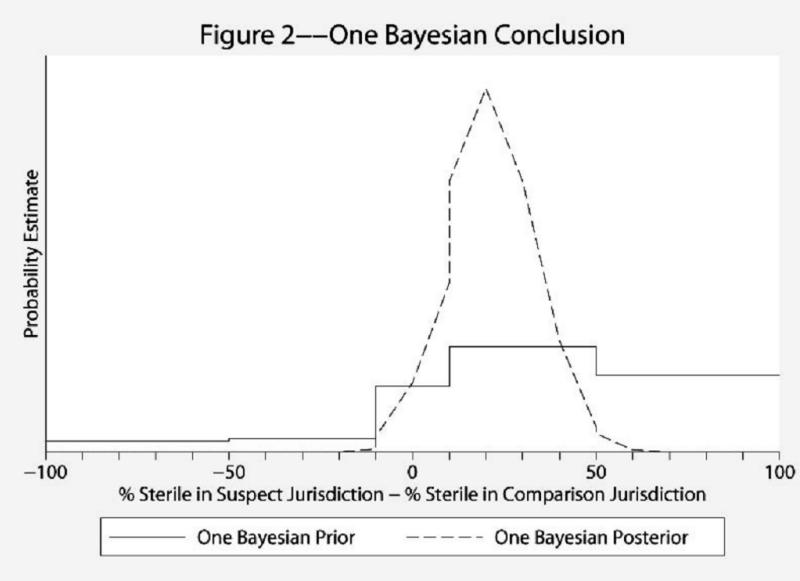
[206] SCOTT MICHAEL LYNCH, INTRODUCTION TO APPLIED BAYESIAN STATISTICS AND ESTIMATION FOR SOCIAL SCIENTISTS 55 (2007).

[207] *Id*. at 50.

one can see how prior expectations are altered by consideration of the data. Combining data and prior greatly reduces some probability estimates–it is not perceptible in the figures, but none of the possibilities have zero probability, although much of the range has posterior probability close to zero. However, belief that estimates fall between 10 and 50 percentage points is now very large. In fact, the bayesian finds an approximately 89.6 percent chance that sterilization incidence in the suspect context is 10 to 50 percentage points greater than in the comparison context, and a 90.6 percent chance that sterilization is at least 10 percentage points greater in the suspect context. A very small chance exists that the difference is negative. Thus, while the frequentist concludes there is insufficient evidence of genocidal sterilization, the bayesian finds an overwhelming preponderance of the evidence and what some might regard as evidence beyond a reasonable doubt of more sterilization in the suspect locale.[208] Thus, the bayesian would find the evidence supports the likelihood of genocidal sterilization.[209]

[208] *See* Rita James Simon, *"Beyond a Reasonable Doubt"–An Experimental Attempt at Quantification*, 6 J. APPLIED BEHAVIORAL SCIENCE 203 (1970) (demonstrating how laypersons may quantitatively interpret the "reasonable doubt" standard).

[209] No conclusions should be drawn concerning the direction of the illustrative results. Here the bayesian analysis suggests genocide probably occurred, but in other cases the pattern could reverse, with the bayesian analysis suggesting that genocidal sterilization did not occur.



This illustration indicates that bayesians and frequentists do not simply use different methods to reach the same conclusion. Sometimes they reach the same conclusion, but often they do not. Thus, the debate between these two paradigms is not merely an academic curiosity. In fact, one's chosen statistical paradigm will determine, in part, what one will decide in any given instance. This possibility undercuts the claim of objectivity for every statistical paradigm.

Bayesians claim that, at best, frequentists indefensibly treat each case as an entirely new circumstance and ignore related information and that, at worst, frequentists inadvertently and haphazardly use prior information.[210] In response, bayesians systematically incorporate prior information into their analysis. Yet, this systematic introduction of prior beliefs means that, because individual analysts have different perspectives, tastes, priorities, and experiences, even if different

[210] E.g., E. T. Jaynes, PROBABILITY THEORY: THE LOGIC OF SCIENCE 280 (2003).

bayesians had the same exact prior information they could construct different prior distributions, and thus, even using the exact same data they could reach different conclusions. Hence, according to frequentists bayesians simply produce contaminated, i.e., biased analyses.[211] For frequentists, a technique that leads two analysts with the same information to reach different conclusions is a textbook case of a subjective situation. Frequentists recoil from this possibility.

In sum, frequentists attempt to purge subjectivity, while bayesians attempt to make subjectivity transparent.[212] The claim here is not that bayesian statistics are better or worse than frequentist statistics. The claim here is that *statisticians* disagree on whether bayesian statistics are better or worse than frequentist statistics.[213] This presents a very different paradigmatic terrain than that upon which bio-scientists stand in virtual unanimity concerning DNA testing and evolutionary theory or chemists occupy concerning the nature of the elements. Without elaborating additional differences between the bayesian and frequentist paradigms, we can note that the existence of this fundamental difference–is prior information essential to systematic analysis or, instead, a contaminant of systematic analysis?–implies that the statistics proposal is under-specified, for even after promulgation of the statistics proposal the question remains: *which* statistics?

One might regard the bayesian/frequentist dispute as easily reconciled. While bayesians believe it is indefensible to ignore existing information, they also claim that if there is little existing information one may specify a non-informative prior that is robust to estimation efforts and gives maximum weight to the data. In the limit, then, one might, at least in this regard, see bayesian statistics with a non-informative prior as

---

[211] Robert L. Winkler, *Why Bayesian Analysis Hasn't Caught on in Healthcare Decision Making*, 17 INT'L J. TECH. ASSESSMENT HEALTH CARE 56, 60 (2001).

[212] Matters are complex, so this is a general statement. For example, some bayesians accept the frequentist criticism and seek to respond with a form of bayesian objectivity. *E.g.* James Berger, *The Case for Objective Bayesian Analysis*, 1 BAYESIAN ANALYSIS 385 (2006). Further, large datasets can swamp priors, although, as mentioned earlier, sample size is unlikely to be within the control of litigators. *See, e.g.*, Frank Tuyl et al., *A Comparison of Bayes-Laplace, Jeffreys, and Other Priors: The Case of Zero Events*, 62 AM. STATISTICAN 40 (2008). Finally, bayesians could point out that the selection of the critical cut-off of 95% in frequentist reasoning, while inter-subjectively maintained, remains arbitrary and, to the extent it is arbitrary it is not really objective. Even with these qualifications, the example illustrates one of the fundamental and most potentially consequential differences between frequentists and bayesians.

[213] For example, Francisco J. Samaniego & Dana M. Reneau, *Toward a Reconciliation of the Bayesian and Frequentist Approaches to Point Estimation*, 89 J. AM. STAT. ASS'N 947 (1994), implies there is a disagreement to resolve or reconcile.

converging to frequentist statistics. However, three observations need be made.

First, non-informative priors are not panaceas. In the example above, do we really believe it possible that 100 percent of the women in the comparison jurisdiction are sterile? If not, then we cannot believe the difference in the percentages across the two jurisdictions has any chance to be -100. However, if we reflect this belief in our prior, clearly a defensible (and to the bayesian, necessary) response, we make our prior an informative one. Accordingly, non-informative priors are not necessarily defensible priors.[214]

Second, use of explicit priors is only one of the easier-to-convey differences between these paradigms. Other differences concern such fundamental issues as the appropriate basis for statistical tests, [215] the meaning of confidence intervals,[216] and how one explicitly specifies the costs of making an erroneous inference in decision analyses.[217] Were we to elaborate such differences the incommensurability of bayesian and frequentist approaches would deepen.

Third, a bayesian who prefers a non-informative prior could confront a bayesian who prefers an informative prior. The proponent of an informative prior may see the issue as a largely mathematical requirement, as above, or, instead, as a sociologically necessary response to a history in which one group has long oppressed the other. Faced with such a claim, how could the court determine which approach is statistically appropriate, when the experts fundamentally disagree? Thus, by extension, if a frequentist with no explicit prior confronts a bayesian with an explicit informative prior, how can the court evaluate which approach is best? If the analyses reach the same conclusion, then all is

---

[214] The ratio of the estimates of the incidence of sterilization, or some function of that ratio, would be a better index for comparing the two jurisdictions. I used the difference in percentages for the illustration to ease interpretation and to keep the focus on the epistemological differences between the two paradigms, and because the general point holds even if other indices are utilized. Still, see Yu Xie, *Measuring Regional Variation in Sex Preference in China: A Cautionary Note*, 18 SOC. SCI. RES. 291, 303 (1989), for a cogent discussion of the problems with a difference-in-percentages measure as well as possible alternative indices.

[215] Eric-Jan Wagenmakers et al., *Bayesian Versus Frequentist Inference*, *in* BAYESIAN EVALUATION OF INFORMATIVE HYPOTHESES 181 (Herbert Hoijtink et al. eds., 2008).

[216] Bayesians often regard confidence intervals as probability intervals. See ANDREW GELMAN ET AL., BAYESIAN DATA ANALYSIS 4 (2d ed. 2004), for a brief discussion of the difference between Bayesian and frequentist interpretation of such intervals.

[217] Decision analysts attempt to explicitly account for the costs of being incorrect. This is very different from how statistics is usually used in academic studies. See JAMES O. BERGER, STATISTICAL DECISION THEORY AND BAYESIAN ANALYSIS (2d ed., 1985), for differences between Bayesian and frequentist approaches to specifying the cost of erroneous inference.

well. However, often analyses before the court will conflict, with exceedingly visible implications of selecting one or the other paradigm for victims' emotional repair, the chance of obtaining other victims' testimony, as well as defendants' freedom or imprisonment all hanging in the balance.

The difference between the posture of statistics and the secure footing of DNA evidence and much other natural science evidence is striking. It is common for experts to battle in court, and that commonality makes it easy to miss qualitative differences between the kind of battle routinely observed in court amongst forensic pathologists, for example, and the kind of battle statisticians may wage in court.

Most expert versus expert court battles are intra-paradigmatic, such that analysts agree on the foundational aspects of their field, yet disagree on some details or interpretive aspects concerning the specific case before them. Because the fundamental underpinnings are accepted, it is conceivable that one expert could persuade the other, because they agree on the rules that matter (although they may disagree about their application). Faced with bounded disagreement, courts defer to the experts on disciplinary matters (the expert's major premise).[218] Thus, courts' use of findings on, for example, DNA evidence, does not depend on their adjudicating between evolutionary theory and some other theory of the fundamental basis of life.

In contrast, statistician versus statistician conflict in a case can easily involve inter-paradigmatic disputes. Bayesian statisticians do not accept frequentist approaches as appropriate, and frequentist statisticians do not accept bayesian approaches as appropriate. Both are conducting statistical analyses, but their *foundations* contradict each other. Thus, in every judicial decision involving the statistical evidence the ICC will be forced to take sides in the dispute between paradigms.

This is treacherous terrain for a nascent court. Requiring the court to resolve inter-paradigmatic scholarly disputes to conduct its focal business is to endanger the legitimacy of the court, for, in that context, its rulings will be at great risk of being seen as reflecting prejudice for or against particular parties and groups rather than reflecting the weight of

---

[218] *See* Edward J. Imwinkelried, *The "Bases" of Expert Testimony: The Syllogistic Structure of Scientific Testimony*, 67 N. C. L. REV. 1, 3 (1988) (distinguishing major and minor premises of expert witness testimony). For the ensuing debate concerning appropriate and inappropriate judicial deference to experts, see Ronald J. Allen & Joseph S. Miller, *The Common Law Theory of Experts: Deference or Education?*, 87 NW U. L. REV. 1131 (1993), and Edward J. Imwinkelried, *The Educational Significance of the Syllogistic Structure of Expert Testimony*, 87 NW U. L. REV. 1148 (1993).

the evidence only. Reconciliation of the statistical paradigms seems the only way courts could be saved from this particular threat. Yet, reconciliation of the contradictory paradigms does not seem imminent. Absent reconciliation, the statistics proposal is in essence asking jurists to rely heavily on a discipline that is insufficiently mature for the task.

C. CAUSAL UNCERTAINTY AND PARADIGMATIC DISSENSION: IMPLICATIONS FOR THE STATISTICS PROPOSAL

These disagreements have important implications. First, although the different views of whether statistics can establish causality are, from one point of view, just another manifestation of the routine differences of opinion that professionals express, even this conclusion suggests that prosecutors will not long be the only actors bringing expert statistical analyses to bear in cases concerning alleged crimes against humanity or genocide. Thus, the proposal offers great advantages to defendants, especially to guilty defendants who can use statistical analyses to cast doubt on the prosecution's causal claims. This is another pathway through which the statistical bias toward indeterminance is produced.

In addition, the implications of paradigmatic statistician disagreement are large. Inter-paradigmatic disputes are irresolvable on the basis of evidence because a fundamental incommensurability holds.[219] The statistics proposal is placed before us even though statisticians are currently in the midst of a dispute whose resolution may not be possible on the basis of evidence because the dispute concerns fundamentally different paradigms. As Kuhn maintains, the dispute may only be resolved once advocates of the dominant position are replaced in the field through the natural demographic processes of death and replenishment.[220] Yet, a demographic resolution may be unlikely in this case, as the bayesian/frequentist dispute has lasted for over two and a half centuries–bayesians dominated the nineteenth century, frequentists dominated the twentieth century, and bayesians are presently making a resurgence.[221] If a thoroughgoing resolution is not soon provided, the statistics proposal will eventually bring the academic dispute into the international courts.

---

[219] THOMAS S. KUHN, THE STRUCTURE OF SCIENTIFIC REVOLUTIONS (3d ed. 1970).

[220] *Id.* at 150–51.

[221] B. Efron, *Why Isn't Everyone a Bayesian?*, 40 AM. STATISTICIAN 1 (1986); Bradley Efron, *Bayesians, Frequentists, and Scientists*, 100 J. AM. STAT. ASS'N 1 (2005).

Thus, despite Spirer and Seltzer's claim that statistical assessment of human rights allegations can be as useful as autopsies and DNA evidence, [222] statistical evidence is decidedly less useful. The evidentiary value of chemical testing is based on the solid consensus reflected in the periodic table; the evidentiary value of forensic work rests on physiological study of how the human body responds to assaults and decomposes under various conditions, a field of study reliant on chemistry. The evidentiary value of DNA testing flows from its basis in Mendelian principles of genetics and, thus, ultimately from the theory of evolution. Although the principles of chemistry, DNA study, and forensic science are not etched in stone,[223] they are considerably more solid than statistics, because the content of statistical work is infinitely-arrangeable information, and statisticians do not have an equivalent to the periodic table nor anything approaching the rigor or consensus of the theory of evolution to aid their analysis of that information.[224] Thus, statistics cannot provide to jurists anything approaching the value of autopsies, chemical analysis, or DNA testing.[225]

Experts often disagree. But, once fields riven by fundamentally different paradigms are placed in a position to certify the veracity of all other evidence, then the last crucial step will have been taken, such that, ultimately, no accusation may ever survive scrutiny because the alternative paradigm is always available to impugn the logic and thus the credibility of the analysis that established the point with which one

---

[222] Spirer & Seltzer, *supra* note 5, at 195.

[223] *E.g.*, Hinne Hettema & Theo A. F. Kuipers, *The Periodic Table–Its Formalization, Status, and Relation to Atomic Theory*, 28 ERKENNTNIS 387 (1988); Angmary Brito et al., *A reconstruction of development of the periodic table based on history and philosophy of science and its implications for general chemistry textbooks*, 42 J. RES. SCI. TEACHING 84, 92 (2005).

[224] Even so, the strength of DNA evidence is often over-stated in trials. *See* Jonathan J. Koehler, *Error and Exaggeration in the Presentation of DNA Evidence at Trial*, 34 JURIMETRICS 21, 22 (1993). Note that DNA evidence draws upon and is evaluated using a statistical model. However, the primary inputs to that model are well-understood physical parameters drawn from evolutionary theory and massive knowledge of the human genome. Yet, even here we see the way in which statistics is not objective; for example, different assumptions of the cleanliness of the lab (or the degradation of the physical sample from a crime scene) can translate into different estimates of error rates in the DNA test, and these differences can lead different analysts to draw different conclusions on the basis of the same physical DNA test as to the likelihood a particular person is singly identified by the test. Thus, to the extent DNA evidence depends on statistics, even DNA evidence is less informative than it is often idealized to be.

[225] Statistics is not the only field that might make such demands on the court. For example, psychiatry and sociology are fields with multiple conflicting paradigms. However, statistics proposal advocates did not compare the value of statistics to the value of psychological or sociological examination; instead, they explicitly claimed that statistics would prove as illuminating as DNA and forensic science. Evidence suggests this claim is incorrect.

disagrees. This paradigmatic dispute provides yet another way by which the statistical bias toward indeterminacy is produced.

Taking these examples together, it may be the case that after adoption of the statistics proposal the evidence might *never* be strong enough to extract reparations for victims, consign regimes to eternal opprobrium, and send perpetrators to their punishment, even as the field collapses into irresolvable conflict in the process. If this future unfolds, two key aims of many professionalization projects–the increasing stature and employment of those deemed professionals–will have been secured.[226] The more conflict that surrounds the statistical analyses allegedly needed to certify and adjudicate charges of crimes against humanity or genocide, the more necessary statisticians are for deciphering the meaning of the analyses and conveying their implications to others. In that context, prosecutorial failure to consult a statistician will be deemed malpractice, even as indicting suspects on such evidence will necessarily weaken the case against defendants by legitimating an alternative basis of defense.

Of course, one might argue that a field that collapses into irresolvable conflict will never gain the credibility in the courts needed to eventually undermine all other evidence. This supposition would be true, if the collapse into conflict were immediate, widely visible, and accurately interpreted as irresolvable. However, at present prosecutors are the only one's presenting statistical analyses, and they are obtaining convictions in those trials.[227] The one-sided use of statistical evidence erroneously suggests that statistics lacks conflict and is objective and thus disadvantages guilty defendants. Yet, defendants have yet to exploit the many advantages the statistics proposal offers to their position, guilty or not. Thus, the conflict has yet to be fully engaged, slowing apprehension of statistics' disposition toward indeterminance and the accompanying bias of statistics toward acquittal in a legal system in which persons are presumed innocent until proven guilty.

At the same time, the long-running professionalization project advocates training jurists and other personnel involved in human rights adjudication in statistics and statistical logic.[228] Once that occurs on a large enough scale one may be forced to articulate accusations concerning crimes against humanity and genocide in statistical terms in

---

[226] Klegon, *supra* note 121.

[227] *E.g.*, Prosecutor v. Momćilo Krajišnik, Case No. IT-00-39-A, Judgment (Mar. 17, 2009), http://www.icty.org/x/cases/krajisnik/acjug/en/090317.pdf.

[228] Spirer & Seltzer, *supra* note 5, at 215.

order to communicate to the institutionalized actors in terms they understand and value. It will not matter that the content one can unproblematically communicate has been severely narrowed; it will not matter that solid grounds for reliance on statistical analysis were never present and thus the proposal constituted over-reaching from the beginning. What will matter is that jurists will be caught between two incommensurable paradigms *after* the historic solution has had its legitimacy drained away by the steady assault on its sufficiency that the statistics proposal constitutes. At best all that will remain will be "victor's justice," a form of human rights bearing the label of the ideals, but in practice mocking the possibility of universal recognition.[229] Some might view this as the current reality,[230] but, even if so, adoption of the statistics proposal will only intensify and solidify that reality. If these developments unfold as I suggest they may, adoption of the statistics proposal will have eventuated in a Final Solution to the sovereign's Human Rights Question.

VI. HISTORICAL EXPERIENCE OF THE STATISTICS PROPOSAL: DISCRIMINATION ADJUDICATION IN THE UNITED STATES AS A DEMONSTRATION

One might maintain that matters cannot unfold as I suggest, because statistics would not be allowed to so deeply impact a field of legal inquiry that it threatened the legitimacy of every decision and possibly rendered convictions virtually unobtainable. In that connection, discrimination adjudication in the United States provides an intriguing test-case of the value of statistical analysis for human rights adjudication.

Both discrimination on the one hand, and crimes against humanity and genocide on the other, concern the oppression of vulnerable populations in a wholesale manner, an oppression based in social relations that transcend the individuals who are party to any specific act;[231] as such, the patterns of treatment can be central to establishing the existence of the crime. Both crimes can be (and have

---

[229] *See* AXEL HONNETH, THE STRUGGLE FOR RECOGNITION: THE MORAL GRAMMAR OF SOCIAL CONFLICTS 119–20 (Joel Anderson trans., 1996); Nancy Fraser, *Recognition Without Ethics?*, THEORY, CULTURE & SOC'Y, Apr.–June 2001, at 24, 24–28.

[230] See Victor Peskin, *Beyond Victor's Justice? The Challenge of Prosecuting the Winners at the International Criminal Tribunals for the Former Yugoslavia and Rwanda*, 4 J. HUM. RTS. 213 (2005), for a cogent study of the victor's justice problem.

[231] *See* Lucas, *supra* note 10, at 175−91.

been) legalized by the duly constituted authority in a jurisdiction such that critics of the oppressive circumstances often must look beyond local law for grounding to contest the arrangements.[232] Indeed, both crimes have been conceived as lying on a common continuum that includes and is undergirded by violence.[233] Further, both crimes are such that powerful actors might prefer to avoid serious enforcement efforts.[234]

These observations do not imply that the crimes are exactly the same, nor that the environment in which statistics approached each court is the same, either. Notably, statistics entered US discrimination litigation nearly two centuries after the establishment of the Supreme Court; in contrast, statistics is being brought to the ICC from its outset.[235] These different histories pose decidedly different implications for each Courts' ability to secure and maintain their legitimacy in the face of the danger statistics can pose to that legitimacy. Moreover, it was mid-twentieth century statistics that was brought to US courts to aid assessment of discrimination. Statisticians and social analysts have learned a great deal about the complexity of inference since that time, and more knowledge has produced more doubt. Thus, despite the claims of statistics proposal advocates, many twenty-first century statisticians are far less confident of the utility of statistics for aiding adjudication of such cases than were their mid-twentieth century peers.[236]

These differences in the environments are such that one cannot unreflectively extrapolate from discrimination litigation in the United States to human rights adjudication at the ICC. Even so, the similarities suggest that how discrimination law has worked under certain conditions can inform our understanding of how law against crimes against humanity and genocide will work to the extent similar conditions come to prevail in that arena of law. And, although the particulars and environment differ in some respects, discrimination adjudication in the United States is buffeted by many of the dynamics mentioned above. Thus, the exploration may be informative.

---

[232] *See id.* at 86−88, 100−01, 130−42.

[233] *E.g.*, ALLPORT, *supra* note 10; *see* Lucas, *supra* note 10.

[234] Dudziak, *supra* note 55, at 45; POWER, *supra* note 16, at xvi–xvii..

[235] *Compare* Spirer & Seltzer, *supra* note 5, at 214–15, *with* David E. Bloom & Mark R. Killingsworth, *Pay Discrimination Research and Litigation: The Use of Regression*, 21 INT'L STUD. MGMT. & ORG. 318, 318–19 (1982).

[236] FREEDMAN, *supra* note 189, at 212.

A. STATISTICS IN DISCRIMINATION LITIGATION AS A PROFESSIONALIZATION PROJECT

Bloom and Killingsworth convey the history of the inclusion of one statistical method–regression modeling–in discrimination cases, with the first such Federal employment case decided in 1973. Regression is a specific case of the general linear model, and is used in legal proceedings as an inferential technique.[237] In the case of discrimination, it was plaintiffs that opened the statistical door, not defendants. Similarly, in the case of human rights adjudication, it is prosecutors, not defendants, who are taking the initiative.

It appears that plaintiffs' inclusion of statistical analyses diffused widely and quickly,[238] such that by 1976 one judge castigated plaintiffs for failing to include regression analyses.[239] That case, *Patterson v. Western Development Laboratories*, suggests that judges had already come to expect statistical evidence in discrimination cases.[240] This is just as is predicted to occur for human rights adjudication; court authorities will come to expect statistical corroboration, and its absence will weaken the case.

Since the mid-1970s a cadre of statistical consultants has become available for attorneys to hire. Multiple directories of such consultants exist, indicating the institutionalization of statistical analysis for discrimination cases.[241] That institutionalization signals an increase in employment opportunities relative to what the employment opportunities would be were statistics not accepted as evidence in discrimination cases. As the argument about whether to include regression (or statistical) models devolved into case-by-case debate concerning the properties a specific statistical analysis need have,[242] the employment impact of using statistics in discrimination litigation likely increased. Thus, one

---

[237] Note that Ball, et. al, *supra* 102, at 5–6 used the general linear model to analyze counts of killings in Kosovo. Thus, the same inferential statistical techniques are under consideration in the discrimination and human rights cases. Bloom & Killingsworth, *supra* note 235, at 319.

[238] Bloom & Killingsworth, *supra* note 235, at 320.

[239] *Id.*

[240] Patterson v. W. Dev. Labs, 13 Fair Empl. Prac. Cas. 772 (N.D. Cal. 1976).

[241] Statistical experts for discrimination cases join many other potential expert witnesses on directories such as JurisPro and ExpertPages. *See,* JURISPRO, http://www.jurispro.com/category/discrimination-s-141/ (last visited Feb. 25, 2012); EXPERT PAGES, http://expertpages.com/ (last visited Feb. 25, 2012). Certainly, many other kinds of experts service many other kinds of cases; this is not to denigrate the use of experts to aid litigation.

[242] Bloom & Killingsworth, *supra* note 235, at 320.

implication of the use of statistics in discrimination has been to increase employment of statisticians.

B. SELECTED EPISTEMOLOGICAL IMPLICATIONS OF STATISTICS IN DISCRIMINATION LITIGATION

I have maintained that statistics would become the central screen for rulings concerning human rights. In *Teamsters v. United States* the US Supreme Court affirmed its ruling that statistical evidence could establish the existence of discrimination.[243] And, in *Hazelwood School District v. United States*, also decided in 1977, the Court noted that statistical evidence alone was sufficient to prove the prima facia case of discrimination.[244] Analysts have concluded that statistics, usually in the form of regression models, has become the main method plaintiffs use to establish discrimination such as wage discrimination, which has led defendants to provide alternative models which undercut plaintiffs' claim that discrimination is occurring.[245] Indeed, some courts may discount testimonial evidence of discrimination when it is not buttressed by statistical evidence of discrimination,[246] which is consistent with predictions of the ascendance of statistical evidence in human rights adjudication.

Bazemore v. Friday justifies ignoring omitted variables and thus justifies consideration of only what can be placed in a model, suggesting how what is countable will be counted and what is not may be ignored, at least by the statistical analyst.[247] Although this case was decided in favor of plaintiffs, the logic is as predicted for human rights adjudication after full implementation of the statistics proposal–the countable will ascend in importance. And, I submit that the implications for human rights will favor defendants, especially guilty ones, for the reasons conveyed above.

Evidence also exists that defendants have responded with statistical techniques that *look* appropriate but, upon closer inspection, are shown to almost guarantee plaintiff defeat. For example, Bielby and Coukos show that efforts at class certification in discrimination cases are

---

[243] Int'l Brotherhood of Teamsters v. United States, 431 U.S. 324, 339 (1977).

[244] Hazelwood Sch. Dist. V. United States, 433 U.S. 299, 308−09 (1977).

[245] James T. McKeown, *Statistics for Wage Discrimination Cases: Why the Statistical Models Used Cannot Prove or Disprove Sex Discrimination*, 67 IND. L.J. 633, 633 (1992).

[246] Jenson v. Eveleth Taconite Co., 824 F. Supp. 847, 866 (1993), *later remanded for re-trial*, 130 F. 3d 1287 (1997).

[247] Bazemore v. Friday, 478 U.S. 385, 400 (1986).

virtually guaranteed to fail if courts accept defendant strategies to disaggregate analyses into site-specific assessments.[248] By rebutting the class certification claim, discrimination defendants force each member of the class to pursue an individual case, thus undermining the ability to pool resources, greatly increasing individuals' costs and thus greatly reducing the chance of continued litigation. Such techniques essentially shield discrimination by so raising the bar for proof of discrimination that a great deal of discrimination can occur without tripping concern. This is exactly what has been predicted for future human rights adjudication should the statistics proposal be fully implemented.

At the same time, observers lament the courts' embrace of statistical significance as a screen for effects of discrimination.[249] Often this embrace is uncritical[250] and some claim it is uninformed.[251] And, because statistical significance is a function of sample size, courts that have adopted statistical significance as a screen have essentially accepted some level of discrimination simply because of sample (or establishment) size.[252]

Evidence indicates that courts seek evidence of discrimination effects to conclude that discrimination occurred. Observers note that "[p]roving that a certain policy has broad discriminatory *effects* without statistics is obviously quite difficult."[253] This change in focus, from occurrences to effects, is just as predicted for human rights adjudication with full implementation of the statistics proposal.[254]

---

[248] William T. Bielby & Pamela Coukos, *"Statistical Dueling" with Unconventional Weapons: What Courts Should Know About Experts in Employment Discrimination Class Actions*, 56 EMORY L.J. 1563, (2007).

[249] *E.g.*, Allan G. King, *"Gross Statistical Disparities" as Evidence of a Pattern and Practice of Discrimination: Statistical Versus Legal Significance*, 22 LAB. LAW. 271(2007).

[250] David H. Kaye, *Is Proof of Statistical Significance Relevant?*, 61 WASH. L. REV. 1333 (1986).

[251] Arnold Barnett, *An Underestimated Threat to Multiple Regression Analyses Used in Job Discrimination Cases*, 5 INDUS. REL. L.J. 156 (1982).

[252] Ramona L. Paetzold, *Problems with Statistical Significance in Employment Discrimination Litigation*, 26 NEW ENG. L. REV. 395, 402−03 (1991).

[253] Emphasis added. Michael Aleo & Pablo Svirsky, *Foreclosure Fallout: The Banking Industry's Attack on Disparate Impact Race Discrimination Claims under the Fair Housing Act and the Equal Credit Opportunity Act*, 18 B.U. PUB. INT. L.J. 1, 27 (2008).

[254] There are multiple ways to justify judicial interest in effects. For example, courts might want to estimate effects of discriminatory treatment to aid determination of compensatory damages. Thus, investigating effects is not wholly inappropriate. A problem exists, however, when the exceedingly difficult task of estimating effects is treated as the way to establish the existence of discrimination. A better approach would be to use studies of effects only at the "damages" stage of the proceedings, using other evidence to establish whether discrimination occurred. See Lucas, *supra* note 10, at 206−16, for further consideration of the problems with estimating effects as a means to establish the existence of discrimination.

Finally, Browne, lamenting that shortcomings of statistical analysis had not diminished enthusiasm for statistics in discrimination cases, claimed that, "[I]nstead, limitations on the utility of statistical evidence have been relied upon not to limit the use of such evidence, but rather to modify evidentiary rules in ways that tend to obscure its lack of usefulness, and in the process to shift the burden of proof, albeit surreptitiously . . . ."[255] Browne further claims that although some courts are skeptical of statistics, "a proper amount of skepticism in cases where the proof is largely statistical would result, in virtually all cases, in judgment for the defendant."[256] It is just such a statistics-driven change in the interpretation of law, which some claim has occurred in the case of discrimination that is predicted for the case of human rights adjudication, with the result that, in virtually all cases, judgment would favor the defendant.

C. CAUSAL AND PARADIGMATIC CONTESTATIONS AND STATISTICS IN DISCRIMINATION LITIGATION

When US civil rights attorneys began to turn to statistical analysts for assistance in the early 1970s, key claims in twenty-first century debates concerning causal inference and bayesian statistics were just beginning to be developed. Thus, 1970s applied statistics did not engage these issues, such that courts of the 1970s and mid-1980s affirmed the value of a statistics that, at that very moment, was being radically challenged.

At the same time as some sociologists were raising serious questions about the possibility of statistical analyses to reveal causal effects, [257] statisticians were developing a thorough-going framework for re-thinking the statistical assessment of causality. Landmark theoretical and practical developments pioneered by Paul Holland and Donald B.

---

[255] Kingsley R. Browne, *Statistical Proof of Discrimination: Beyond "Damned Lies"*, 68 WASH. L. REV. 477, 478 (1993). In this excerpt Browne claims the burden shifts to defendants, but it seems what is meant is that the legal interpretations facilitate plaintiffs' efforts to certify an unequal outcome (but not its cause). As I show below, any other interpretation would be odd given available evidence on plaintiff win-rates in discrimination cases. Later, Browne contends courts, if they operated in line with statistical skepticism, would disadvantage plaintiffs.

[256] *Id.* at 557.

[257] *E.g.*, STANLEY LIEBERSON, MAKING IT COUNT: THE IMPROVEMENT OF SOCIAL RESEARCH AND THEORY (1985); Andrew Abbott, *Transcending General Linear Reality*, 6 SOC. THEORY 169, 169−86 (1998).

Rubin in the mid-1970s and early 1980s greatly disturbed previous confidence in the ability of statistical analyses to reveal causal effects.[258] The causal inference dust from this upheaval still has not settled, as analysts recognize that eminent statisticians strongly disagree as to whether statistical analysis may excavate causal effects and, if such is possible, how to proceed to excavate those effects.[259] In the academic discussion of discrimination cases that conflict is replicated.[260]

As for the bayesian/frequentist debate, that also became more pronounced in the late 1980s as bayesian analyses became increasingly feasible with increases in computing power and speed[261] which allowed implementation of what had been largely theoretical estimation methods such as the EM algorithm[262] and Markov Chain Monte Carlo (MCMC)[263] estimation. The frequentist/bayesian dispute has yet to be fully aired in court, but some analysts have proposed a shift to bayesian approaches in discrimination cases because of alleged incoherent aspects of classical statistical tests,[264] the use of bayesian methods for correcting for missing data in discrimination trials,[265] and more.[266] As US courts accepted statistical analysis for discrimination adjudication and set the rules of its use before the resurgence of bayesian statistics, simple inertia may have slowed its embroilment in the bayesian/frequentist conflict. But, that conflict may find its way into discrimination litigation eventually.

---

[258] Donald B. Rubin, *Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies* 66 J. EDUC. PSYCHOL. 688, 688−701 (1974); Paul W. Holland & Donald B. Rubin, *On Lord's Paradox, in* PRINCIPLES OF MODERN PSYCHOLOGICAL MEASUREMENT: FESTSCHRIFT FOR FREDERICK M. LORD 3 (H. Wainer & S. Messick eds., 1983).

[259] *E.g.*, Juni Palmgren, *Introduction to Causal Modeling and Inference*, 31 SCANDINAVIAN J. STAT. 159, 159−60 (2004).

[260] *E.g.*, Mary W. Gray, *Can Statistics Tell Us What We Do Not Want to Hear? The Case of Complex Salary Structures*, 8 STAT. SCI. 144, (1993); Harry V. Roberts, Comment, 8 STAT. SCI., 171 (1993); Mary W. Gray, Comment, 8 STAT. SCI., 177 (1993).

[261] Ira H. Fuchs, *Prospects and Possibilities of the Digital Age*, 145 PROC. AM. PHIL. SOC'Y 45, 45−53 (2001); Jean E. Vuillemin, *On Computing Power, in* PROGRAMMING LANGUAGES AND SYSTEM ARCHITECTURES 69, 70 (Jürg Gutknecht ed., 1994).

[262] Bruce W. Turnbull, *The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data*, JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B (METHODOLOGICAL) (1976) 38: 290−95; A. P. Dempster et al., *Maximum Likelihood from Incomplete Data via the* EM *Algorithm*, 39 J. ROYAL STAT. SOC'Y 1 (1977).

[263] Persi Diaconis, *The Markov Chain Monte Carlo Revolution*, 46 BULL. AM. MATHEMATICAL SOC'Y 179, 179–205 (2009).

[264] Paetzold, *supra* note 252.

[265] Joseph B. Kadane & Norma Terrin, *Missing Data in the Forensic Context*, 160 J. ROYAL STAT. SOC'Y 351 (1997).

[266] Browne, *supra* note 255, at 488 n.39.

D. PLAINTIFFS' PROSPECTS AND STATISTICS IN DISCRIMINATION LITIGATION

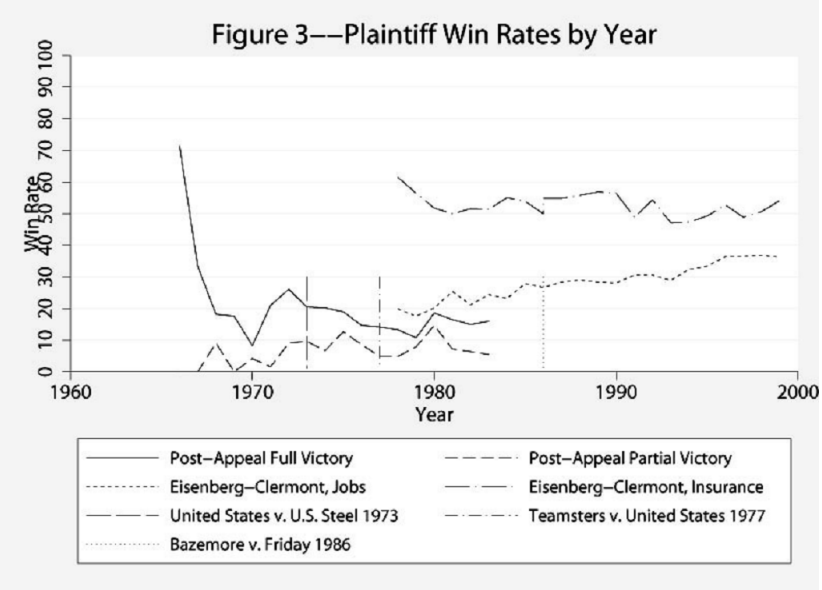
All these dynamics partly establish the terrain upon which plaintiffs confront defendants. Figure 3 reveals the over-time pattern of plaintiff success, as it contains four time series' of plaintiff win rates in Federal cases between 1966 and 1999. Two series covering the earlier period reflect post-appeal full and partial victory figures for discrimination plaintiffs.[267] As indicated, the first Federal employment discrimination case to use regression models was decided in 1973.[268] Use of regression statistics was more widely-affirmed in 1977, as the US Supreme Court clearly noted the sufficiency of statistical analyses for making a case for the existence of discrimination. From the 1970s apex in 1972, post-appeal victories declined steadily. Most of the decline followed *United States v. U.S. Steel*. The claim is not causal, but the post-appeal data show that the chances of prevailing fully fell short of 30 percent both immediately before and after that case. Full victories continued to decline after *Teamsters v. United States*, and the series does not recover to 20 percent in the data observed. Although the data available does not allow one to assess the role of statistics on a case-by-case basis, the low win rate is somewhat inconsistent with statistics providing powerful support for plaintiffs' efforts, especially as evidence indicates statistical analyses were becoming more common in plaintiffs' filings during this period.[269]

---

[267] Paul Burstein & Kathleen Monaghan, *Equal Employment Opportunity and the Mobilization of Law*, 20 LAW & SOC'Y REV. 355, 372 (1986).

[268] Bloom & Killingsworth, *supra* note 235.

[269] *Id.* at 320.



The later period is captured in the Eisenberg-Clermont data; one series contains employment discrimination cases (labeled jobs), one series contains insurance cases.[270] Neither series identifies final victories after appeal. Further, the jobs category is only a subset of discrimination claims. Still, as statistics are often used in employment discrimination cases, the category provides useful information as to whether statistical evidence has aided plaintiffs.

The Eisenberg-Clermont data reveal a steady increase in win rates for plaintiffs. However, the trend predates *Bazemore v. Friday* and extends at an apparently similar rate afterwards. Because the trend started before the cited case the pattern is not solid evidence that

[270] Theodore Eisenberg & Kevin M. Clermont, Judicial Statistical Inquiry, CORNELL.EDU (JUNE 2002), http://legal1.cit.cornell.edu:8090; My comparing discrimination plaintiff win rates with insurance plaintiff win rates replicates over time some aspects of Michael Selmi, *Why are Employment Discrimination Cases So Hard to Win?*, 61 LA. L. REV. 555 (2001).

statistics provides support for plaintiffs, especially as we realize the latter series is not limited to final win-rates after appeals. Indeed, as evidence indicates that discrimination plaintiffs are more likely to lose pre-trail motions, and, should they win at trial, they are more likely to lose on appeal than they are to prevail,[271] this omission biases results, if anything, toward more plaintiff victories. Given the rate of defeat on appeal, it is likely that post-appeal data would be even less consistent with the claim that statistics improves plaintiffs chances of prevailing.

Indeed, once one considers more of the dispute pyramid, matters are even less sanguine for discrimination plaintiffs. Laura Beth Nielsen, Robert L. Nelson, and Ryon Lancaster intensively studied a probability sample of 1,672 cases filed in Federal Court during the 1988-2003 period, and found nearly 20 percent of cases were dismissed outright, while another 18 percent of plaintiffs lost on summary judgment and only 6 percent of the cases went to trial.[272] Of all the plaintiffs filing cases–already a selected subset of those who might do so–only 2 percent prevailed at trial.[273] Indeed, evidence indicates that discrimination plaintiffs are less likely to win than are any plaintiffs other than prison convicts.[274] These results hardly provide a ringing endorsement of the kinds of evidence plaintiffs are able to present, including their ability to present statistical evidence.

One alternative explanation for the rarity of plaintiff victories is that discrimination is rare. Yet, evidence suggests the opposite. Study of seventy-one different audits of housing markets, [275] coupled with research from audits of insurance companies,[276] car dealerships,[277] as well as employers[278] all suggest that discrimination is widespread. Thus,

---

[271] Kevin M. Clermont & Stewart J. Schwab, *How Employment Discrimination Plaintiffs Fare in Federal Court*, 1 J. EMPIRICAL LEGAL STUD. 429 (2004).

[272] Laura Beth Nielsen et al., *Individual Justice or Collective Legal Mobilization? Employment Discrimination Litigation in the Post Civil Rights United States*, 7 J. EMPIRICAL LEGAL STUD. 175, 187 (2010).

[273] *Id.* at 175−201.

[274] Theodore Eisenberg, *Litigation Models and Trial Outcomes in Civil Rights and Prisoner Cases*, 77 GEO. L.J. 1567, 1578 (1989).

[275] George Galster, *Racial Discrimination in Housing Markets during the 1980s: A Review of the Audit Evidence*, 9 J. PLAN. EDUC. & RES. 165, 165–75 (1990).

[276] Gregory D. Squires & William Velez, *Insurance Redlining and the Process of Discrimination*, REV. BLACK POL. ECON., Winter 1988, at 63, 63–75 (1988).

[277] Ian Ayres & Peter Siegelman, *Race and Gender Discrimination in Bargaining for a New Car*, 85 AM. ECON. REV. 304 (1995).

[278] Marianne Bertrand & Sendhil Mullainathan, *Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination*, 94 AM. ECON. REV. 991 (2004).

it is unlikely that the 2 percent plaintiff success rate at trial is an indication of a low level of discrimination in US society.

An alternative explanation is that discrimination is not rare, but weak cases (either clearly guilty defendants or clearly frivolous plaintiffs) settle.[279] However, most grievants do not formally file a grievance, [280] and most settlements are small despite media coverage suggesting otherwise.[281] Collectively, these findings are consonant with the claim that the use of statistics in discrimination litigation does not seem to aid plaintiffs to consistently obtain redress.

That statistics has not aided plaintiffs is readily understandable. In the normal course of business potential discriminators have opportunities to produce records that can make it difficult for statistical analyses to discover discrimination. Potential discriminators in many litigation situations are necessarily gatekeepers, and as such they can systematically collect some information and systematically *not* collect other information, simply as a matter of business practice. Further, they can employ consultants prior to making or implementing a decision to help them do what they would like to do in a way that minimizes the likelihood of discovery and, if discovered, minimizes the likelihood of culpability attaching to the potential defendant. Thus, despite the noteworthy differences between international efforts to bring genocidal actors to justice and the court system of a particular nation-state, findings from the case of discrimination litigation in the United States should give one pause. A broad assessment of the case data indicates that statistics has not clearly aided plaintiffs. Further, that failure to aid has developed in an environment produced by successful efforts on the part of dominant elites to contain movements for deeper social change.[282] Similar dynamics buffet the environment within which the contemporary effort

---

[279] George L Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 13 J. LEGAL STUD. 1 (1984). Note, however, that the weakness of a case is partly determined by the state of the law– if the legal terrain makes prevailing very difficult, clear sociological cases of discrimination will be weak legal cases to file.

[280] Richard E. Miller & Austin Sarat, *Grievances, Claims, and Disputes: Assessing the Adversary Culture*, 15 LAW & SOC'Y REV. 525 (1980).

[281] Laura Beth Nielsen & Aaron Beim, *Media Misrepresentation: Title VII, Print Media, and Public Perceptions of Discrimination Litigation*, 15 STAN. L. & POL'Y REV. 237 (2004) (showing substantial over-estimates of win-rates and awards in media reports).

[282] *See, e.g.*, Derrick A. Bell, Jr., Brown v. Board of Education *and the Interest-Convergence Dilemma*, 93 HARV. L. REV. 518, 518–33 (1980); Alan David Freeman, *Legitimizing Racial Discrimination Through Antidiscrimination Law: A Critical Review of Supreme Court Doctrine*, 62 MINN. L. REV. 1049 (1982); Cheryl I. Harris, *Whiteness as Property*, 106 HARV. L. REV. 1707 (1993); DUDZIAK, *supra* note 55.

to empower institutions to enforce transnational human rights is occurring.[283] Given these environmental and substantive similarities, I submit that we need not wait for the process to fully unfold at the ICC before we may apprehend the likely result.

E. LESSONS OF HISTORY: STATISTICAL EVIDENCE AND DISCRIMINATION

Discrimination litigation in the United States provides many lessons for how full implementation of the statistics proposal may impact human rights adjudication. Although the historical and intellectual environments differ, the discrimination/human rights comparison is apt because both concern categorical treatment; and enforcement of protective measures has been and continues to be resisted by some powerful actors.

These and other aspects of the complex dynamics culminate in exceedingly low plaintiff win rates. Certainly multiple factors determine win rates, including changes in the law, the changing composition of the judiciary, changing economic conditions,[284] and more. However, plaintiff win-rates in discrimination cases are not zero. It is important to acknowledge a key difference between the discrimination and human rights arenas that advantages discrimination plaintiffs in comparison to human rights prosecutors. Discrimination trial procedures were crafted in the United States in line with 1970's era statistics. As more recent statistics is aware of more sophisticated threats to causal inference, it will be harder to certify cause if contemporary statistics is institutionalized in the legal system. This suggests discrimination win-rates probably over-estimate, and perhaps substantially over-estimate, the win-rates one can expect for human rights prosecutors once the statistics proposal is fully implemented.

Are low plaintiff win-rates a result of reliance on statistics? Plaintiff win rates do not seem to have improved owing to the availability of opportunities to present statistical evidence. In the end, assessing this question is most difficult, because the most straightforward

---

[283] *E.g.*, YOO, *supra* note 28; MARGULIES *supra* note 28; *see* Marlene Wind, *Challenging Sovereignty? The USA and the Establishment of the International Criminal Court*, 2 ETHICS & GLOBAL POL. 83 (2009).

[284] Peter Siegelman & John J. Donohue, III, *The Selection of Employment Discrimination Disputes for Litigation: Using Business Cycle Effects to Test the Priest-Klein Hypothesis*, 24 J. LEGAL STUD. 427 (1995).

assessment is not currently possible and not clearly correct. It is not possible at this time to separate a broad sample of discrimination cases into those with plaintiffs and/or defendants who presented statistics and those with plaintiffs and/or defendants who did not. If one could make this separation, one could assess win rates in the different types of cases to see whether statistical evidence is associated with higher plaintiff win rates. However, this proposed analysis ignores that one powerful implication of the interjection of statistics into discrimination litigation is that judges were thereby socialized to expect statistical evidence and thus to see discrimination claims that lack statistical support as suspect. A plaintiff cannot avoid the implications of that expectation by failing to submit statistical evidence. Consequently, it is unlikely that the question of statistics' utility can be answered by comparing the outcome of cases that use and do not use statistics. Instead, analysis of the changing environment is preferable. Given this observation, the analysis above is informative.

VII. CONCLUSION: THE STATISTICS PROPOSAL AS FINAL SOLUTION

In the foregoing analysis the statistics proposal is interpreted as a moment in statisticians' professionalization project and, thus, despite the good intentions of the statisticians involved, the statistics proposal bears the interest of statisticians as a class in rendering what they have to offer indispensable to the court. In order to accomplish this aim, statisticians must undermine the sufficiency of the oral, textual, and physical evidence that has traditionally formed the basis of investigation, prosecution, and judgment. As statisticians undermine that evidence, they will become the validators of victims' accounts primarily because victims cannot generally testify about the pattern of atrocities. Yet, as the early cases show, this predicament has been no bar to judges' ability to discern patterns; consequently, the statistics proposal offers little to victims. However, by providing a hitherto non-existent means to defend against the allegations, the statistics proposal offers much to defendants, especially to guilty defendants who may take the opportunity to prepare a statistical defense prior to and during the period of atrocities. This new means of defense is made real even as statisticians subtly alter the *de facto* content of crime against humanity and genocide, most notably by conflating incidents and effects in a way that requires the estimation of effects in order to document the existence of incidents. Further, as effects

are much harder to establish than incidents, the statistics proposal implicitly, perhaps unwittingly, raises the bar for establishing the existence of a crime against humanity or genocide to a nearly unattainable height. At the same time, statistics de-emphasizes the horror dimension and heightens the number dimension, yet the number dimension will only capture atrocities that are feasible to count. This feature will motivate criminals to: 1) emphasize assaults that are difficult to count, 2) prepare the way for their atrocities ahead of time by under-counting the target population, 3) calibrate their actions during the infliction of atrocities in an effort to keep the number of victims low enough (or perhaps random enough) to undermine detection, and, 4) afterward, to explain away the results with alternative statistical analyses. All this and more will occur even as statisticians, divided paradigmatically and collectively doubting the ability of statistics to secure causal inferences, demonstrably lack a solid basis for advising the court. Consequently, if statistical evidence is required for proceedings to go forward, the chance of obtaining convictions on charges of having committed a crime against humanity or genocide may almost completely evaporate as may the ability to defensibly certify the very existence of transnationally actionable human rights violations, as the entire adjudicatory process becomes captive to an increasingly obfuscatory battle of statisticians. Finally, although sovereigns seeking to avoid human rights enforcement obligations may not have sponsored the statistics proposal, they may embrace its implementation as its ability to reduce the number of prosecutable cases becomes evident.

One might ask, however, what is the alternative? Fully addressing that question is beyond the scope of the analysis, but one can briefly observe that an alternative is eminently visible, for the historic solution offers a developing example of another way.[285] Notably, the historic solution is based on a principle the statistics proposal mutely, almost imperceptibly dismisses–every person counts. Because in the historic solution every person counts, in principle one incident, indeed, one violation of one person, can convict a general, a president. Clearly, that has not generally been the case in the emerging system of transnational justice, but the gap between principle and reality reflects the incomplete realization of the principle of the inviolability of every human life, not the rejection of the principle of the inviolability of every

[285] *E.g.*, Audrey I. Benison, *War Crimes: A Human Rights Approach to a Humanitarian Law Problem at the International Criminal Court*, 88 GEO. L.J. 141 (1999).

human life. However, the statistics proposal implicitly rejects this principle, as the analysis above has shown. This is a high price to pay for the dubious promise of statistical justice. Instead, if the historic solution is further elaborated, while the inviolability principle is continually affirmed as an operating commitment of the developing system, backed eventually by prosecutions and convictions that demonstrate the sincerity with which the principle is held, generals and presidents must become concerned about the lives and experiences of every single civilian in every single village they administer or through which their military units pass. Long and difficult though the process of bringing this possibility to fruition may be, the statistics proposal aborts this developing possibility, and substitutes a far less universal admonition–no need to protect everyone; instead, simply avoid abusing "too many."

Even partial adoption of the statistics proposal lodges tolerance for atrocities at the core of human rights enforcement. Beyond that, statistics' bias toward indeterminance will make it extremely difficult, if not impossible, to convict even guilty defendants of human rights violations once defendants enlist the aid of statisticians in their defense. The indeterminance makes the statistics proposal an ideal instrument for securing *de facto* sovereign immunity, for several reasons. First, statistical analysis procedures are technical enough to limit the ability of non-statisticians to evaluate the analyses. Second, the procedures treat uncertainty as the default position, a position it will be even harder to reject once analysts employed for the purpose of discovering uncertainty are enlisted on the side of the accused. Such implications would unfold in a context sculpted, in part, by the dual role of the nation-state in relation to human rights and two convergent interests to which the dual role gives force: 1) an interest to maintain human rights rhetoric as a resource nations may deploy at will, and 2) an interest to legitimate the unfettered action of sovereigns. Both the rhetorical resource line and the sovereign immunity line reflect and construct sovereign interest in resisting the establishment of effective transnational human rights adjudication. In a context shaped in part by these interests, preserving sovereign immunity is a structural predisposition of key actors, and the statistics proposal would likely further that aim despite the very different intentions of proposal architects. In such a context it is no exaggeration to observe that full implementation of the statistics proposal may constitute a Final Solution to the problem human rights law poses to sovereignty.

Thus, of what use is statistics to those seeking to provide rights to defendants while preserving courts' ability to convict the guilty? Statistics may be useful as a means through which, post-atrocity, analysts ascertain community desires,[286] if processes of data collection are inclusive, other means to determine victim desires are also employed, and the analysis reports simple descriptive statistics rather than complex parameter estimates or findings dependent upon methods of data imputation. Statistics is also useful for academic research that endeavors to discern large patterns amidst the noise of social life. This information can aid efforts to consider various policy proposals, though the complex feedback loops that attend policy implementation make proposing any specific policy intervention a task that requires caution.

Yet, major limits on the use of statistics exist. Despite all the technical trappings of precision it can marshal–standard errors, confidence intervals, α-levels for Type-I error, β-levels for Type-II error, p-values, likelihood ratios, and more–statistics remains both insufficiently precise *and* insufficiently broad to answer whether a crime against humanity or genocide was committed and, if so, who committed it upon whom. We reach this conclusion without introducing many other challenges that attend efforts to establish causation with observational data, such as missing data problems,[287] measurement error,[288] identification difficulties,[289]selection bias,[290] and more, on which any defendant may also draw to rebut accusatory analyses.

Certainly, the challenges raised by the prospect of determining the veracity of allegations of crimes against humanity and genocide and, if true, prosecuting persons accused of crimes against humanity or genocide, are daunting. Meeting those challenges is essential if the web weaved by the promulgated statutes in support of human rights is to ultimately undergird even the physical safety, not to mention the full flourishing, of persons. Still, those challenges, serious as they are, pose

---

[286] Phuong Pham & Patrick Vinck, *Empirical Research and the Development and Assessment of Transitional Justice Mechanisms*, 1 INT'L J. TRANSITIONAL JUST. 231 (2007).

[287] RODERICK J.A. LITTLE & DONALD B. RUBIN, STATISTICAL ANALYSIS WITH MISSING DATA (2d ed. 2002).

[288] *E.g.*, KENNETH A. BOLLEN, STRUCTURAL EQUATIONS WITH LATENT VARIABLES (1989).

[289] *E.g.*, Robert M. Hauser, *Context and Consex: A Cautionary Tale*, 75 AM. J. SOC. 645 (1970); CHARLES F. MANSKI, IDENTIFICATION PROBLEMS IN THE SOCIAL SCIENCES (1995).

[290] *E.g.*, Richard A. Berk, *An Introduction to Sample Selection Bias in Sociological Data*, 48 AM. SOC. REV. 386 (1983); Barbara Geddes, *How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics*, 2 POL. ANALYSIS 131 (1990).

far less of a threat to securing human rights than does the statistics proposal as a response.

However, now that the statistics proposal is supported through a burgeoning institutional structure, and has begun to be adopted by prosecutors in cases, defendants may seek to introduce statistical evidence in their defense. Defendants can and should petition to present whatever evidence they deem relevant. However, what this means is that the only hope now for resisting the statistics proposal is to withhold legitimacy from statistics for use in adjudicating allegations of crimes against humanity and genocide, and the only way to withhold such legitimacy is for prosecutors to refrain from using such evidence in their original indictment and in the development and prosecution of cases. Statistical analysis must not be the prosecutor's chosen tool, although in response to defendant use of statistics the prosecution can certainly bring forth statisticians to undercut the claims of the defendant's statistical expert, especially the claim that statistics has enough precision and solidity behind it to provide causal insight for the court. The difference, of course, is that using statistics to rebut defendants' use of statistics by showing that statistics is unhelpful does not legitimate statistics as a basis for adjudicating human rights allegations, whereas basing prosecutions on statistics does legitimize statistics for adjudicating human rights allegations. Indeed, prosecutors who rely on statistics in some cases may thereby encourage each defendant to employ statistical experts to undermine the prosecution's position in every case before the court, even in cases for which the prosecution refrained from using statistical analysis in its indictment. Accordingly, perhaps the most eloquent way for prosecutors to convey that statistics has virtually no value for human rights adjudication is to eschew statistics as they build solid cases for the ICC to consider, relying only on statistics if the need arises to rebut a defendant's implicit claim that statistics has value for the proceedings. I admit that this response still increases the employment (and perhaps the stature) of statisticians, at least in the short term, but at least the damaging impact of the statistics proposal on human rights enforcement would be resisted with this response.

Resisting the statistics proposal is of paramount importance. For, ultimately, to adopt the statistics proposal is likely to unleash a lamentable Final Solution, to transform judges' chambers into death chambers within which the incipient institutionalization of transnational human rights will be perhaps forever extinguished. All the reasons herein developed, and more, point in this direction. Sadly, and despite

appearances and the best intentions of its advocates, if sociology, epistemology, statistics, history, or logic are any guide, it appears that adopting the statistics proposal *is* to set out on the road to hell, a destination, much closer than it appears, where, outside of victor's justice, no allegation of a crime against humanity or genocide can ever be substantiated, no sovereign, soldier, or civilian need ever fear having to answer for any human rights violation they might commit and, thus, where no innocent can trust the deterrent power of a prospect of future justice to secure their safety should times of trouble threaten to descend upon them and those they love.