Employment Discrimination Law and Industrial Psychology: Social Science as Social Authority and the Co-Production of Law and Science*

Robin Stryker, Department of Sociology and Rogers College of Law, University of Arizona

Danielle Docka, Department of Sociology, University of Minnesota

Pamela Wald, Marketing Department, University of the Sciences

January 30, 2011

*Previous versions of this paper were presented at the Conference on the Discoveries of the Discrimination Research Group, Stanford Law School, November 7-8, 2008, the Conference on Law and Regulation of Economic Activity, École normale supérieure de Cachan, Cachan, France, October 2-3, 2008, and the American Sociological Association, Atlanta, August 14-17, 2010. We thank conference participants for their comments and critiques. We also thank Mary Driscoll, Barry Goldstein, Marc Bendick, Lauren Edelman, Joshua Guetzkow and Sandra Kalev for their comments, and we thank J. Taylor Danielson and Zachary Schrank for additional research assistance. National Science Foundation Grant #SES-0514700 (2005-2009) and a John Simon Guggenheim Foundation Fellowship (2008) to the first author supported this research. This research benefited from the first author’s participation in a research working group entitled, "Social Science Perspectives on Employment Discrimination in Organizations,” which is part of the Discrimination Research Group. The Discrimination Research Group is a joint effort funded by the American Bar Foundation, the Center for Advanced Study in the Behavioral Sciences, and the Ford Foundation (grant #1045-0189). This research also benefitted from the first author’s participation in the Workplace Diversity Conference, organized by Frank Dobbin, at the Radcliffe Institute of Advanced Study, Cambridge, Massachusetts, April 25-26, 2008. Please direct all correspondence to Robin Stryker at rstryker@email.arizona.edu
Employment Discrimination Law and Industrial Psychology: Social Science as Social Authority and the Co-Production of Law and Science

ABSTRACT

This article examines how and with what consequences Title VII and industrial psychology became co-produced. Combining Monahan and Walker’s classification of social facts, social authority and social frameworks with political-institutionalism’s view of law and science as competing institutional logics helps explain this. New legislation provides opportunities for social science knowledge to penetrate law and become translated into legal terms. When social science is incorporated into interpretation and enforcement of legislative law as social authority--rationale for judicial rule-making—law’s institutional logic of relying on precedent and reasoning by analogy ensures that social science will have ongoing influence on law’s development. By helping set research agendas and providing new professional opportunities, institutionalized legal doctrine shapes social science knowledge. But because of differences in institutional logic, wherein legal cumulation is backward looking whereas scientific cumulation is forward looking, co-production of law and science may produce mismatch between legal doctrine and scientific knowledge. Building on their empirical research, the authors suggest it would be fruitful to develop a more general socio-legal theory of conditions for and consequences of law-science co-production.
In April, 2010, the 9th Circuit affirmed class certification in a well-publicized class action gender discrimination case against Wal-Mart, exposing the company to potential damages over one billion dollars. In December, 2010, the Supreme Court granted certiorari to hear some, but not all of Wal-Mart’s objections (Dukes v Wal-Mart [9th Cir. 2010], cert granted 562 US PAGE). One ongoing controversy highlighted by the case involves law-social science intersection. Social science generates knowledge of probabilistic patterns of co-variation or (loosely speaking) cause-effect; adjudication determines “specific causation” pertaining to the dispute (Stryker 1994; Faigman 2008). Controversy is over whether and how much relevance and weight, under the Supreme Court’s Daubert standards for judicial gate-keeping of expert testimony, should be accorded what Walker and Monahan 1987; Monahan and Walker 2010a) termed social framework evidence—expert-provided empirical context for judicial fact-finding and determining liability (Bagenstos 2007; Borgida and Fiske 2008; Faigman 2008; Mitchell and Tetlock 2009; Hart and Secunda 2009). While not resolving this controversy, we locate it in broader socio-legal terrain including a less controversial, but perhaps more consequential law-science intersection: social science as social authority—rationale for legal rule-making (Monahan and Walker 1986, 2010a).

Long before Monahan and Walker coined the term, industrial psychologists provided key elements justifying the disparate impact theory of
discrimination under Title VII of the Civil Rights Act (CRA) of 1964, prohibiting employment discrimination based on race, sex, color, religion and national origin. With roots in the late 19th century, industrial psychology was relevant to employment discrimination law because industrial psychologists were developing concepts, theories and cumulative research about abilities, skills, and other performance-related worker characteristics, applicant recruitment, screening, socialization and training, and performance evaluation for employees and human resource management. Industrial psychologists also helped develop job and task analysis, aptitude and achievement tests, test validation, performance appraisal and new data analytic techniques, including meta-analysis (Thornton and Wingate; Kehoe and Olsen 2005; Murphy 2000).

First labeled economic/business psychology and then industrial psychology, the field became known as industrial-organizational (I-O) psychology in 1973, when the American Psychological Association’s Division of Industrial Psychology (Division 14) added the word “organizational” to its title. Since 1982, Division 14 has parallel existence as the independently incorporated Society for Industrial-Organizational Psychology (SIOP) (Koppes 2003). We use the terms I-O psychology and industrial psychology interchangeably.

In the remainder of this article, we use public and private documents supplemented by a small number of in-depth interviews to show how and with what consequences Title VII and I-O psychology intertwined to build the
foundational effects-based theory and method of proving employment discrimination called disparate or adverse impact. Combining Monahan and Walker’s (2010a) classification of social facts, social authority and social frameworks with insights gained by viewing law and science as competing institutional logics helps explain why I-O psychology had such profound impact on Title VII and vice versa.

CO-PRODUCTION AND COMPETING INSTITUTIONAL LOGICS

Max Weber appreciated that law shaped—and was shaped by—society, with legal rationalization coterminous with rationalization in other realms, including science (Edelman and Stryker 2005). Jasanoff (2004, pp. 1, 3) developed “the idiom of co-production” for the state and science, exploring how state governance incorporates knowledge making and reciprocally, how knowledge shapes governance. In science studies, Jasanoff (1995), Lynch (2004) and Gieryn, Zevins and Behr (1985) examined how courts shaped boundary-making between science and non-science in litigation over evolution, and when courts decide expert testimony’s admissibility and credibility at trial. Expert testimony issues are central to socio-legal studies, especially after the Supreme Court decision in Daubert v. Merrell-Dow Pharmaceuticals (1993) making judges gate-keepers for “real” vs. “junk” science (Bielby 2003; Faigman and Monahan 2005). For over thirty years, socio-legal research has focused on science’s use/misuse in court, functions it serves at trial and appellate review, and barriers faced in promoting law’s reliance
on good science (Collins 1978; Saks 1980; Monahan and Walker 1986; Walker and Monahan 1987; Tanford 1990; Chesler, Sanders and Kalmuss 1988; Schuck 1993; Goldberg 1994; Faigman 1999; Krieger and Fiske 2006; Fiske and Borgida 2008).

Stryker (1994, 2000) developed a political institutional perspective showing how incorporating scientific reasoning and experts into law affected law’s internal logic of rule-making and interpretation, occasioning lawyer-scientist conflict, and affecting law’s legitimacy and power. Savelsberg (1994, Savelsberg et al. 2002) focused on how states, including courts, shaped science. Dezalay and Garth (2002) showed how power and legitimacy competitions in the United States and Latin America externalized supply and demand for legal and economic expertise. International flows of economists and economic knowledge transformed national fields of state power.

Building on such scholarship, we offer a co-production perspective on Title VII and industrial psychology, focusing on mutual construction of legal and social-scientific knowledge and institutions. We link co-production to power and legitimacy issues in political-institutional perspectives on law and society (see Dezalay and Garth 2002; Edelman and Stryker 2005), and especially to Stryker’s (1994, 2000) political-institutional view of law and science as ideal-typical competing institutional logics (Stryker 2000).

Socio-legal scholars long pointed to tensions between principles and practices of law and science (Loevinger 1966; Schuck 1993; Goldberg 1994). For Anglo-
American case law, some emphasized barriers to court reliance on science from socializing lawyers and scientists into different world-views and beliefs (Chesler, Sanders and Kalmuss 1988; Stryker 1989). Others noted expert witnessing in adversary proceedings creates role conflicts and ethical dilemmas for scientists. Social scientists with objectivity, disinterestedness and neutrality norms feel discomfort when lawyers push them to eschew nuance and deliver unequivocal support for one side (Chesler et al 1998). Tanford (1990) points to tension between science’s idea of truth and law’s idea of legal accuracy; unreliable information such as from eye-witness accounts is included, but relevant information can be excluded, because, e.g., law considers it unfairly prejudicial to a party or obtaining it violated a criminal defendant’s Fourth Amendment rights. Still others emphasized barriers to effective communication across discourse communities (Sohn 2005).

Stryker’s (2000) political-institutional perspective identified competing institutional logics, systematizing law-science tensions in ideal-typical form for adversary, case-law systems. Tensions are manifested through what are characteristic, legitimate and authoritative field-specific reasoning and practices.

As socio-legal scholars know, legal reasoning in case law is precedent-based, analogical reasoning interpreting and applying judge-made rules by
establishing similarities and differences between current and prior cases. Since no two cases are identical, deciding what cases the current case is like or unlike requires extracting what is deemed pertinent to the analogy. Prior cases a court does not want to consider as precedent are distinguished away by explaining why analogizing them to the current case is inappropriate. The court’s holding/ruled becomes precedent refined in later cases. Interpreting legislation adds concepts and strategies, but these are incorporated into precedent-based reasoning.

In adversary litigation at trial and in appellate advocacy, lawyers advocate the strongest case for clients in facts they develop and rules they seek. Justifying ideology is that truth emerges when judges act as neutral arbiters and the judge or jury decides facts. Fact finders weigh evidence introduced by examining and cross examining witnesses in accord with procedural rules allocating proof burdens between plaintiff and defendant in civil cases and state and defendant in criminal cases consistent with such values as due process, individual rights and presumed innocence. Trials satisfy rules of a closed system but are not designed to produce factual truth (Tanford 1990).

\[1\] Unless noted, we rely on Stryker (1994, 2000). Disputes in code law create cases, but power balance between courts and legislatures is different. Co-production in code law is beyond our scope.
Other organizational features are hierarchy, judicial review and division of labor between trial and appellate courts. A federal employment discrimination case is heard first in federal district court. The losing party may appeal matters of law to the appropriate federal circuit court. The circuit court loser may petition the Supreme Court for certiorari, but relatively few cert petitions are granted.

From a political-institutional perspective, most legal change occurs through small-step incremental decision-making leading to evolutionary case law accretion. There is path dependency wherein precedent defines likely legal issues and argumentation frames (Stone Sweet 2002). Case outcomes cannot be known in advance, but argument terrain is predictable, coming from extant argument frames and case sequencing in them. That lower courts must apply as precedent prior rulings from higher courts covering their jurisdictions also promotes path dependency. Courts can overrule their own precedent, but this is rare. The US Supreme Court set stringent conditions for when this may occur legitimately.

---

2 Prior to the Civil Rights Act of 1991, amending Title VII to allow money damages, all employment discrimination cases were bench trials.

3 Judges must consider extent to which the public has relied on the prior rule, its workability, other relevant doctrinal change, and change in facts or perceptions of facts (Friedrichs 2006, pp. 59-60).
New legislation sets parameters for new argument frames in case law enforcing it. Courts apply legislation with “common law plus” techniques including construing legislative intent. In initial cases under new legislation—cases of first impression—judges refer directly to law’s text and legislative history to infer legislative intent. But once interpreted by courts for several years, legislative intent typically is imputed by invoking precedent.

Two previously implicit features now can be underscored. Case law’s cumulative logic is backward looking, constrained by precedent (Krieger 2004; Krieger and Fiske 2006). Overruling prior argumentation frames and new legislation are exogenous shocks to prior frames. (Only the second is exogenous to courts). Such “shock” events should be especially conducive to entry of new elements, including scientific ones, into judicial reasoning. Cases of first impression under new legislation set parameters for new argumentation frameworks typically reinforced and elaborated by later case law.

How might ideal-typical scientific institutional logic conflict with this ideal-typical legal logic? Science and law are not unitary, but appropriate ideal-typical

---

4 This does not preclude new legislation from “undoing” statutory interpretation the legislature does not condone, as with the 1991 CRA. Once enacted, new legislation resets parameters on argument frames addressed.
characterization sensitizes us to tensions shaping law-science intersection and co-production.  

First, though originality often cannot jump far ahead of extant knowledge or break paradigms, science emphasizes and rewards new discovery and can be seen as forward looking, against backward looking legal reasoning (Polanyi 1962, Krieger 2004; Krieger and Fiske 2006). Second, as suggested, scientific aims of truth production may conflict with what Tanford (199) called legal accuracy.

Third, rivalry may spur discovery, but scientists also cooperate, contrasting with common law’s purposeful embrace of adversariness, wherein each side slants relevant, dispositive evidence to a pre-determined outcome—client victory. Scientific theories guide empirical discoveries, but slanting data to protect a pet theory is illegitimate. Scientists must consider how and why they could be wrong.

Fourth, scientists sometimes think analogically, but social science’s dominant reasoning is causal analytic—searching for patterns of association and cause and effect. Causal logic has many styles, including putatively universal propositions and

---

5 Hacking (1996, p. 65), suggested that science includes “algorithmic” and “postulational” mathematical reasoning; experimentally exploring and measuring “complex detectable relations;” “hypothetical modeling;” “ordering of variety by comparison and taxonomy;” “statistical analysis of populations;” and “historical derivation of genetic development.”
limited, contextualized and conditionalized statements. Broadened thusly, causal analysis demarcates social science from cultural studies (Mahoney and Rueschemeyer 2003, p. 23).

Causal logic with empirical research captures social science’s core for legal policy making because it identifies means to achieve politically-selected legislative ends, evaluating whether those ends are—or can be—met through current precedent-based enforcement (Stryker 1994). If not, conflict must be resolved either by following precedent or—as in Brown v. Board of Education’s (1954) rejection of “separate but equal,” following implications of science to create a new argumentation framework.

To convince lawyers and judges to use social science, as in Brown, to justify making new legal rules, Monahan and Walker (1986, 2010a) coined the term social authority, in analogy to precedent’s role as legal authority. They

6 One reviewer of an earlier draft found this analogy unpersuasive because there often is no consensus among social scientists about theories or findings. On one hand, we build on the concept of social authority because it now is widely accepted and institutionalized in legal training and discourse (Monahan and Walker 2010a, b). On the other hand, in our political-institutionalist perspective on law, the degree of scientific consensus is a key variable affecting whether courts invoke particular scientific theories or research as social authority to help
considered legitimate both social authority and social fact —introducing social science into evidence to help judicial fact finding. Naming by analogy may increase lawyers’ comfort, but does not erase tensions when causal vs. precedent-based reasoning are resources for opposing outcomes.

A fifth tension revolves around courts’ need to determine—within burden of proof parameters—what did happen in the litigated case relative to scientists’ capacity to provide general, probabilistic cause-effect statements or at least reliable co-variation patterns (Stryker 1994; Faigman 2008). Science looks for “universals hiding among the particulars; trial courts look for “particulars hiding among the universals” (Faigman 1999, p. 69). This distinction structures debate over definition and reach of social framework evidence.

Walker and Monahan (1987) proposed the social framework concept to explain how cumulative, valid and reliable, peer-reviewed research providing probabilistic knowledge of co-variation or cause-effect legitimately provided conceptual and empirical context informing case-specific fact finding by judges justify their legal rulings (Stryker 1994). This article shows how scientific consensus figures into the development of disparate impact theory. More generally, we suggest that, all else equal, the greater the consensus around the legally relevant science at the time of court decisions, the more likely that courts invoke it to help justify legal rule-makings.
and juries. Legal scholars differentiate general causation familiar to scientists (there is an association pattern between smoking and lung cancer and we know probabilities and mechanisms by which the former produces the latter), and case-specific causation (smoking, rather than other possibilities, did cause lung cancer in this patient). They debate what is required to prove each, the role such proof plays in deciding whether defendant acted illegally, and how fact-finding authority on general vs. specific causation should be allocated between scientific-technical experts, and judges and juries (Faigman and Monahan 2005; Borgida and Fiske 2008). Some argue that social framework testimony never should extend beyond reviewing research on general causation to opine that it was more likely than not that the pattern applied to this particular case (Monahan, Walker and Mitchell 2008). Others argue that social scientists legitimately can use scientific concepts to interpret case-specific evidence about how employer organizations operate. In this view, experts may opine about how such case-specific analysis combines with peer-reviewed science establishing causal patterns to strengthen (or weaken) support for finding legal liability (Bielby 2003; Borgida and Fiske 2008; Hart and Secunda 2009). We return to this controversy in our conclusion, after showing how industrial psychology became social authority for Title VII enforcement, promoting law-science co-production.
METHODS AND DATA

Our analytic narrative is based primarily on archival data, including public and non-public primary records from in- and outside government, including expert and social interest organizations, administrative agencies and courts. Histories of Title VII enforcement and of I-O psychology, law reviews, scientific articles, handbooks, symposia, scientific and policy reviews and assessments, and human resource and personnel practice texts ordinarily would be secondary data. For us, many such documents are primary evidence helping establish mutual influence between law and I-O psychology. Many Title VII actors published retrospective accounts. We used these cautiously, knowing that such accounts may be self-serving or may reinterpret events with 20-20 hindsight.

To fill gaps, the first author interviewed a small purposive sample of public and private sector lawyers and social scientists, including I-O psychologists who served as expert witnesses or employees of anti-discrimination agencies.\(^7\) The

\(^7\) This article is part of a broader project on social science in government regulation of equal employment opportunity. About 140 interview hours have been completed with 30 persons. For each, interview length and number depended on whether: he/she had broad or narrow involvement in and knowledge of Title VII or Executive Order 11246 enforcement; was involved over a longer or
document record made it unnecessary to do more interviews. Interviewees included persons from multiple disciplines and organizations who knew about events at different times. To cross validate, we included multiple persons speaking to the same events/issues. Given likely biases from adversary enforcement, we included lawyers and scientists who worked for plaintiff-employees or government enforcers and for defendant-employers. We cross-checked and validated our data. We used event-concurrent documents to anchor memories and query interviewee accounts, and interview data to help locate documents to cross check and extend the record.

**HISTORICAL ANALYSIS**

**Opportunity for Field Intersection**

During Congress’ debate on the proposed 1964 CRA, an Illinois Fair Employment Practices hearing examiner ruled the Motorola Corporation in Chicago discriminated against Leon Myart. Myart, a black man, applied for an assembly line job but was rejected because of poor performance on a multiple choice general ability test. Supporting his order that Motorola hire Myart and cease using the test, the hearing examiner stated “the test itself was unfair to ‘culturally deprived and disadvantaged groups’ because it did not take into account ‘inequalities and differences in environment’” (Graham 1990, p. 149, shorter time; and occupied multiple enforcement roles. Eight interviews are used here and cited where appropriate.

The Tower Amendment  From this controversy emerged Title VII Section 703(h). To safeguard testing, emphasize ability and qualification-based hiring, and guard against making employers liable for past discrimination in education or race inequalities in society, Congress enacted Senator John Tower’s (R-Texas) amendment to the House-passed bill (Graham 1990; Pedriana and Stryker 1997, 2004). It stated that employers could “give and act upon the results of any professionally developed ability test, provided that such test...is not designed, intended or used to discriminate because of race, color, religion, sex or national origin” (BNA 1964, App. A). Bi-partisan co-sponsors of Title VII, Senators Joseph Clark (D-PA) and Clifford Case (R-NJ) explained: “There is no requirement in Title VII that employers abandon bona fide qualification tests where, because of differences in background and education, members of some groups are able to perform better on these tests than members of other groups. An employer may set his qualifications as high as he likes, he may test to determine which applicants have these qualifications, and he may hire, assign, and promote on the basis of test performance” (US Congress 1964, p. 7213, April 8, 1964).

Neither I-O psychology nor psychologists were mentioned when Congress discussed S 703(h), but the words “professionally developed ability test” gave
them an opportunity. Employment tests long predated Title VII and were the acknowledged turf of industrial psychologists (Hale 1982). But when Title VII became law, many employers who had refused to hire blacks or segregated them in low paying, less skilled jobs, enacted education and test requirements for blue and white collar jobs (Ash 1969; Conference Board 1965). When Title VII ended overt discrimination, the predictable result of ostensibly neutral tests was that black applicants, who—especially in the south attended inferior schools and disproportionately lacked a high school degree—were denied jobs at much higher rates than whites (Cooper and Sobel 1969; Pedriana and Stryker 2004). These race inequalities combined with Section 703(h) set the stage for I-O psychologists to provide social authority for Title VII enforcement.

By compromise, Title VII would be enforced mostly by persons filing private claims in federal court. (Graham 1990). The newly created EEOC could only process and conciliate individual complaints. If the employer refused conciliation, the EEOC could issue notice that there existed “reasonable cause” to believe Title VII was violated. But reasonable cause findings provided no redress. The complainant had to file suit in federal court. Pedriana and Stryker (2004) document how, despite minimal enforcement power, the EEOC turned weakness

---

8Psychologists testified on both sides at the Motorola FEPC hearing. In 1966, the Illinois Supreme Court reversed the FEPC’s discrimination finding.
into substantial strength during Title VII’s early years. We show that use of I-O psychology as social authority was a key part of this process.

Scholars across the spectrum agree that fair employment’s traditional goal was color-blind equal treatment (Graham 1990; Pedriana and Stryker 2004). Title VII’s language and legislative history made it hard to interpret in a way that did not require showing employers purposely intended race-based discrimination against individuals (Pedriana and Stryker 1997, 2004). But once the NAACP Legal Defense Fund, (LDF) provided the fledgling EEOC with thousands of complaints, EEOC staff understood that discrimination was “structured, patterned and systemic” so that traditional interpretation could not combat it fully (Pedriana and Stryker 2004, p. 725). I-O psychology used as social authority helped resolve this enforcement dilemma in favor of aggressive enforcement.

Early EEOC priorities included seniority systems and employment tests, but these could not clearly be reached by Title VII because they were apparently race-neutral. EEOC partnered with LDF and both made I-O psychology central to attacking employment tests. The EEOC issued and disseminated informal general guidelines interpreting testing’s legality, used these in reasonable cause findings, and wrote amicus briefs supporting private plaintiffs in federal court. The LDF litigated tests (Pedriana and Stryker 2004; Cooper and Sobel 1969).

*Lawyer-I-O Psychologist Networks Help Create Social Authority* Many credit the legal doctrine of disparate or adverse impact—an effects-based theory
of employment discrimination argued in litigation by the LDF, supported by the EEOC, and endorsed by the Supreme Court in *Griggs v. Duke Power* (401 U.S. 42 [1971])—as the single most important vehicle of early, aggressive anti-discrimination enforcement (Pedriana and Stryker 2004). LDF and EEOC-centered multi-disciplinary networks and enforcement alliances of lawyers and industrial psychologists created this doctrine.

When Title VII became effective, the EEOC organized a small General Counsel’s office whose lawyers decided to monitor all private Title VII cases filed, while liaising with attorneys and organizations bringing the cases (Cashdan 2008; Blumrosen 2004). These included the LDF, directed by Jack Greenberg, and also Richard Sobel, a law professor then at Michigan. With his Columbia Law School colleague George Cooper, Sobel joined with the LDF—a public interest litigation organization—as co-counsels to plaintiffs in *Griggs* (Cashdan 2008; Cooper and Sobel 1969; Belton 2005).

While representing plaintiffs in *Griggs* and other cases, Cooper and Sobel published a *Harvard Law Review* article on legality of testing and seniority systems under Title VII (Cooper and Sobel 1969). Seniority and testing initially were linked in disparate impact’s development but lines of precedent for the two later diverged (Belton 2005). By the time the 1977 Supreme Court re-evaluated and overturned the “present-effects of past discrimination” doctrine applied to seniority, the *Griggs*
effects-based doctrine pertaining to tests and employment practices more generally was institutionalized. Testing was I-O psychology’s turf.

About 15-20% of early Title VII charges involved testing, so effective strategy required understanding how tests worked. Cooper and Sobel (1969, p. 1637) interviewed “industrial psychologists and personnel directors of more than 50 selected businesses,” expressing special appreciation to Drs. Richard Barrett, Brent Baxter and William Enneis, who gave extensive comments and suggestions. Barrett was the LDF testing expert in *Griggs*. Enneis was an early EEOC consultant, then employee who, once on staff, led in constructing and publicizing EEOC’s view of tests. Baxter chaired the 1969 American Psychological Association (APA) Task Force on Employment Testing of Minorities, and was part of an expert psychologist panel asked by EEOC to report on tests and their consequences (EEOC 1966; APA Task Force 1969; Belton 2005; Golub 2005; Rose 2005; Enneis 1967, 1969, 1971).

Employment tests varied from testing relatively concrete, narrow job skills to more abstract, general skills, but many employers relied on general cognitive tests to screen for lower blue collar to upper white collar jobs (Cooper and Sobel 1969). On such tests, including the widely used Wonderlic Personnel Test, the evidence was clear: blacks generally did not perform as well as whites (Legal Implications 1968; Ash 1969; APA Task Force 1969). Citing many empirical studies, Cooper and Sobel (1969, pp. 1640) emphasized *expert consensus* that
such race differences reflected “discrimination, a lack of educational and cultural opportunities, and “cultural separatism’” preventing blacks from succeeding.

Because of tests’ different impact on blacks and whites, there was much writing in the mid-1960s about dangers to equal opportunity if tests were used absent appropriate validation—assessment of whether and degree to which tests reflected real differences in capacity to do jobs for which employers hired (e.g., Cooper and Sobel 1969; Guion 1966; Prentice-Hall Research Staff & Am. Soc. for Personnel Admin 1966; Wallace et al 1966-1967; Krandal 1968; Lockwood 1965; APA Task Force 1969). By 1966, test validation science was developed and consensual enough that the American Psychological Association, American Educational Research Association and Council of Measurement in Education jointly formalized good professional practice, in *Standards for Educational Tests and Psychological Tests and Manuals* (1966).

Cooper and Sobel (1969) communicated across legal and psychological fields, and their partnership with the LDF helped drive LDF strategy. EEOC attorneys developed relationships with Richard Sobel and LDF attorneys (Cashdan 2008). Once the EEOC had recruited economist Phyllis Wallace to organize and direct a sub-unit of its research office to focus on *empirical*, rather than legal research, Wallace hired William Enneis—who had advised Cooper and Sobel—to come in-house as EEOC chief psychologist (Blumrosen 2004, Golub 2005, 2008, Copus 2008). Wallace also recruited the expert panel of psychologists on which
Brent Baxter served, to report on testing to the EEOC. Baxter’s fellow panel psychologists were Dr. Richard Docter, Associate Administrative Officer, State and Professional Affairs, APA, and Dr. George Elias, Associate Professor of Psychology and Education at Assumption College.

Alfred Blumrosen, first EEOC Director of Conciliations, recalled that the panel established “technical foundations for the [EEOC] testing guidelines” (Blumrosen 2004, p. 1). Because Title VII authorized “professionally developed ability tests” provided they did not discriminate, but did not define what such a test was, “it was sensible to go to the Professional Association to find out what they thought a professional ability test was. So Charlie [EEOC Office of Research Director Charles Markham] asked Phyllis to get together a group of industrial psychologists who were knowledgeable about this subject – test validation—to come to Washington and give us the benefit of their understanding” (Blumrosen 2005, p. 41). In a one-day meeting Wallace organized at the EEOC, the panel expressed “unanimity of opinion” (Blumrosen 2005, p. 4; EEOC 1966). Cautioning that tests must be developed and applied professionally, the panel said a test had to be proper for each use. Advice drew on then scientific consensus around situational specificity, the theory that a test must be validated for each specific situation, including job, employer and population, for which it is used (Ghiselli 1959; Guion 1966; Landy 2003).
The panel also recommended that the EEOC adopt the APA’s (1966) Standards. A test’s “professional nature” involved determining job requirements through “careful job analysis,” “selection and/or development of instruments to measure…critically important abilities,” administering the instruments to job applicants or current employees, “identify[ing] or develop[ing] effective job performance (the criteria), [and] compare[ing] individual employee scores with their criterion performance.” Unless test scores “correlated (more than chance would indicate)” with reliable measures of job performance, the test was not valid: “Tests should be selected based on validation against performance requirements of the job, that is, criterion related validity.” The same test could be expected to have differing validities for each “job situation,” in which it were used. Likewise, employment tests should be revalidated frequently to keep validity evidence up-to-date (EEOC 1966, p. 96).

Especially important, validity was established relative to the population used for norming test scores. Acknowledging that cultural factors affected how well different groups performed, the EEOC panel also stated that culture could “affect test reliability and validity for that segment of the population whose culture differs appreciably from the normative group” (EEOC 1966, p. 96).
short, the panel expected differential validity and differential prediction. Given that validation had been done almost exclusively on whites, the panel urged the EEOC to push for validation on minorities, “using measures of cultural background as moderator variables” (EEOC 1966, p. 96).

In August, 1966, the EEOC adopted its *Testing Guidelines*, importing into them wholesale the expert panel’s recommendations. According to precedent, courts were to defer to regulations issued by federal agencies to which Congress gave formal administrative rule-making power. The EEOC had no such power; it issued early interpretations of tests as informal interpretive guidelines (Pedriana and

---

9 The two terms sometimes are used inter-changeably because test validation rests heavily on predictive/criterion validity. Differential validity refers to differences in size of the correlation coefficient between a predictor and a criterion; differential prediction refers to needing a different regression equation across different populations to predict the criterion. A differential validation study typically would look for significant differences in correlation size *and* significant differences in prediction equations for multiple groups. Both differential validity and prediction are aspects of “test fairness.” Both measure bias, but experts now usually emphasize differential prediction because it answers whether and how an equation computed for one group would over- or under-predict performance for another group (Hartigan and Wigdor 1989; US Civil Rights Commission 1993).
Stryker 2004). Without certainty that courts would defer as they did to formal regulations, the EEOC nonetheless relied on the Guidelines to issue many case-specific reasonable cause findings stating employers had violated Title VII, while the LDF relied on the Guidelines to attack tests in court (Pedriana and Stryker 2004; Cooper and Sobel 1969; Belton 2005).

According to the Guidelines, a test did not automatically qualify as “professionally developed” under Section 703(h) just because it was designed by someone “claiming expertise in test preparation.” “The Commission…interprets ‘professionally developed ability test’ to mean a test which fairly measures the knowledge or skills required by the particular job or class of jobs which the applicant seeks, or which fairly affords the employer a chance to measure the applicant’s ability to perform a particular job or class of jobs.” Prior discriminatory employers who now tested were told that “use of tests in such circumstances will be scrutinized carefully” (EEOC 1966, p. 95).

Employment tests did not always adversely affect blacks; some early cases used tests to challenge denial of jobs to blacks who passed tests (Cooper and Sobel 1969, p. 1664, n. 124). Here tests benefitted minorities by limiting subjective decision-making. I-O psychologists of the 1960s generally assumed that well-designed tests promoted equal opportunity by removing unfettered employer subjectivity and race bias (APA Task Force 1969). Thus, I-O psychologists could promote EEO norms while also expanding professional turf (Ash 1966, p. 803).
But the *Guidelines* intended to guard against inappropriate testing excluding qualified minorities, giving employers a “scientifically sound, industrially proven and equitable basis for matching manpower requirements with human aptitudes and abilities.” Tests used should be based on “specific job-related criteria,” there should be “comparison of test performance versus job performance,” and an opportunity to retest those who got more training or experience. Tests must be validated for minorities because “only a test that has been validated for minorities can be assumed to be free of inadvertent bias” (EEOC 1996, p. 95).  

---

10 Psychologists agreed that for tests to be equally valid across race/ethnic groups, each needed equal exposure to tested skills/knowledge. In 1966, few finished empirical studies of differential validation by race existed, but many were ongoing. Multiple studies soon showed differential validity Kirkpatrick and Tenopyr 1967; Ruda and Albright 1968), convincing psychologists, including those consulted by the EEOC, that differential prediction would be born out widely. Many psychologists urged the legal presumption of differential prediction unless the contrary were proved in the case. But some thought that because studies had not yet shown the exact extent of differential prediction, plaintiffs should have to prove it. EEOC psychologist Enneis (1971) believed that differential validation studies lay at equal opportunity’s heart. Ralph Berdie (1965, p. 146), writing on behalf of the Ad Hoc Committee on the Social Impact of Psychological
By the time Cooper and Sobel (1969) wrote their article, they too would rely on the *EEOC Guidelines* to support their position on testing. For them, legality under Title VII rested on a tripartite analysis. Did the test have adverse impact on minority group members? If it did, could the test be justified by business needs? If the test had adverse impact but could be justified by business needs, was there an alternative selection device meeting business needs with less adverse impact?

Not all Cooper and Sobel’s (1969) arguments became court doctrine. But disparate impact adopted by *Griggs* conformed to their analysis. Plaintiffs using disparate impact theory would bear the burden of showing adverse impact of the test or other selection device, establishing a prima facie case of discrimination; defendants would bear the burden of proving the “business necessity” defense. If

Assessment, said that, given tests were expected to have differential validity for minorities and whites, employers typically used tests in ways that “encouraged” discrimination. The APA Task Force on Employment Testing of Minorities (1969, p. 641) noted that differential prediction was a key implication of extant science, but that more research was needed.
defendants succeeded, plaintiffs would bear the burden of proving the existence of
a “lesser discriminatory alternative” (Stryker 2001).11

Legality’s lynchpin would be validation. Validated tests that substantially
(and statistically significantly) correlated with job performance would meet the
business need/necessity test and not be ruled discriminatory even if they had
adverse impact on minorities (Stryker 2001). A test not properly evaluated for
employer’s job and workforce could not qualify as professionally developed within
Section 703(h). Litigation ported the Cooper-Sobel proof model into case law.

*Industrial Psychology as Social Authority in Griggs* Notwithstanding
consensus on test validation’s importance for equal opportunity, disparate impact’s
success was not assured; a large array of standard legal-interpretive techniques
worked against court adoption. Testing cases leading to *Griggs* were cases of first
impression *under Title VII*, but prior dominant legal understanding presumed that
non-discrimination was color-blind equal treatment. Even though Title VII did not
define discrimination explicitly, it was easy to argue that Title VII’s language and
legislative history showed Congress required purposeful racial animus/bad intent

---

11 Boundaries between parts two and three may blur because courts may want
employer to show it investigated potential less discriminatory alternatives before
granting that business necessity is met. Once employer shows business necessity,
plaintiff may still win by proving there was a less discriminatory alternative.
for employers to be found in violation (Pedriana and Stryker 1997, 2004).

Employers insisted that Section 703(h) protected tests if tests were “color blind and not used purposely to discriminate” (Pedriana and Stryker 2004, p. 733).

Because the most obvious legal strategy worked for defendants but not for plaintiffs, plaintiffs’ lawyers had an incentive to find and use an alternative, and they relied on social science to construct an effects-based legal theory. That *consensus* existed among testing psychologists about applicable scientific concepts and their import, and that this unequivocally favored minorities claiming discrimination because of tests’ adverse impact, increased incentive on plaintiffs’ lawyers to treat I-O psychology as *social authority*, providing key rationale for disparate impact. It also increased what were still small expected odds of plaintiffs’ success.

Law enforcers reasonably could have seen disparate impact as hopeless and given up—there is evidence of EEOC and LDF equivocation. But LDF lawyers did not give up, nor did the EEOC give up support (Pedriana and Stryker 2004; Belton 2005; Graham 1990, p. 385 and n. 79, 80, p. 551). The LDF hedged its bets and “decided to make alternative arguments in the District Court in *Griggs*,” maintaining that intent was *not* required to prove Title VII violation, but that Duke Power could be found in violation *even if* the Court decided only intent-based liability could pass muster (Cooper and Sobel 1969; Belton 2005, p. 448).
Griggs involved tests of general cognitive ability. Prior to Title VII, Duke Power segregated facilities and jobs, restricting blacks to its lowest “labor” department. With Title VII’s advent, Duke Power desegregated, adopted formal non-discrimination, but required applicants and employees to pass the Wonderlic Personnel Test and Bennett Mechanical Comprehension Test—or have a high school diploma (Belton 2005; Pedriana and Stryker 2004). Only incumbent employees in the labor department who passed the Wonderlic and Bennett tests or had a high school diploma could transfer out to higher ranked departments.

Plaintiffs challenged the tests and diploma, arguing that while these did not discriminate overtly, they discriminated because they had adverse impact and, contrary to EEOC Guidelines, were not job related. The 1960 census showed 12 percent of North Carolina blacks graduated high school, compared to 34% of whites. In its amicus brief for trial, the EEOC emphasized its study for another case, showing that 58% of whites but 6% of blacks passed the Wonderlic and Bennett tests (420 F.2nd 1225 [1970], 1233).

At trial, the LDF relied on Richard Barrett to reiterate the position of the Guidelines’ and the consensus among I-O psychologists. Duke Power’s expert was University of North Carolina industrial psychology professor Dannie Moffie. At trial Moffie agreed that professional standards required that tests be valid and reliable and conceded that neither the Wonderlic nor Bennett tests measured capacity to perform a specific job or job category (420 F. 2d 1255, 1244-45). But

The District Court found for Duke Power, ruling that plaintiffs failed to carry their burden of proving the company intentionally discriminated and that an effect-based concept of discrimination was inappropriate. The key issue on appeal was whether Section 703(h) required tests to be job-related? Both the EEOC and Justice Department argued that it did. But the Fourth Circuit dismissed the EEOC Guidelines out of hand (420 F. 2nd 1255, 1234).

The LDF was cheered somewhat by Judge Simon Sobeloff’s dissent. He found the EEOC interpretation reasonable, informed by relevant expertise, entitled to deference and “well supported by [Title VII’s] legislative history” (420 F. 2nd 1225, 1243-1244). No other interpretation could be effective in realizing Title VII’s anti-discrimination goals (420 F. 2nd 1225, 1241).

Judge Sobeloff’s opinion convinced the LDF to seek Supreme Court review despite some concern that Griggs might not be the right case at the right time (Belton 2005, p. 453). The gamble paid off. In a unanimous 8-0 ruling, the Supreme Court reversed the Fourth Circuit, following Judge Sobeloff’s dissent.
In *Griggs*, a case its LDF litigators put on par with *Brown v. Board of Education* (Belton 2005; Rose 1989), the Supreme Court unequivocally provided an effects-based interpretation of Title VII, endorsed the job-relatedness concept of “professionally developed tests,” and applied job-relatedness broadly to all employment practices (401 U.S. 424 [1971], 429-32). The Supreme Court “rehabilitated and deferred” to the EEOC Guidelines and decoupled disparate impact from the “present-effects” theory used for seniority (Pedriana and Stryker 2004, p. 738, Rose 1989, Belton 2005).

*Griggs*’ addition of effects-based enforcement to supplement intent-based liability was crucial to attacking institutionalized employment practices that systematically harmed minorities (Pedriana and Stryker 2004). Because effects-based enforcement is necessary (though not sufficient) to maximize effectiveness of legislation that, like Title VII, aims to reduce inequality (Stryker 2007), the practical import of testing psychology was large. Its *social authority* role in *Griggs* helped secure institutionalization of I-O psychology in later case law.

LDF expert Barrett testified that Duke Power’s tests had adverse impact and were not job-related. The company did not challenge these expert-provided *social facts*; the District Court accepted them as adjudicated fact. But because the Court endorsed only intent-based liability, these facts were not significant for its rule-making. Endorsing effects-based jurisprudence and job-relatedness, the Supreme Court restored legal significance to Barrett’s testimony. The Court
followed the *EEOC Guidelines*, elevating them to *social authority* for the legal meaning of ‘professionally developed: absent job-relatedness, tests were not professionally developed.

Once adopted as social authority in *Griggs*, I-O psychology could become part of a precedent-based legal argumentation frame. *Albemarle v. Moody* (422 US 405 [1975]) reinforced and further specified *Griggs*, ensuring the ongoing presence of I-O psychology in Title VII litigation. This institutionalized presence, in turn, was an essential condition for Title VII-I-O psychology co-production.

*Cementing the Foothold; Insuring Co-Production* In *Albemarle*, the Supreme Court cemented and further specified adverse impact to ensure I-O psychology’s institutionalization as social authority, triggering I-O psychology’s institutionalized role of providing social facts to aid judicial fact-finding. This opened a niche for I-O psychologists in overlapping legal and scientific institutional fields. Once I-O psychology routinely helped resolve questions of legal liability, legal issues likewise shaped scientific questions and research.

*Albemarle*, like *Griggs*, was an LDF-litigated class action. Albemarle Paper Company used a high school degree and two general ability tests, the Wonderlic Test and Revised Beta Examination, to screen. Barrett again served as LDF’s expert. Because of *Griggs*, Albemarle knew four months before trial that it must show job-relatedness. It tried to validate its tests by hiring a consultant who
visited the company for half a day and “developed a criterion validation strategy but did not supervise the actual work” (Gutman 2005, p. 28).

Satisfied, the District Court did not enjoin Albemarle’s tests. The Circuit Court reversed, leaving the Supreme Court with the task of further interpreting job-relatedness under *Griggs*. The Supreme Court said that showing a test’s job-relatedness required “a statistical correlation between test scores and job performance” (Player 1988, pp. 367-68).

The Court held that Albemarle’s validation was materially defective when stacked up against the *Guidelines*. Tests failed to correlate significantly with some job progression lines. Supervisors “rank[ed] employees by an “extremely vague ‘standard’” “fatally open to divergent interpretations” (405 U.S. 405, 433). Albemarle focused on job groups “near the top” of company job ladders; the Court “endors[ed]” the *Guidelines*’ “sensible” view that, when job progression was not almost automatic, applicants must be judged using requirements of entry or near-entry jobs (422 U.S. 405, 434). The Court did not invoke the scientific term *situational specificity*, but failure to adhere to this idea drove Court concern that the test score-entry level job relationship was not validated.

Finally, Albemarle’s study “dealt only with job-experienced white workers,” but “tests…were given to new job applicants,” who were “younger, largely inexperienced and in many instances nonwhite” (422 U.S. 405, 434). The Supreme Court invoked the *Guidelines*’ differential validation requirement for
minorities when technically feasible, and also the recently updated APA
Standards (1974) deeming it “essential” that validation be done for persons with
the same characteristics as applicants. The Court thought Albemarle could have
studied validity by race at least for lower level jobs in which it employed blacks,
and questioned whether validity would hold for much of the applicant pool.

Thus, Albemarle cemented and built on Griggs, defining legal job-
relatedness through the scientific concept of test validation. With scientific
concept transposed to legal requirement, the stage was set for attorneys in later
cases to use I-O psychologists to argue that professional, scientific standards for
test validation had—or had not—been met. Based on the EEOC Guidelines and
to some extent on allied APA Standards, the Supreme Court of the early to mid-
1970s embedded in formative Title VII enforcement the broadly consensual
scientific assumption of situational specificity, and the not-quite-so-consensual
but logically allied assumption of differential validity.

Griggs and Albemarle remain foundational. By September 2008, Griggs
had been cited in 2,926 later opinions, Albemarle in 2,650.12 Barrett’s success led

12 Not all cases invoked Griggs or Albemarle for test validation. In Wards Cove v.
Atonio (1989), the Supreme Court greatly reduced defendant’s burden to rebut
plaintiff’s prima facie case. But the 1991 CRA restored proof burdens to those in
Griggs and courts continue to wrestle with test validation. In Washington v.
him to testify in 100+ more cases (Barrett 1998). *Albemarle* signaled that the co-production of I-O psychology and Title VII had begun.

**Co-Production of Title VII Enforcement and I-O Psychology**

Given scientific consensus around situational specificity, it made legal sense that employment tests be validated against performance in jobs for which employees were sought. If un-validated tests could screen out minorities, this violated equal opportunity. The logically derivative theory of differential validity by race was not as firmly rooted. But given situational specificity, it made good scientific sense for the 1960s EEOC and the 1969 APA Task Force to promote differential validation. Given United States’ racial history, it also made good legal and moral sense. But this apparently tension-free alignment of law and science in foundational testing law would not last.

*I-O Psychology Institutionalized as Social Fact:  Post-Albemarle*, I-O psychologists became routine in Title VII testing litigation. Congress’ 1972 amendments extended Title VII to state and local government. The Employment Litigation Section of the Justice Department’s Civil Rights Division (ELS) used *Griggs* to challenge “police, fire and sheriff’s departments that used...tests of cognitive ability that had disparate impact on minority applicants” (Ugelow 2005, p. 125). The Supreme Court ruled that proving discrimination in equal protection cases required proving intent.
Longtime ELS Section Chief Richard Ugelow said: “The most important expert witness in a testing case was the I-O psychologist …there were a core of four or five I-Os who were regularly…retained by the DoJ [Department of Justice] …another three or four experts….would be hired by the ELS at any given time” (Ugelow 2005, p. 47). The ELS also provided its lawyers with in-house training in test development, use and validation (Ugelow 2005).

*Griggs and Albemarle* guaranteed that later litigation would involve issues calling for I-O psychologists. Among these were: how large a correlation between test scores and job performance would be required to satisfy employer’s business necessity burden; what constituted adequate job analysis; how to set appropriate cut-off scores for employee selection; whether and what kinds of banding would be allowed to set cut-off scores; whether and when pass-fail vs. ranking procedures were preferred; whether and how validation could be generalized; whether or not differential prediction by race, ethnicity or gender was likely; legitimacy of ranking candidates relative to adjusted mean scores computed within race, ethnic or gender groups (i.e., race or gender-normed tests); what defined test fairness; how to design alternative procedures with less adverse impact; and the import of different procedures for adverse impact reduction and legitimate business interests, including productivity and safety (Kehoe and Olsen 2005; Ugelow 2005; Hartigan
and Wigdor 1989; Player 1988; US Civil Rights Commission 1993; Landy 2003; Seymour 1988).\(^{13}\)

All these issues required using I-O psychology as social fact, to help trial judges adjudicate what had, or had not, been done in testing. I-O psychologists provided expert testimony or consulting based upon their own “original scientific research” or their “education and experience,” and they “critique[ed] testimony” offered by the opposing party’s experts (Thornton and Wingate 2005, p. 180). Indeed, because disparate impact lawsuits typically are class actions involving

\(^{13}\) Contribution of I-O psychology to litigating these issues merits more research.

*Albemarle* set tough business necessity standards but *New York City Transit Authority v. Beazer* (1979) adopted a more relaxed approach. But *Beazer* did not involve tests, so testing remained guided by *Albemarle*. For unskilled jobs posing little risk from poor performance, courts require higher statistically significant test-score/job performance correlations. For high skill jobs in which poor performance is risky—e.g., flight engineers—courts lighten employers’ burden (Stryker 2001). Disparate impact became highly politicized in the 1980s-early 1990s, resulting in legislation to undo the Supreme Court’s *Wards Cove* ruling changing the allocation of proof burdens in disparate impact cases (see n. 12). The 1991 CRA also outlawed race norming pre-employment tests (US Civil Rights Commission 1993).
potentially large liability costs, both plaintiffs and defendants often hire experts to help develop and present social facts to the court (Stryker 2001; Ugelow 2005, Rose 1989). Experts help attorneys prepare for cross examination, and even when cases are settled before trial on the merits, I-O psychologists typically play a large role.

In a discrimination case the psychologist can help the lawyer virtually from day one…The psychologist can prove invaluable in formulating a discovery plan. She will know what sort of documentation should exist, what particular tasks should have been performed, and what potential holes are in your opponent’s case. Similarly, there is no one better to help prepare for a deposition of the I-O psychologist than another I-O psychologist…To effectively examine any expert, a lawyer needs to understand the expert’s unique language as much as possible, know when he doesn’t, and really know where and how to score points and stick to those areas. Without your psychologist, this is difficult if not impossible (Sohn 2005, pp. 499-500).

Because novices may not know what facts about tests should be introduced, sometimes attorneys will ask I-O psychologists to “summarize [for them or the court] findings from past scientific studies” (Thornton and Wingate 2005, p. 180). This corresponds to Monahan and Walker’s (2010a) social framework role, that is, general context to interpret legal meaning of facts about a specific employer’s test.
Finally, with I-O psychology embedded as *social authority*, lawyers need I-O psychologists to help determine what case law means. It is impossible to understand precedent and effectively make analogies without I-O psychologists. There are myriad [testing] cases…But [beyond such seminal cases as *Griggs* and *Albemarle*], how would the lawyer know which of these cases is apposite? Is a *cut score* the same as a *critical score*? The same as a *pass score*? Is *generalizability* the same as *reliability*? The same as *consistency*? The same as *accuracy*? It is the I-O psychologist who leads the lawyer through the forest of technical terms so that the lawyer may collect, use and truly understand the relevant case law…most Title VII cases do not describe in detail the psychometric analyses required for a court to proclaim whether a test is valid or violates Title VII. One must dig beneath the veneer of the opinion for these inferences, and this is when it is valuable to have the I-O psychologist digging along with the lawyer…[I-O psychologists help by] offering input in figuring out what legal standard may apply… (Sohn 2005, pp. 496-97, emphases in original).

Thus, Title VII provides I-O psychologists with a new market niche, while allowing them to help shape policy by providing social facts in litigation. Many I-O psychologists do little or no expert witnessing because they find adversary work uncomfortable or disruptive to their science-based priorities and reputations...
(Schmidt 2007; Rose 2005). Others—like Richard Barrett or the small group of ELS experts—do much of it (Ugelow 2005; Schmidt 2007)

**I-O Psychology Institutionalized in Administrative Rule-Making**

Following their role in constructing the original *EEOC Testing Guidelines*, I-O psychologists helped write other government guidelines and formal regulations. In 1969, the Labor Department’s Office of Federal Contract Compliance (OFCC), charged with enforcing Executive Order 11246 mandating affirmative action in government contracting, involved I-O psychologists in creating testing regulations similar to those of the EEOC. The OFCC, which had power to issue formal regulations, issued revised rules in 1971 (Guion 1997). When the 1972 EEO Act gave the EEOC prosecuting power, the Civil Service Commission (CSC, later the Office of Personnel Management, OPM), and Justice Department also had regulations. Differences among agencies reduced rationality and increased employer complaints, so the 1972 Act established an inter-agency Equal Employment Opportunity Coordinating Council (EEOCC) to negotiate uniform rules. In 1978, after the Carter Administration kick-started deliberations, the four agencies, all of which put I-O psychologists on the EEOCC, adopted concurrently the *Uniform Guidelines on Employee Selection Procedures* (Jeanneret 2005; Rose
Professional organizations representing I-O psychologists, including APA and SIOP, provided input and advice (Camara 1996, Rose 1988, 2005; 42 Federal Register 66542 [1977]; 43 Federal Register 38289, 38292-38293 [1978]). Except for a 1981 provision mandating OMB control numbers for record keeping, the 1978 Guidelines are un-amended, but remain in force, setting the stage for mismatch between developing science and static testing regulations.

**Title VII’s Impact on I-O Psychology**

Historians of I-O psychology concur that Title VII enhanced field growth, advancing theory, methods and research (Koppes 2003; Vinchur 2007; Katzell and Austin 1992; Lowenberg and Konrad 1998; Zickar and Gibby 2007). SIOP past President Angelo DeNisi said Title VII “began a tidal wave of work on test validation, beginning with discussion of criterion issues and culminating in…work on validity generalization …the race was on to discover forces that led to lower than desired validities and ways to validate tests more efficiently” (quoted in Koppes 2003, p. 380).

---

14 William Enneis joined Alfred Blumrosen and EEOC Deputy Director Al Golub on the staff committee for the EEOC (Rose 2005; Blumrosen 2005; Enneis 1967, 1971). The EEOC hired Richard Barrett as a temporary employee to help draft the 1978 Guidelines and 1979 Questions and Answers (Barrett 1998, p. xi.) I-O psychologist Donald Schwartz later replaced Enneis (Rose 1988, 2005). The main CSC/OPM psychologist was research head Bill Gorham.
I-O psychologists thought Title VII important for their profession. In 1965, the APA’s annual meeting had a symposium on legal issues confronting psychologists and society. Lawyers presumed testing would be a minor issue, but I-O psychologists thought business use of tests could fall prey to extensive litigation (Viteles 1964; Ash 1966). Notwithstanding some concern that litigation could “cripple” I-O psychologists working for industry, many industrial psychologists perceived opportunity rather than constraint (Berdie 1965). Philip Ash (1966, p. 803), then of Inland Steel, told his APA audience:

Psychologists face, in the matter of civil rights, not a threat to their instruments but a challenge to their talents: to serve as resource people and advisers, to organize and administer programs. To make effective use of our tools and to do research to clarify the problem of unintentional but nonetheless insidious discrimination through the adverse impact of tests that are unrelated to job performance…We have always deplored the use of un-validated instruments; the time has come to remedy the all-too common practice of taking a convenient brief intelligence test off the shelf and using it for all jobs, without local norms, without criterion relationships, and in fact frequently without any demonstrable relevance to the selection program at hand.
Ralph Berdie (1965), writing for The Ad Hoc Committee on Social Impact of Psychological Tests, Ash (1966) and others emphasized studying differential validation for whites and minorities. Some declared it unethical to use tests unless the tests were validated for the appropriate sub-group (Krug 1966).

I-O psychologists answered the call in force. In the 1960s-1970s, much research focused on determining bias in employment tests (Zickar and Gibby 2007; Katzell and Austin 1992). I-O psychologists in academia, government and industry researched differential prediction by race across jobs, companies and industries (e.g., Krug 1966; Kirkpatrick and Tenopyr 1967; Ruda and Albright 1968; Kirkpatrick et al 1968; Guion 1966). Validation became more involved, often including multiple predictors and job performance criteria.

Undergirding how tests related to EEO was scientific consensus that “validity [was] situation specific” (Guion 1976, p. 783). Reinforcing this “tenet of orthodoxy” (Guion 1976, p. 783) were Edwin Ghiselli’s (1955, 1966) reviews of variability in validation results across jobs and employers. By the 1970s, situational specificity was so enshrined that it often was seen as fact, not theory (Copus 2006; Schmidt and Hunter 2003).

But in 1977, I-O psychologists Frank Schmidt and John Hunter published a path breaking frontal assault on it. In “Development of a General Solution to the Problem of Validity Generalization” (1977), they reviewed published and unpublished research on validity of ability tests, reinterpreting variable validity in
light of measurement and sampling error. Error in criterion (job) performance would lower observed correlation between test scores and job performance. Sampling error would lead to random variation in outcomes. These error sources could lead to apparent invalidity of a test that was valid and to apparent variability in a test’s predictive validity, even if validity were constant across all studies.

In short, findings of invalidity/differential validity could be artifacts of sampling and measurement error. Using statistical knowledge about estimating the impact of sampling error and unreliable measures on findings, Schmidt and Hunter (1977) corrected results from prior studies for these errors and conducted “meta-analysis,” in which variables and results of prior studies provided data for cumulative assessment. Social scientists still argue about appropriate corrections for sampling and measurement errors and whether studies meet conditions for sensible cumulation. But the basic technique of meta-analysis now is “well described, accepted and accompanied by little controversy” among I-O psychologists (Landy 2003, p. 156).

Looking back, Schmidt and Hunter (1998, p. 264) said their 1977 findings showed that most differences across prior studies were artifacts, that “variability in validity was ….small or zero across settings for the same type of job,” and “small across different kinds of jobs.” These findings, if accepted as true, had revolutionary implications, because “the traditional belief in situational specificity would be seen to be erroneous and the conclusion would be that validity findings
can be generalized” (Schmidt 1992, p. 1177). The special meta-analysis application pioneered by Schmidt and Hunter (1977) is called validity generalization—VG for short.

Schmidt and Hunter’s techniques and conclusions were contested. Some who staked careers on situational specificity or trained when it was consensual maintained its truth until they died (Landy 2003; Schmidt 2007). A 1984 Annual Review article expressed substantial concern about VG (Zedeck and Cascio 1984). Landy’s (2003) dispassionate review suggests VG was more controversial among I-O psychologists than it otherwise might have been because Schmidt and Hunter sometimes underplayed uncertainty on assumptions made to correct for statistical error, gave mental ability primary or exclusive status in tests and exaggerated productivity gains from testing and rank ordering applicants on mental ability. But, “even without the Schmidt and Hunter style, VG would have been a new and uncomfortable doctrine for many to grasp, after so many years of situationism” (Landy 2003, p. 170). Some disagreement between VG’s strongest and more cautious advocates remains, but the 1977 article was “the beginning of the end for the doctrine of situational specificity” (Copus 2006, p. 11).

Unintended Consequence: VG is the New Scientific Consensus

Revolutionary science must be disseminated widely and subjected to rigorous critique and further empirical testing before gaining wide acceptance. By 1987 “more than 200 meta-analyses examining the relationship between cognitive
ability and performance had been completed [and] provided overwhelming evidence that cognitive ability tests generalized across virtually all jobs and settings” (Rothstein 2003, p. 125). Schmidt and Hunter (1981, p. 1132-33) reported their meta-analysis of 500 criterion validation studies showing that “cognitive abilities are valid for all jobs…there is no empirical basis for requiring separate validity studies for each job.”

Looking back, psychologists wrote of VG’s “enormous,” “revolutionary” impact (Rothstein 2003, p. 126; Sackett 2003, p. 11). Cumulative empirical support for VG “turned on its head” “the paradigm of situational specificity” (Rothstein 2003, p. 126). Over time, “the basic propositions of VG” became accepted scientific wisdom (Landy 2003, p. 174; Murphy 2003; Jeanneret 2005).

The APA Standards and those by the APA’s Division of Industrial Psychology (after 1982, SIOP) testify to evolving consensus. The 1974 Standards were updated in 1985 and 1999 (APA 1985, 1999). In 1975, The Division of Industrial and Organization Psychology published its first standards: Principles for the Validation and Use of Personnel Selection Procedures. There were three later editions (Division of Industrial and Organization Psychology 1980; SIOP 1987, 2003). Both the Standards and Principles are formal policies of the profession

In the late 1980s, the National Academy of Sciences convened an expert interdisciplinary committee to assess validity generalization used by the VG-GATB Referral System. The GATB—General Aptitude Test Battery—was a federally sponsored employment test matching job seekers to employer requests for applicants. The GATB-VG referral system used VG to extrapolate from empirically estimated GATB validities predicting performance in 500 jobs to all 12,000 jobs in the US economy. The Labor Department promoted GATB for all jobs and the US Employment Service encouraged states to give the GATB to all registrants, using test scores to fill job orders (Hartigan and Wigdor 1989). In its 1989 report, the NAS committee “accept[ed] the general thesis of validity

---

The *APA Standards and Division 14/SIOP Principles* are intended to be consistent on test validity, reliability and bias. Both are research-based and express professional consensus. The *Principles* give greater guidance on validation in employment settings. The *Standards* have broader scope, applying also to measurement in education testing, program evaluation and public policy (Jeanneret 2005, p. 69).
generalization—that the results of validity studies can be generalized to many jobs not actually studied” (Hartigan and Wigdor 1989, p. 8).\textsuperscript{16}

**What about Differential Validation?** What did situational specificity’s demise mean for differential validity by race? In 1981, Schmidt and Hunter (1981, pp. 1128, 1131) reported: “[c]ognitive ability tests are equally valid for minority and majority applicants and…do not underestimate the expected job performance of minority groups… evidence for single group validity by race does not occur any more frequently in samples than one would expect by chance.” As research cumulated, most of it supported Schmidt-Hunter (Hackel 1986). By 1997, a scientist who in 1966 had assumed differential validity reviewed the cumulative evidence and concluded it “rare” to find differential validity under-predicting job performance for blacks (Guion 1966, 1997, p. 439).

\textsuperscript{16} The Committee sided with VG’s more cautious advocates, indicating that GATB validities for supervisor ratings were “in the range of .2 to .4… although we have seen no evidence…that the test battery is a valid predictor of all 12,000 jobs in the economy…we urge a cautious approach of generalizing validities only to appropriately similar jobs (Hartigan and Wigdor 1989, p. 8). GATB scores were useful to screen applicants but predictive power was not strong enough to warrant using the test alone to fill jobs. The NAS also urged caution in estimating productivity gains using the GATB.
Official SIOP pronouncements echoed Guion (1997). In 1985, SIOP President Benjamin Schneider told a House Committee (1985, p. 76) that research failed to support that “selection procedures work[ed] differently for persons of different racial subgroups.” The 1987 SIOP Principles (p. 18) “[found] little convincing evidence that well constructed and competently administered tests are more valid predictors for one population sub-group than another.” The 2003 Principles (p. 25) stated: “Predictive bias has been examined extensively in the cognitive ability domain. For White-African-American and White-Hispanic comparisons,” predictive bias is “rarely found.”

The 1989 National Academy Committee also reviewed evidence on differential prediction by race, reanalyzing 78 of 200 Employment Service GATB studies including at least 50 blacks and 50 whites. Average correlations between test scores and supervisor ratings were somewhat less for blacks than whites, .12 for the former, .19 for the latter.¹⁷ The best prediction equation for blacks was a bit different than for whites. But “use of a single formula…would not be biased against black applicants: if anything, it would slightly over-predict their performance, particularly in the higher score ranges” (Hartigan and Wigdor 1989, p. 6).  

¹⁷ The correlations of .2-.4 (see n. 16) correcting for sampling error were based on all 750 criterion-related studies of the GATB (Hartigan and Wigdor 1989).
There is an important caveat. If measures of the criterion—job performance—were biased systematically to under-estimate performance of minorities relative to whites, concluding that test scores over-predicted job performance for minorities would be suspect. The Committee trod gingerly, noting that slight over-prediction for blacks “need[ed] to be treated with some caution [because] there may be bias against blacks in the primary criterion used in these studies—supervisor ratings. Usually supervisors were white. There is some empirical evidence …that supervisors…favor employees of their own race” (Hartigan and Wigdor 1989, p. 6).18

In sum, scientific consensus today endorses VG, though not the strongest statements made by its most avid proponents. Mental ability is an agreed upon valid predictor of job performance, but there are others, including work sample tests measuring job-specific skills, that have as high or higher validities as cognitive tests and typically also have less adverse impact (Schmidt and Hunter 1998, Table 1, p. 18). The Committee also emphasized that without score adjustments, false rejection would be higher for minority applicants: majority applicants would benefit from higher false acceptance. This is because prediction error combines with average group differences in test scores (the average is lower for minority groups), and modest validities of scores for supervisor ratings, so there is less difference in average job performance between majority and minority job applicants than between respective test performances (Hartigan and Wigdor 1989, p. 6-7).
The cumulative record assessed by the National Academy and researchers with varying stakes in falsifying the differential validation hypothesis suggests that differential prediction of cognitive tests working to under-predict minority job performance is the exception rather than the rule.

**VG, Uniform Guidelines and Precedent: Unintended Law/Science Mismatch?**

With 20-20 hindsight, the *Uniform Guidelines* emerged just as the contest over situational specificity began.\(^{19}\) VG consigned situational specificity to science’s dustbin, creating what many see as serious Title VII-I-O psychology mismatch. Even if the *Guidelines* are seen to provide an opening for VG, extant case law rejects it.

*Guidelines* section 5C explicitly signals consistency with the 1974 APA *Standards*, adopting the presumption of situational specificity (43 *Federal Register*, 38298). Because the *Standards* also stated it essential that validation be done for persons with the same characteristics as job applicants taking the tests, differential validation also is a *Guidelines* premise. The *Guidelines* require “conduct[ing] a

\(^{19}\) The *Guidelines* implicitly signal tensions of law-science hybridization, proclaiming adherence to “standards of the psychological profession,” but acknowledging that “court decisions, the previously issued guidelines of the agencies, and the practical experience of the agencies” also provided building blocks for the *Guidelines* (43 *Federal Register* 38289, 38292, [1978]).
fairness analysis whenever disparate impact is found. Unfairness occurs when lower minority scores on a selection procedure are not reflected in lower scores on the criterion or index of job performance” (Jeanneret 2005, p. 81). “Before applying validity evidence from one situation to a new situation,” selection procedures must be evaluated for fairness “if technically feasible” (Jeanneret 2005, p. 81). Regulations embody serious concern about under-predicting minority performance, preferring local validation studies examining differential prediction by race.

The Guidelines allow transportability of validity evidence from one situation to another only in very limited conditions—predicting performance in basically the same job and infeasible local validation. But because Section 5A states that “new strategies for showing the validity of selection procedures will be evaluated as they become accepted by the psychological profession,” lawyers advocating VG can point to an opening, even without amending the Guidelines (43 Federal Register 38289, 38298). Because the EEOCC inserted Section 5A close to negotiations’ end and after getting APA and Division 14 comments and testimony,20 it is possible that the EEOCC nodded toward reminders that science can evolve in unexpected ways. But existing case law is unfriendly to VG precisely because it contradicts Griggs and Albemarle. By the 1980s, plaintiffs’ lawyers thought VG threatened their capacity to

challenge tests successfully in court, but their fears largely went unrealized (Seymour 1988; Landy 2003). Few employers defended tests using VG, perhaps because they thought it futile given institutionalized precedent (Landy 2003).

Consistent with the Guidelines, case law approves transporting validation from one setting to another when job analyses show that the jobs in question are essentially the same as the jobs on which the selection procedure was validated (Landy 2003; Taylor v. James River Corporation, 51 Fair Emp. Prac. Cases, BNA, 893 [1989]). In dicta in a 1983 case in which John Hunter testified as expert for the employer, the judge understood clearly VG’s implications: “According to research, even gross changes in job duties did not destroy validity” (Landy 2003, p. 187 quoting Pegues v. Mississippi State Employment Service).

But in EEOC v. Atlas Paper Co. (6th Cir. 1989), Hunter testified to the validity of a more current version of the Wonderlic test at issue in Griggs and Albemarle, stating that given VG and that Wonderlic was a standard mental ability test, it could be assumed that Atlas’ use of it was valid. The trial judge ruled for Atlas, but the Circuit Court reversed, noting that Hunter never visited Atlas or studied its jobs. Generalizing validity was inappropriate without on-site investigation to show jobs’ similarity. “Upon remand…if the defendant has not demonstrated a justifiable business basis for its practices, particularly in light of…inapplicability of the validity generalization theory, then judgment should be entered for plaintiff EEOC” (868 F
A concurring judge bluntly emphasized the mismatch between VG and controlling precedent:

The premise of the validity generalization theory...is that intelligence tests are always valid. The first major problem with a validity generalization approach is that it is radically at odds with Albemarle Paper Co. v Moody, Griggs v. Duke Power, relevant case law within this circuit, and the EEOC Guidelines, all of which require a showing that the test is actually predictive of performance at a specific job...Albemarle and Griggs are particularly important precedents since each of them involved the Wonderlic Test...Atlas’ validity generalization theory ignores the teachings of Albemarle by implying that no linkage or similarity between those jobs which had been previously researched by Hunter and those at the Company need be shown...As a matter of law, Hunter’s validity generalization theory is totally unacceptable under the relevant case law (868 F 2nd 1499, 1501, emphases ours).

Atlas Paper remains the clearest direct ruling on the evidentiary sufficiency of full-blown VG theory, so it is not surprising that Landy (2003, p. 189) counseled his fellow professionals against relying on VG as the “sole defense for a test or test type.” Analysis of transportability still is needed.

DISCUSSION AND IMPLICATIONS

Our account sheds new light on the mutual influence between Title VII and I-O psychology and has more general implications. First, social authority has
import for socio-legal dynamics when combined with the insight that law and science can be viewed as competing institutional logics. Second, it is useful to incorporate Monahan and Walker’s (2010a) tripartite categorization of social authority, social facts and social frameworks into building a political-institutional theory of conditions for and consequences of law-science co-production. While it never is possible to generalize based on one case, we can use that case to suggest some empirically-grounded building blocks, issues and hypotheses for further socio-legal research. Before we do so, however, we want to clarify where our contribution to understanding legal dynamics fits within socio-legal research.

Building directly on American strands of cultural and political institutionalism described in detail elsewhere (Stryker 2000, 2003; Edelman and Stryker 2005), the political-institutionalist perspective advanced here is consistent with Bourdieu’s (1987) focus on the politics and culture(s) of the juridical field, and with use of the basic concepts of Bourdieu’s field theory by Dezalay and Garth (2002). Basic parallels between Bourdieu’s field theory and other institutional perspectives on law—albeit expressed in different language—may be yet another example of similar fundamental insights arrived at mostly independently in different branches of scholarship. The same may be true for Luhmann’s (1985) arguments about the self-referential, relatively closed nature of law and political-institutionalist arguments about path within legal argumentation frames (Stone Sweet 2002). What no general theory of the juridical field has
done to date—and what we try to do—is construct specific hypotheses for future empirical research based on integrating Monahan and Walker’s (2010a) conceptual apparatus into a broader political-institutional perspective on law. This is especially important because Monahan and Walker’s (2010a) concepts are institutionalized into US legal training. At least in Title VII, these concepts help focus current debate in legal scholarship and in courts, both in connection with Daubert and its progeny, and beyond (e.g., Bagenstos 2007; Faigman 2008; Monahan, Walker and Mitchell 2008; Mitchell and Tetlock 2009; Hart and Secunda 2009; Monahan and Walker 2010b; Mitchell, Walker and Monahan Forthcoming). To paraphrase and elaborate on a well known sociological aphorism, what is perceived as real by actors in an institutional field will have real empirical consequences for construction and transformation of that field.21

21 This will be true whether or not these perceptions are accurate or mistaken.

There are many nuanced differences among diverse “institutional,” “field” or “systems” approaches to law that may carry useful implications for socio-legal research. But these are the subject of other articles, as would be debating diverse interpretations of Bourdieu (1987), Luhmann (1985) or other grand theorists. Similarly, drawing implications from similarities between (some) legal and scientific discourses, whether expressed in terms of rationalization (e.g., Stryker 1994), deconstruction (Fuchs and Ward 1994) or some variant of realism or
Returning to specific implications of our case study, we have shown that new legislation represents exogenous shock to prior argumentation frames, facilitating entry of new elements, including those from science, in cases of first impression. We suggest that, if resolving such cases includes social science used as social authority, as did Griggs and Albemarle, then social science is likely to play an ongoing role in path-dependent unfolding of later cases in the same argumentation framework. Once institutionalized as legal precedent, elements entering law from science are translated into legal terms, restoring legal reasoning’s self-referential nature. In our case study, this translation spurred Title VII-I-O psychology co-production.

If we are correct, a more general theory of conditions for law-social science co-production must specify when social science is more vs. less likely to enter debate over—and resolution of—cases of first impression under new legislation, and when it is more vs. less likely to operate as social authority, rather than positivism (Mercer 2002) is beyond this article’s scope, as are debates about match or mismatch between visions of science embodied in Daubert and in the sociology of science (e.g., Mercer 2002; Edmonds 2003; Edmonds and Mercer 1997). Elsewhere, the first author has written on key themes of this broader literature, including power, legitimacy, and the plurality of legal cultures and of scientific methods (Stryker 1989, 1994, 1996, 2000, Pedriana and Stryker 1997).
than, or added to, *social facts* or *social frameworks*. Because Title VII and its legislative history contained no mandate to consult psychology (or another science), our research shows that explicit legislative instruction to consult social science is *not* necessary for intersection and co-production of social science and litigation-based enforcement of legislation. But legislative bans on social science in law enforcement will, at minimum, reduce use of social science in law dramatically (Stryker 1989).

More speculatively, we suggest that initial law-social science intersection is more likely when at least one actor in the adversary system: 1) perceives that a statute to be enforced provides an opening for social science, *and* that 2) using relevant social science will promote his/her side’s interests, *and* that 3) alternative strategies will not do so. Title VII’s words “professionally developed ability tests” suggested the relevance of I-O psychology to law enforcers and created opportunities to use I-O psychology for legal interpretation. Incentive to do so was high, because readily available, more traditional statutory construction techniques worked *against* plaintiffs.

Because cases of first impression under new statutes often involve construing legislative intent, it seems likely that, *given* reliance on social science in these cases, the role played will include serving as social authority for statutory construction. Assuming this is endorsed and institutionalized, social science is assured continuing roles as social authority and as social fact. For example, once
job-relatedness was institutionalized with scientific ideas of test validity, I-O psychologists were needed to help courts determine if litigated tests had been validated.

As long as the relevant legal argumentation frame is in part social science-based, ensuring routine expert dueling for court fact-finding, it also is likely that social scientists will help educate novice trial lawyers, science as social authority, triggering an institutionalized role for science as social fact, also may lead to an ancillary social framework role for science.

As illustrated by technical issues set in motion by Griggs and Albemarle, once social science routinely is used to help resolve legal liability, scientific issues become legal issues and vice versa. Illustrated by psychologists’ enthusiastic pursuit of differential validation, law henceforth shapes scientific commitments and research agendas. Science continues to shape law, enabling foundational precedent’s further elaboration.

Just because science-based legal interpretation is offered does not mean courts accept it. Contingencies, including case timing and sequence in and across argument frames, affect whether or not social science becomes institutionalized social authority. Had Griggs not been decided before the Supreme Court rejected present-effects theory, and had Griggs not constructed disparate impact separate from present effects, Albemarle, wherein Title VII definitively incorporated test validation might have been decided differently or might not have arisen.
As long as legislation regulates or redistributes legal rights or social costs and benefits, the party benefitting from effects-based statutory interpretation has incentive to use cause-effect reasoning to evaluate how to get maximum impact (Stryker 2007). This does not mean the same party invariably benefits from extant science. Had consensus among 1960s psychologists presumed VG, employers’—not plaintiffs’ lawyers—would have had incentive to use testing science. Follow-through on science-based legal reasoning requires the party contemplating it to think the statute’s text or legislative history provides opening for and substantial benefits from science, while perhaps also thinking it will be much harder to win in court with resources confined to alternative, non-science based reasoning.

This suggests how and why scientific content and the degree of scientific consensus matters. Content determines which party benefits from science’s use in litigation. Unless substantial scientific consensus points in just one direction, there may be no actor with strong incentive to be “first mover” using science in litigation. Unless there is substantial scientific consensus when cases of first impression arise, courts may hesitate to rely on social science as social authority.

Yet another factor inhibiting or facilitating using social science in litigation is a party’s recognition that there is relevant, available science to further his/her interests. Multi-disciplinary law-social science networks and alliances are crucial for this. In our study, overlapping multi-disciplinary networks centering on the EEOC and LDF communicated about practices, concepts, research and
findings in I-O psychology. These networks operated as strategic alliances for legal innovation based in part on I-O science, and then as diffusion mechanisms for disparate impact re-interpreting discrimination’s legal meaning,

So far, we have considered how the situation presents to law enforcers, but not to scientists. Beyond opportunity to exploit a new market niche, why were I-O psychologists collectively, through their professional associations, so interested in informing legal policy, and why were at least some I-O psychologists interested in allying with the EEOC and LDF to ensure aggressive enforcement? It was not because I-O psychologists historically identified with the disadvantaged. Industrial psychology has been more (self- and other-) identified with management than with employees (Gordon and Burt 1981; Katzell and Austin 1992). Once I-O psychologists began routinely to consult and testify in Title VII litigation, many became employers’ experts.

While some I-O psychologists were drawn to Title VII enforcement because they advocated minority rights, it was consistent with the profession’s interests to track and provide enforcement input, as the APA and SIOP did on testing. As we showed, scientific and professional leaders in I-O psychology understood that Title VII would affect I-O researchers in industry. They understood that, at that time, employers often did not seek to validate tests. When employers did try validation, often they did not conform to best or appropriate professional practices.
I-O psychologists could recognize Title VII provided opportunities to promote their science’s utility, enhance professional power and prestige, and expand private sector employment and research. Given scientific consensus on situational specificity and that situational specificity’s corollary was differential validity/prediction by race, 1960s I-O psychologists simultaneously could promote their professional interests and values, and equal opportunity and business productivity and efficiency. This fortuitous three-fold confluence resulting from timing of Title VII’s enactment relative to developments in the then separate field of I-O psychology could not outlast the VG revolution.

Regarding conditions for co-production, preceding discussion suggests that, once social science is institutionalized as social authority in a legal argument framework, then law-science co-production is assured as long as the foundational argumentation framework is not overruled judicially or replaced by new legislation. But things may not be that simple. I-O psychologists may have been especially prone to Title VII shaping their research because many worked for regulated employers, so that Title VII regulated how they deployed their professional knowledge.22 Had I-O psychologists been fully cloistered in academe, perhaps Title VII would have exerted less influence on them.

22 See Author (forthcoming) for private industry’s employment of I-O psychologists. By 1990, about 36% of SIOP members worked in academia, 7%
As well, empirical research is I-O psychology’s core. If it more resembled US economics—theory heavy, with assumed rather than tested canonical principles and a prestige hierarchy downgrading empirical research (Fourcade 2009)—using I-O psychology in litigation may have had little feedback on scientific research. We are conducting comparative research to follow this lead.

Monahan and Walker (1986, 1987, 2010a) proposed social authority, social fact and social framework to elaborate appropriate roles for social science in law, but the competing institutional logics framework applies to law and both social and natural science (Stryker 2000). When natural science is used in litigation [as in environmental law], using it as social authority for court rule-making may have import for socio-legal dynamics somewhat similar to social science as social authority. But given differences as well as similarities in how social and natural science relate to broader political-economy and culture (Moore 2008), we leave law-science co-production beyond the social sciences to future research.

Having discussed conditions of I-O psychology-Title VII co-production, we turn to consequences. First, because lawyer-scientist interaction in litigation and administrative rule-making was ongoing, it created opportunities to diffuse

for public sector organizations, 21% for private business and 29% for private consulting firms.
new roles, leading to further meaning and norm construction, e.g., expert-witnessing promoted APA discussion about norms for psychologists as expert witnesses (Brodsky 1991). Second, I-O psychologists and their associations advised and lobbied government regulators, and testified before Congress and agency hearings about revising legislation and regulations (Camara 1996). When scientists influence formal politics and policy making, this may enhance their competition for status, power and legitimacy, thus extending and intensifying the internal politics of academic science (Stryker 1994, 2000). We are conducting further research investigating this.

Third, competing institutional logics help us explore co-production’s unintended consequences. Once law institutionalizes using science, the result is a hybridized logic internalizing in law opportunities for tensions and mismatch between elements characterizing legal reasoning and elements characterizing science. We developed five types of mismatch: 1) between backward looking, precedent-based reasoning and forward looking scientific discovery and development; 2) between precedent-based analogical reasoning and causal-analytic reasoning; 3) between case-deterministic specific causation familiar to lawyers and the probabilistic general causation familiar to scientists; 4) between legal accuracy and factual truth seeking; and 5) between advocacy toward predetermined ends within a purposely adversarial framework and non-adversary institutions encouraging non-predetermined ends and considering other views.
Our research highlighted especially the first two mismatch types, because Title VII’s embrace of situational specificity and push for differential validation unintentionally promoted scientific revolution. Today’s consensus around VG points Title VII away from foundational precedent’s situation-specific idea of legally appropriate test validation. On one hand, this provides opportunity and resources for further legal innovation—e.g., rewriting the *Uniform Guidelines* to endorse VG or overruling foundational precedent to incorporate new scientific consensus. On the other hand, it provides strong constraints on—and powerful resources to use against—innovation, *given* foundational precedent and thirty+ years of elaborated argumentation frame.

Our research shows that some actors in now overlapping legal and scientific fields urge the first possibility and others the second. Research must examine further the institutional politics stemming from these competing orientations. Future research should examine factors shaping whether results of such institutional politics are *reaffirming*—with *evolutionary modification* of—extant judicial rules, or instead produce *revolutionary change* in institutionalized legal argumentation frameworks.

When the latter happens, law-science co-production parallels somewhat Edelman et al.’s (1999) endogeneity of Title VII to regulated business. But here formal law is reinterpreted by internalizing conceptual change in *science*—and through overt contests of meanings, norms and interests—rather than by
internalizing business compliance strategies. When an institutional politics of actors mobilizing precedent (incorporating the old science) vs. actors using the new scientific paradigm results in retaining precedent, law-science co-production has not resulted in full endogeneity of law to science. In co-production of Title VII and I-O psychology, science’s earlier social authority role paradoxically now limits legal innovation consistent with scientific revolution.

Title VII-industrial psychology co-production did not promote substantial conflict around mismatch between general, probabilistic and case-specific causation, but such conflict structures current debate about social-cognitive and sociological organization science in class actions such as that against Wal-Mart (Bielby 2003; Borgida and Fiske 2008; Monahan, Walker and Mitchell 2008, 2009). When the psychology of race and gender stereotyping and implicit bias enter Title VII litigation, it is as social framework providing probabilistic knowledge of cause-effect relations to inform legal fact-finding. Because plaintiffs’ lawyers have employed psychologists and sociologists to develop social framework evidence about cognitive bias to persuade judges that major US employers engage in gender discrimination in pay and promotion on a massive scale, and because this strategy had some success in lower federal courts, the defense bar fought back (Hart and Secunda 2009). It tried to limit admissibility of plaintiffs’ expert testimony to cause-effect patterns established in peer-reviewed science and to forbid plaintiffs’ experts from using organizational materials
produced in discovery to opine about fit between the general science and the case at hand (Monahan, Walker and Mitchell 2008; Hart and Secunda 2009). Such issues implicate scientific and normative questions unlikely to be resolved against a backdrop of high stakes litigation, even were there a somewhat more collaborative regime for expert involvement.

Implicit bias theory and research also is used to promote “behavioral realism” in Title VII, expanding intentional discrimination beyond focus on overt/conscious intent to reach implicit biases that psychologists suggest operate when not checked by differentiating information, accountability and supervision (Krieger and Fiske 2006). Those promoting behavioral realism urge using social-cognitive psychology as social authority to innovate new judicial rules promoting aggressive Title VII enforcement, just as I-O psychology did when situational specificity reigned. But while behavioral realism with respect to current social-cognitive psychology tends to be plaintiff-friendly, behavioral realism with respect to current test validation science would bring a more management friendly approach to key legal issues pertaining to testing. Thus, calls to bring Title VII up to date using social science as social authority need not result in tilting the playing field more in plaintiffs’ favor.

Finally with respect to legal policy, we urge pursuing connections between legal argumentation frameworks pertaining to discrimination in test-based employee selection and those pertaining to discrimination in pay and promotion
operating through cognitive biases shaping performance evaluation. In 1989, the National Academy expressed concern that while cognitive tests were validated, their predictive capacity was very modest (typically explaining just 4-16% of variation in job performance), with selection errors favoring the white majority. Adding concern were possible biases in measuring validation criteria, typically supervisor evaluation of job performance (Hartigan and Wigdor 1989).

Today evaluating job performance—performance appraisal—is a central topic for I-O psychologists, social-cognitive psychologists, organizational sociologists, managers and human resource professionals (Farr and Levy 2007; Borgida and Fiske 2008). Research suggests that biases in performance appraisals are even more important than the 1989 National Academy report assumed (e.g., Hunt et al 2002; Heilman and Hayes 2008). If some combination of conscious and unconscious biases infected now dated measurement of criterion variables in validation studies providing input for I-O psychologists’ meta-analytic conclusions supporting VG, there might have been sufficient under-estimation of minorities’ job performance that cognitive tests are not the racially unbiased performance predictors I-O psychologists typically presume. Perhaps VG is not so strongly supported empirically as it seems.

REFERENCES


Dezalay, Yves and Bryant Garth. 2002. *Internationalization of Palace Wars: Lawyers, Economists and the Contest to Transform Latin American States.* Chicago: University of Chicago.

Division of Industrial and Organizational Psychology. 1975. *Principles for Validation and Use of Personnel Selection Procedures.*


-----, 2007. Author conducted interview. Iowa City, IA, May 13.


