

Organizational Supererogation and the Transformation of Nonprofit Accountability¹

Aaron Horvath
Stanford University

This article advances the moral philosophical concept of supererogation as a sociological process through which organizational action outstrips externally imposed evaluative demands. It illuminates this process by following 200 nonprofits in the San Francisco Bay Area from 2002 to 2016, when they became subject to myriad accountability obligations. Interviews, surveys, and content analyses show that nonprofits chafed against compulsory reporting and evaluative standards, regarding them as ill-suited to their work and goals. Contrary to current perspectives in organizational theory, however, nonprofits neither minimized external scrutiny nor conformed to external criteria. Instead, they disclosed more information than required. Despite nonprofits' misgivings about oversight obligations, they nevertheless identified with the broader ideal of accountability these demands represented. Navigating this tension, nonprofits repurposed newly salient evaluative practices as tools through which unique goals and values could be given facticity. Through supererogation, nonprofits surpassed mere compliance and pursued accountability in organizationally distinctive ways.

INTRODUCTION

In recent decades, the principles and methods of accounting have permeated organizational life, extending beyond the balance sheet and manifesting in tools for evaluating the efficiency, transparency, and accountability of

¹ This article has benefitted from comments from Alex Barnard, James Chu, Alexandra Kaley, Barbara Kiviat, Anders Krabbe, Ted Lechterman, John Meyer, Woody Powell, Mike Power, Sarah Soule, Viviana Zelizer, and the *AJS* reviewers. Additionally, I would like to thank members of the Stanford Networks and Organizations Workshop as well as participants at the 2020 SASE and ASA annual meetings. The broader project of which this article is a part has been a collaborative effort by me, Christof Brandtner, Denise Gammal,

organizations (Miller and Power 2013; Mennicken and Espeland 2019). This development is consistent with a shift—variously characterized as the “audit society” (Power 1997), the “audit culture” (Strathern 2000), or more provocatively the “tyranny of metrics” (Muller 2018)—from a society of abundant trust and scarce audit to one of scarce trust and abundant audit. The torrent of routinized checking is evident in the growing collection of accountability measures assuring the public of organizational integrity. Legal requirements push organizations to expand reporting, rating agencies set standards and foster interorganizational comparability, and a quickly growing field of monitoring and evaluation investigates the candor of organizational claims. From schools to tech companies to newspapers, modern organizations are being pried open and their inner workings exposed (Espeland and Sauder 2016; Turco 2016; Christin 2018).

Sociologists have long debated how organizations operate under scrutiny. Anticipated by Weber’s (1978) classic statement on bureaucratic secrecy and Selznick’s (1957) insights on the precarity of organizational values, prevailing theoretical accounts emphasize how organizations endure dissonant encounters with externally legitimated criteria of worth. One perspective holds that, because inspections are “assertions of societal control,” organizations will “seek to minimize” their influence through low-commitment structural changes and account giving (Meyer and Rowan 1977, p. 359). To insulate the exigencies of work from the peering eyes of the outside world, organizations engage in decoupling, providing accounts bearing only faint resemblance to organizational realities (Brunsson 1993). Isomorphic with environmental demands, they file forms, adopt policies, and meet standards to satisfy inquisitive outsiders and inoculate against further inspection (DiMaggio and Powell 1983; Edelman 1992; Dobbin and Kelly 2007).

Another perspective proceeds from the observation that mounting cultural emphasis on monitoring and evaluation renders symbolic compliance difficult (Bromley and Powell 2012). Because the ideal of accountability has permeated society, audits are largely “sheltered from critique” and difficult for organizations to resist (Christensen and Cornelissen 2015, p. 133). Unable to compartmentalize the influence of scrutiny into peripheral practices, organizations internalize the criteria by which they are judged, abandoning distinctive goals and conforming their practices accordingly (Espeland and Sauder 2007). Powerless against the unwavering gaze of monitoring and

Hoky Hwang, Woody Powell, Caroline Simard, and David Suárez and has been generously supported by the Stanford Center on Philanthropy and Civil Society. Katey Webber was instrumental in facilitating contacts with nonprofit sector public accountants. Her background in auditing and nonprofit accounting proved invaluable for analyzing interview data and interpreting the subtleties of nonprofit reporting, both financial and otherwise. Direct correspondence to Aaron Horvath, Center on Philanthropy and Civil Society, Stanford University, 559 Nathan Abbott Way, Stanford, California 94305. Email: ahorvath@stanford.edu

evaluation, organizations may become disciplined (Sauder and Espeland 2009). In this fashion, ratings, rankings, and other evaluative mechanisms can cause diverse organizations to adopt similar practices and espouse uniform goals (Sharkey and Bromley 2015).

From either perspective, when institutional environments define the terms of legitimate action, organizations, either ceremonially or substantively, hasten to conform. Decoupled organizations do the bare minimum to satisfy public expectations and ensure legitimacy while buffering organizational goals from external influence. Disciplined organizations conform to external evaluative standards, using them to set priorities and guide decision-making. In either case, when common criteria of worth become objects of cathexis, inter-organizational convergence is likely to result. Indeed, Espeland and Sauder (2016, p. 201)—invoking Weber’s iron cage and echoing Selznick’s language of value subversion—forcefully argue this point: because evaluation “tends to foreclose other modes of creating and expressing value [it] leads to greater homogenization among those being held accountable.”

Although decoupling and disciplining helpfully characterize numerous organizational responses to the intrusions of their environment, adjacent lines of economic and organizational sociology suggest an alternative account where encounters between organizational values and external demands are highly contingent and suffused with generative potential. Contingency is evident in the historically and culturally distinctive approaches to monetary valuation (Zelizer 1978; Quinn 2008; Fourcade 2011) and in the divergent ways organizations make sense of similar market pressures and evaluative technologies (Reich 2014; Christin 2018). Generative potential is evident when supposedly corrosive evaluative techniques are transformed to demarcate social relations and institutional boundaries (Zelizer 1994; Murray 2010) or when encounters between dissonant evaluative principles catalyze new behavioral protocols (Stark 2009). As such, the imposition of evaluative standards may not flatten social relationships so much as it provides novel frameworks through which underlying differences are reproduced (Fourcade 2016). Seen in this light, a new account emerges in which the isomorphic pressures of evaluative criteria mark starting points from which assertions of distinctiveness proceed—not endpoints toward which organizations converge.

Considering the contingent and generative encounters between incompatible conceptions of worth, this article explores how organizations experienced the explosion of measurement, evaluation, and regulatory oversight in the U.S. nonprofit sector during the early years of the 21st century. Following several highly publicized scandals in the late 1990s and early 2000s, nonprofits were inundated with audits, financial ratings, and impact evaluations, each designed to assess fidelity to charitable missions (Reiser 2005). By following 200 randomly sampled nonprofits in the San Francisco Bay Area from 2002 to 2016, I examine how organizations coped with this evolving evaluative

environment. Using data from two waves of interviews with nonprofit directors, surveys, longitudinal content analyses of annual tax returns, and interviews with accountants, I find that nonprofits chafed against obligatory evaluations, regarding them as time-consuming, ill-suited to organizational goals, and misrepresentative. Despite these pernicious effects, organizations neither minimized external scrutiny through shows of pro forma compliance nor conformed internal goals to fit external demands. Instead, they repurposed the compulsory tools, metrics, and discourses of external evaluation in ways that made their unique work and values visible to the broader public. In so doing, organizations not only satisfied external accountability demands; they also volunteered information far in excess of what was required and did so in ways that reflected the distinctive and locally meaningful nature of their work. Organizational supererogation—the name I give to the process underlying these outcomes—illuminates a major transformation in the practice of nonprofit accountability.

This process of supererogation resulted in a heterogeneous array of accountability practices reflecting nonprofits' distinctive work, goals, and circumstances. Some organizations designed performance criteria reflecting their particular missions, hired independent analysts to evaluate their programs in terms of these criteria, and publicized the results whether good or bad. Others, worried that external evaluations actually obfuscated the true nature of their work, began using reports, public documents, websites, and social media to disclose the unvarnished details of internal decision-making and share stories of success and failure that reductive metrics otherwise obscured. Many organizations appended evaluation results and intimate accounts of human impact to their annual IRS filings. These once-sedate financial forms now read as lively chronicles of the past year's activities, growing in length and detail since the early 2000s. Because no one requires such extensive disclosures, nonprofit accountants have become concerned their clients are exposing themselves to liability in the pursuit of probity.

From the perspectives of decoupling and disciplining, these outcomes are striking. To be sure, as decoupling would predict, organizations satisfied external demands by checking boxes, submitting reports, and opening their books for inspection. Yet, by disclosing far more information than they were required to reveal, nonprofits transcended expectations of "ritualistic assurance" to a surprising degree (Feldman and March 1981, p. 177). From the lens of disciplining, such radical disclosure might suggest that organizations have become "docile bodies" (Foucault 1977, p. 138), completely surrendered to the authority of evaluative procedures. But organizations did not cater to external criteria of worth so much as they advanced their own standards and assessed themselves accordingly. In short, because decoupling and disciplining emphasize conformity with external demands, both perspectives fail to explain situations where organizational practice outstrips the standards to which it is held.

In order to explain these surprising outcomes, I develop a processual account of supererogation. Such an account starts from the observation that, rather than reject or acquiesce to compulsory oversight, nonprofits—long committed to openness and responsiveness to the communities they served—embraced the broader ideal of accountability on which newly imposed evaluative demands were premised. But what it meant for nonprofits to be accountable was as varied as the goals and activities that comprised the sector. Accordingly, nonprofits perceived newly imposed disclosure requirements and standard-issue evaluative criteria as incompatible with their particular understandings of this ideal. Navigating the mismatch between obligatory reporting and locally meaningful conceptions of accountability, nonprofits repurposed the tools and tropes of external evaluation as means through which unique goals and values could be given facticity. Whereas compliance merely allowed organizations to satisfy accounting demands, supererogation allowed them to pursue accountability in organizationally distinctive ways.

By integrating the concerns of decoupling and disciplining with insights on the contingent and generative encounters between dissonant evaluative principles, this article makes several contributions. First, it extends institutional theory by showing how stifling external demands can provoke innovation and distinctive assertions of worth. Second, it shows how isomorphic pressures, often treated synonymously with homogenization, can propel the reproduction of heterogeneity (Beckert 2010; Timmermans and Epstein 2010; Bromley, Hwang, and Powell 2012). Third, because supererogation demonstrates how regulatory pressures can be reinterpreted and transformed through concrete organizational practices, it offers a new lens through which to view regulatory dynamics and institutional change.

In the following, I draw from moral philosophy and recent sociological scholarship to theorize a processual account of organizational supererogation. I then contextualize my study within the nonprofit sector over the last two decades. After discussing data collection and analysis, I present my findings, showing support for supererogation and providing evidence against decoupling, disciplining, and other alternative accounts. I conclude by considering the generalizability of supererogation, both in terms of its applicability to other social domains and its implications for sociology more generally.

ORGANIZATIONAL SUPEREROGATION

Drawing from moral philosophy, this article theorizes supererogation as a social process through which organizational action comes to exceed externally imposed obligations. The term *supererogation*, largely unfamiliar in the modern sociological lexicon, comes from the Roman Catholic doctrine of *opera supererogationis*, which concerns good acts in excess of those required

for spiritual salvation. Indeed, the Latin *super-erogare* (“to pay out over and above”) appears in the parable of the Good Samaritan (Luke 10:35). Despite these theological roots, sustained philosophical interest in supererogation surfaced relatively recently, inspired by an effort to complement a three-part classification of moral acts: things one is obligated to do because they are good, things one is prohibited from doing because they are bad, and things one is permitted to do because they are neither good nor bad. To this, philosophers added a fourth category of moral action, supererogation: things that are good to do but that are not in any sense required (Urmson 1958). Thus, in philosophical usage, to supererogate is to voluntarily perform good acts that exceed the demands of moral obligation (Mellema 1991).

For philosophers, supererogation lies at the intersection of two dimensions of normativity: obligations and ideals. Whereas obligations are prescriptive, designating the minimal requirements of good action that all must obey, ideals are open-ended, signifying virtues and indeterminate conceptions of the good that “have no fixed measure and can in principle be always improved and further perfected or realized” (Heyd 2019). Philosophical debates about supererogation center on how these distinct bases of morality can be reconciled, that is, What are the limits of duty? What is the nature of virtue?

In order to adapt the philosophical concept of supererogation for sociological use, we must first understand how its constituent dimensions—that is, both obligations and ideals—manifest in institutional pressures and organizational practices. Indeed, the tensions between these dimensions are instructive for sociological analysis, sensitizing us to both the specific standards by which actions are judged and the foundational social values on which such judgments are based.²

By *obligations*, I mean the various institutional mandates with which organizations, formally or informally, must comply. Such mandates include the acts organizations must perform, the standards they must satisfy, the policies they must adopt, and the structures they must establish in order to maintain legitimacy and continue to operate. “Must” is the operative word. Failing to fulfill obligations to regulators, auditors, funders, or the general public comes with considerable legal, financial, and social sanctions that can jeopardize organizational survival and reputational standing (DiMaggio and Powell 1983). For example, failing to comply with minimum wage laws, report political campaign contributions in accordance with campaign finance regulations, or meet environmental sustainability standards could result in substantial

² Here, my break with philosophical treatments of supererogation is sharpest. Many philosophers endorse an objective and universal perspective of morality, advancing theories of normative ethics (e.g., deontology, consequentialism, virtue ethics) to arrive at specific duties. My approach—like other sociological treatments of morality—is relativist, tying specific obligations to sociohistorically contingent values.

finances, state intervention, or lost certifications. Obligations can operate through less direct, but nevertheless powerful, institutional forces as well. Ratings, rankings, evaluations, and other public accounts of organizational performance—what Schneiberg and Bartley (2008, p. 43) call “regulation by information”—can effectively oblige adherence to particular molds by transforming collective conceptions of legitimacy and creating powerful incentives for organizations to fall in line (Espeland and Sauder 2016). Failing to satisfy widely legitimated standards could prove culturally and materially injurious.

But obligations are more than simply prescriptive demands requiring organizations to perform particular tasks or adhere to specific standards. They are also, by their very nature, the means through which external authorities give operational flesh to relatively ambiguous *ideals* and codify how these general aims are to be pursued in practice. In this sense, obligations like minimum wage laws, campaign finance regulations, and sustainability standards are concrete manifestations of indeterminate ideals like economic fairness, democratic equality, and sustainability.³

Implicit here is the idea that repertoires of social action—whether organizational routines, regulatory demands, or evaluative protocols—are inscribed with cultural assumptions and moral categorizations (Giddens 1984; Bourdieu 1990; Sewell 1992; Mohr 1994). Such dualities are especially evident in audits, evaluations, and other sorts of formal or routine public inspection. Mandated reporting protocols are not simply forms, checklists, questionnaires, and their accompanying methodologies. They also represent the “idealized, normative projections” of those who have mandated their use (Power 1997, p. 4).

In sum, obligations serve two interrelated functions: they set the minimum standards that organizations are expected to uphold, and they direct

³ The indeterminacy of ideals raises questions about how one might identify them in future research. Taking a semiotic approach, one can think of ideals as floating signifiers—i.e., words or concepts that, while commonly invoked and ascribed positive value, have no standardized meaning or agreed-upon referents (Levi-Strauss 1987; see also Levine 2017). As such, different organizations can earnestly claim to be pursuing the same ideal despite disagreement over how these nominal aspirations should be pursued in practice. While these common reference points may facilitate coordination across difference, interpretive disputes may nevertheless rise to the surface, especially in moments when platitudes become protocols or when one entity imposes its conception of an ideal onto another. Accordingly, researchers should examine the discourse surrounding obligations, attending specifically to the language through which new demands are justified and the ends they are claimed to serve. The keywords and generalities pervading such discourse should point to the ideal in question; the different actions organizations attach to these concepts should reveal the contextually contingent meanings of the ideal in practice. Conceptual analogues—such as Rose and Miller’s (1992) distinction between programs and technologies or Cyert and March’s (1963, pp. 26–36) discussion of goals, aspiration levels, and objectives—may provide further guidance.

attention to the broader ideals organizations are hoped to embody. However, because ideals like fairness, equality, or sustainability are susceptible to multiple, even contradictory, interpretations, situations may arise where organizations embrace the broader ideal in question while nevertheless disagreeing with how it has been codified in practice. In such situations, organizations may take matters into their own hands and pursue the higher-order ideal in locally meaningful terms. For example,

A restaurant with an expansive understanding of economic fairness may see an increase in the legal minimum wage as a step in the right direction but one that, on its own, does not go far enough. Although the new law only requires a \$15.40 minimum wage, the restaurant may nevertheless increase its wages to \$18.40 and begin offering below-market-rate housing to employees as well.

A political campaign dedicated to democratic equality may see new campaign finance regulations as ineffective at leveling the political playing field. Beyond merely capping contributions from PACs and lobbyists as the law requires, the campaign may reject all funding from such sources and provide more extensive disclosures than legally required.

A B corporation committed to environmental sustainability may deem a set of environmental certification standards to be insufficient. In addition to satisfying the requisite criteria, the organization may hold itself to more stringent standards and seek to compensate for what it sees as shortcomings in the certifier's requirements.

In each example, the organization embraces the ideal in question but has misgivings about the compulsory tasks and evaluative criteria in which the ideal has been instantiated. Navigating this tension, the organization not only complies with its external demands but also goes above and beyond these demands either in terms of degree (such as the restaurant paying three dollars more than legally required) or in terms of kind (such as the B corporation advancing its own environmental standards in addition to those the certifier imposed).⁴

⁴ This distinction is informed by philosophical treatments of what it means to go above and beyond a moral obligation. Exceeding obligations in terms of degree lends itself to a quantitative interpretation—i.e., “how much . . . in additive units, the performance exceeded duty” (Feinberg 1961, p. 279). Exceeding obligations in terms of kind lends itself to a qualitative interpretation because, while these actions pursue the higher-order ideal embodied in an obligation, they are not reducible to the obligation's criteria. Sociologically, both senses of above and beyond are valid and analytically useful. The former sensitizes us to organizational actions relative to some external standard; the latter sensitizes us to how such standards can provoke innovative or idiosyncratic responses. I am indebted to an anonymous reviewer for pushing me to clarify this point.

Organizational Supererogation

The sociological question, then, is to understand *why organizations would pursue higher aims when, strictly speaking, the bare minimum would satisfy external demands*. As the above examples suggest, a sociological answer to this question must consider (1) how organizations make sense of the obligations they are expected to fulfill, (2) how organizations perceive alignments and misalignments between these obligations, the ideals they represent, and their own understandings of these ideals, and (3) how, in practice, organizations navigate the tensions among their work, their values, and the demands of their institutional environments. By attending to these items, we can theorize a processual understanding of supererogation and illuminate its value in sociological inquiry.

TOWARD A PROCESSUAL ACCOUNT OF SUPEREROGATION

A processual account of supererogation begins by first considering how organizations come to understand the obligations they face and the broader ideals these obligations represent. Indeed, organizations often imbue ostensibly identical events, regulations, and technologies with strikingly different meanings and implications for practice (Barley 1986). For instance, Christin (2018) demonstrated that American and French newsrooms had divergent reactions to the introduction of analytic technologies designed to measure digital news performance via fluctuations in web traffic. Whereas the commercially oriented U.S. newsroom was able to accommodate the incessant market pressure of audience metrics, the civically and intellectually oriented French newsroom struggled to adapt. Similarly, Reich (2014) showed that variations in founding missions and organizational histories influenced how hospitals navigated the commodification of health care. Whether they saw treatment as a right, religious responsibility, or population-level outcome corresponded to the specific challenges they faced in a competitive marketplace. In short, the course of organizational action is not determined by external pressures alone but by the contingent encounters between these pressures and distinctive organizational values and circumstances. Convergent pressures need not produce convergent responses.

The contingency in how organizations make sense of external demands can ultimately be characterized in terms of dissonance and resonance—a matter of alignment between requisite practices and the organizational settings onto which they are imposed. In some cases, organizations view new obligations as impractical, irrelevant, or out of step with organizational values and moral categorizations (Kiviat 2019). In such moments of dissonance, we typically see attempts to reconcile between organizational interests and the demands of the institutional environment. When obligatory practices and the ideals they represent fail to resonate, organizations may shunt their responses to the organizational periphery. This process of

decoupling results in ceremonial activities that provide public assurance of organizational legitimacy while insulating the organization's work and goals from external influence (Meyer and Rowan 1977). For example, a company with little interest in workplace equality might establish new offices to signal compliance with civil rights laws without substantially altering discriminatory employment practices (Edelman 1992).

Decoupling, however, may prove difficult to sustain, especially when staff or other constituents identify with the ideals in question (Turco 2012). Symbolic activities may eventually penetrate deep, even becoming integral to organizational practice (Hallett 2010). And as organizations orient themselves around institutional demands, they may come to internalize external criteria of worth. This process of disciplining results in organizations striving to align their goals and activities with external expectations and performance standards (Sauder and Espeland 2009).

But we need not assume that organizations are either wholly averse to or totally consumed by the demands of their institutional environment. In fact, new obligations and the broader ideals they represent may resonate with organizational values and activities—assimilated as tools for pursuing goals and overcoming obstacles (see McDonnell, Bail, and Tavory 2017). Furthermore, dissonance and resonance may operate simultaneously. Organizations may recoil at the specific demands of a given obligation even though they identify with the general ideal these demands represent.

The contradictory, turbulent character of such moments—when organizations are both attracted to and repelled by the demands of their environment—affords opportunities for unfamiliar or unwelcome practices to be endowed with new meanings and carried out in ways that better fit organizational ends (White 2008; Padgett and Powell 2012). Organizations may circumvent mandated practices—ignoring some features and emphasizing others—in ways that defy designers' expectations (McDonnell 2010). Murray (2010), for example, shows how academic scientists, facing pressures to commercialize their research, transformed the meaning of patents, repurposing them in ways that preserved scholarly norms and reaffirmed institutional boundaries. Thus, rather than result in resistance or submission, the friction between organizational values and externally imposed obligations may stimulate entirely new forms of organizational action (Stark 2009). Incongruous routines and technologies may be creatively redeployed as tools through which local values and characteristic understandings of widely held but indeterminate ideals (like fairness, equality, or sustainability) can be expressed.

Building on these perspectives, I argue that the process of supererogation occurs when organizations embrace the broader ideal on which an obligation is premised and yet find the specific codifications of that ideal to be incompatible with their work and goals. They may have different understandings of how the ideal ought to be achieved in practice or subscribe to different beliefs

about what behaviors will bring about the idealized state of affairs. Standard-issue accountability criteria, for example, may ring hollow when viewed through the values, histories, and day-to-day experiences of specific organizations. Faced with the push and pull of dissonance and resonance, organizations may forge new practices from the debris of unwelcome techniques and technologies, using them as the “expressive equipment” (Goffman 1959, p. 22) through which their values can be lent facticity and broad ideals can be pursued in locally meaningful ways (Barnard 2016). Through this process of supererogation, organizations come to perform actions that go above and beyond mere compliance with external obligations.⁵

Whereas decoupling results in ceremonial compliance with external demands and disciplining results in the internalization of external demands, supererogation results in organizational action that not only satisfies these demands but also surpasses them in pursuit of the higher-order ideals they represent.

ACCOUNTABILITY OBLIGATIONS IN THE NONPROFIT SECTOR

To illuminate this process of supererogation, I examine the experiences of nonprofits over the first two decades of the 21st century, a period during which monitoring and evaluation obligations multiplied. Public demand for accountability grew sharply following a series of scandals in the 1990s and early 2000s in which several nonprofit leaders were convicted of fraud and embezzlement. Shortly after 9/11, a handful of high-profile organizations came under fire for misleading donors, misusing funds, and overpaying executives. Congressional hearings discussed eradicating “fraudulent” and “wasteful” charity while protecting Americans’ “wonderful spirit of giving.” New York’s attorney general, Elliott Spitzer, testified that, if charities continued unsupervised, citizens would “lose faith in the entire not-for-profit sector” (House Committee on Energy and Commerce 2001, p. 29). Others worried that a “litany of recent miscreancy” had undermined confidence in a sector long celebrated for its trustworthiness (Fleishman 2001, p. 178). Restoring faith in nonprofit accountability was an explicitly moral endeavor: a good nonprofit should have nothing to hide.

Eager to scrutinize nonprofit integrity, various entities—government regulators, nonprofit rating agencies, and philanthropic funders—codified the broad ideal of accountability into oversight mechanisms emphasizing official

⁵ The process-outcome distinction here is important. Supererogation is a specific process through which organizations come to exceed external demands. The mere fact that an organization exceeded such demands does not in and of itself constitute supererogation. I further develop this point in the article’s discussion section.

disclosures, financial ratings, and program evaluation. Such formal and quantitative demands represented a major shift in a sector in which accountability had long been informal and qualitative, achieved via personal experience, face-to-face interactions, overlapping memberships, and word-of-mouth communications (Hammack 1995). The nonprofit sector's shift reflected a broader trend in regulatory approaches to accountability across numerous settings. As Espeland and Vannebo (2007, p. 38) put it, "accountability can mean many things but increasingly we link accountability to quantification." Formalized methods of verification—audits, benchmarking, and the monitoring of performance indicators—came to dominate settings once characterized by highly relational forms of accountability (Power 1997). Accordingly, when nonprofit regulators, raters, and funders pushed for increased oversight, they did so by mandating formal disclosures, imposing financial criteria, and requiring quantitative evaluations. These methods, as one congressional report put it, represented an "integrated package" that would ensure nonprofits were "ethical, responsible stewards of Americans' generosity" (Panel on the Nonprofit Sector 2005, p. 4).

For regulators, accountability was a matter of ensuring that nonprofits were using their tax-exempt status in ways that produced demonstrable public benefits. Citing recent abuses of public trust and taking inspiration from Sarbanes-Oxley, lawmakers around the country mandated that nonprofits undergo financial audits and fulfill extensive formal reporting requirements (Reiser 2005). California passed the Nonprofit Integrity Act in 2004, which mandated annual audits of nonprofits with more than \$2 million in revenue, ordered nonprofits to establish independent audit boards, and required that select fundraisers be registered with the state's attorney general.

Regulators also sought to make nonprofits' official disclosures accessible to the broader public. In 1996, Congress mandated that nonprofits make their federal tax returns (IRS Form 990) available on demand, envisioning these forms as "the primary document providing information about the organization's finances, governance, operation, and programs for federal regulators, the public, and many state charity officials" (Panel on the Nonprofit Sector 2005, p. 26). Working with two data clearinghouses (GuideStar and the National Center for Charitable Statistics), the IRS began digitizing 990s and making them available online. The effect, proponents hoped, would be to thrust nonprofits into public view and "cause rogues to think twice" (Swords 1998, p. 581).

This trove of newly digitized data presented an opportunity for fledgling rating agencies—including Charity Navigator, the Better Business Bureau Wise Giving Alliance, and Charity Watch—all of which drew from 990 data to compute financial ratios and assign accountability scores according to specified benchmarks. Focal in each agency's criteria were two measures: the overhead ratio (the proportion of total expenditures allocated to

administrative and operational costs) and the program ratio (the proportion of total expenditures allocated to direct service provision). A nonprofit spending less than 15% on overhead and more than 75% on programs was deemed exceptionally accountable. A nonprofit spending more than 30% on overhead and less than 60% on programs was deemed questionable (Wise Giving Alliance 2003; Charity Navigator 2004; Charity Watch 2005). Such financial assessments, raters claimed, would help the public make informed decisions about nonprofit accountability and “promote ethical conduct by charitable organizations” (Wise Giving Alliance 2003). In this manner, rating agencies positioned themselves as objective intermediaries between nonprofits and the broader public, even recommending that potential donors dismiss nonprofits’ claims of trustworthiness and rely instead on raters’ official assessments (Charity Watch 2005).

For foundations, accountability was a matter of ensuring that grantees used funds as donors had intended. In practice, this concern materialized in a mix of financial oversight (often with criteria similar to those used at rating agencies) and results-based accountability that emphasized intensive program evaluation and outcome measurement (Center for Effective Philanthropy 2002). Several foundations—including leading Californian funders like Hewlett, Irvine, and Packard—established dedicated evaluation units and incorporated outcome measurement into their grant-making procedures (McMullan and Patrizi 2001; MacIndoe and Barman 2012).⁶ While proponents of this approach called for data-intensive analyses that would draw causal links between nonprofit activities and higher-order social impacts, the sheer difficulty of such analyses led many funders to embrace easily observed indicators of program implementation and completion. Common indicators included the number of people served, the number of units of service provided, and the number of hours spent on grant-funded work.

To determine evaluative criteria, foundation staff often looked beyond the work and goals of specific grantees and opted instead for more general metrics that could be applied across numerous organizations. For some, this practice was rooted in the belief that performance criteria ought to align with funders’ interests. Some went so far as to disparage grantee-centric evaluations that were “not pertinent to the information needs of foundations” (McMullan and Patrizi 2001, p. 14; Center for Effective Philanthropy 2002). For others, general-use outcome indicators would allow foundations to benchmark grantees’ performances against one another in order to determine which grantees

⁶ As a new generation of investors and technology tycoons established foundations, they amplified this approach by merging concepts like return on investment and key performance indicators with mainstream philanthropic interests in performance-based accountability (Barman 2016; Horvath and Powell 2020).

deserved continued support. Under this approach, organizations as diverse as an eldercare facility and a workforce training program might be evaluated according to the same criteria. As one foundation report demonstrates, organizations pursuing similar goals would be evaluated in the same terms (e.g., number of participants recruited and average cost per participant), despite serving different demographics, using different treatment modalities, and operating in different socioeconomic circumstances (see Hernández and Visser 2001).

As regulators, raters, and funders advanced standard-issue criteria for accountability, few nonprofits saw their work or understandings of accountability reflected in these cut-and-dried terms (Carman 2007). Some critics worried that the growing popularity of financial criteria would cause nonprofits to underinvest in their staff, forego long-term improvements, and hesitate in response to unforeseen community needs (Wing and Hager 2004). Others worried that overemphasizing easily measured activities would undermine the relational aspects of nonprofit work, thereby rendering nonprofit-community relationships transactional (Benjamin 2012). And still others worried that results-based accountability would compromise elusive civic goals like citizen engagement and community building (Mosley, Marwell, and Ybarra 2019).

Current sociological perspectives suggest these fears were warranted. Through the lens of decoupling, one might have expected nonprofits to engage in low-cost symbolic activities that signaled compliance while helping to minimize external scrutiny and inoculate against claims of malfeasance (Edelman 1992; Dobbin and Kelly 2007). Or, in a more instrumental variant of decoupling, nonprofits might have made superficial prosocial claims in order to “neutralize” the reputational threat of increased public scrutiny (McDonnell and King 2013, p. 409).

Viewed through the lens of disciplining, one might have expected organizations to internalize the narrow criteria by which they were assessed, using them as guidelines for goal setting and self-evaluation (Sauder and Espeland 2009). Indeed, observers of the early-2000s nonprofit sector believed that new accountability obligations would result in a fieldwide reorientation and interorganizational convergence. For some, this was concerning—the harbinger of mission drift and homogenization (Ebrahim 2005; Hwang and Powell 2009). Others, however, saw the potential reorientation more optimistically, hoping that widespread “shifts in mindset, norms, and practices” would improve how nonprofits made decisions about their work and the people they served (Hernández and Visser 2001, p.2). Proponents even sought to hasten this outcome by calling for “universally agreed upon best practices” for nonprofit oversight (Murray 2001, p. 40), working to build a “common outcome framework to measure nonprofit performance” (Lampkin et al. 2006, p. 3), and creating “common standards for excellence”

(Brest 2020, p. 408). Indeed, “universal” criteria almost received federal imprimatur in 2007 when the IRS—following the lead of rating agencies and foundations—proposed to include financial ratios on the first page of the 990.⁷

Even without a dominant regulatory agency or single set of evaluative standards, the new evaluative landscape augured widespread convergence around short-term, easily measurable, and intervention-centric nonprofit work.⁸ As DiMaggio (2002, p. 269) argued at the time, the growing obsession with “measurable indicators will tend to drive out less easily measured indicators [of] organizational vitality and the quality of services delivered to clients.” Had such a shift occurred, nonprofit historians later concluded, it would have propelled a far-reaching transformation in the sector’s “basic epistemology” (Soskis and Katz 2016, p. 32).

The absence of decoupling and disciplining, however, need not imply the absence of other meaningful effects. As I will show, nonprofits did not take issue with the ideal of accountability—indeed, accessibility and responsiveness to community needs figured prominently in how they understood themselves and their work. But nonprofits did take issue with how the ideal of accountability had been codified into mandatory disclosures, evaluative criteria, and other forms of requisite oversight. In order to pursue the ideal of accountability as they understood it, nonprofits drew on newly salient evaluative techniques and discourses, using them as a socially legitimated vocabulary through which their distinctive understandings of accountability could be made legible to the broader public.

Whereas decoupling sees organizations insulating internal values by minimizing oversight and neutralizing negative attention and disciplining sees organizations subordinating internal values to external criteria, supererogation sees organizations adapting external demands to better fit internal values. Via supererogation, organizations not only satisfy external demands but also extend beyond them through actions that reflect locally meaningful conceptions of widely shared ideals.

⁷ During the public comment period for the proposal, the IRS received a smattering of letters from nonprofit directors opposing the ratios as misleading and misrepresentative of organizational accountability. The proposed changes were never implemented.

⁸ I agree with others’ arguments that evaluative plurality can weaken the potential for interorganizational convergence. In essence, the more evaluative frameworks there are in play, the less disciplining any one framework can be (Espeland and Sauder 2016, pp. 181–92; Brandtner 2017). But it is also important to consider how particular evaluative criteria come to monopolize a field’s evaluative landscape and how the dynamic interplay of the evaluators and the evaluated alter the possibility of such monopolization. I return to this point and supererogation’s implications for regulation at the end of the article.

DATA AND ANALYSIS

To empirically examine the process of supererogation, it is important to understand how nonprofits encountered, made sense of, and responded to demands for monitoring and evaluation. Accordingly, I marshal data from a longitudinal study of 200 randomly sampled nonprofits operating in the 10-county San Francisco Bay Area, a region that includes San Francisco, San Jose, Oakland, and surrounding counties. Organizations were sampled using IRS data on the 10,149 501(c)(3) organizations operating in the region in 2000.⁹ The sample represents the gamut of charitable interests with organizations focused on arts, athletics, education, the environment, health, human services, international interests, and religion. Data were collected in two main waves between 2002 and 2016 with additional data collected throughout.

The first wave consisted of 200 one- to two-hour interviews with nonprofit leaders (executive directors, CEOs, presidents, etc.). Conducted between 2003 and 2005, interviews examined organizations' histories, management practices, public engagement, and experiences with oversight. Using surveys and interviews, a second wave of data collection took place between 2014 and 2016. Surveys included both closed- and open-ended questions paralleling those asked in the first wave and garnered a response rate of 63%.¹⁰ To contextualize survey responses, one- to four-hour interviews were conducted with 26 representatives from 21 organizations. Through interviews, respondents discussed experiences with external oversight and detailed how their accountability practices had changed.

Although interviews shed light on organizational orientations, motivations, and values (essential for understanding how nonprofits made sense of growing demands for oversight), it is important to put these data in dialogue with actual behavior (Jerolmack and Kahn 2014). To do this, I use three additional data sources that offer insight into organizational activities and disclosures over time.

The first additional source is a collection of all available Form 990 filings produced by the sampled organizations between 2002 and 2015 (2,391 forms in total). Although the IRS requires that organizations use these forms to report on finances and governance, nonprofits began using them as media through which they could narrate their unique activities, values, and

⁹ Of 264 organizations, 200 opted to participate in the study (a response rate of 76%); 64 either refused to participate or were dropped after four unsuccessful contact attempts. The sampling frame did not include congregations, private foundations, or organizations with annual revenues under \$25,000. Despite the Bay Area's renown for an active civil society, the sample reflects the general population of U.S. nonprofits in terms of age, size, and activity.

¹⁰ Over the study period, 29 organizations left the sample through closures, mergers, or change in legal form. Of the remaining 171 organizations, 108 participated in the survey.

community contributions—information the IRS does not require. Of particular interest on these forms is the Program Service Accomplishments section, where organizations can provide text describing the prior year's activities. Also of note is the optional Schedule O where organizations can further elaborate on accomplishments statements. I manually coded these sections, recording length (total lines of text) and the use of quantitative detail (e.g., "422 students"; "67% of participants") for each available organization-year. These data offer evidence on both the broad trends of nonprofit disclosure and the specific ways nonprofits pursued accountability in practice.

The second additional data source draws on content analyses of organizations' web pages conducted contemporaneously with the second interview wave. I coded web pages according to functionalities and the types of information nonprofits made public. To assess change over time, I accessed earlier versions of web pages using the Internet Archive's Wayback Machine.

The third additional source draws on in-depth interviews with six partners in three nonprofit sector accounting firms operating in the Bay Area. Each respondent served around 12 nonprofit clients and worked at firms representing hundreds of nonprofits throughout the region. Although these respondents are professionally obligated to provide assurance (i.e., attest to the reasonableness of disclosures and adherence to financial standards), their involvement in nonprofit accounting over the past two decades provides an independent perspective from which to assess how nonprofits understood, came to grips with, and responded to the evaluative demands of their environment. Additionally, accountants provided me with technical knowledge on 990 reporting requirements and helped explain how nonprofits have used the form since the early 2000s. By incorporating accountant interviews with other data sources—interviews, surveys, and content analyses—I am able to examine the transformation of nonprofit accountability from multiple vantage points.

FINDINGS

In the early 2000s, nonprofit leaders described themselves as tangled in a thicket of disclosures, reports, and evaluations. Although they were indignant toward new obligations, their resentment was directed toward onerous reporting demands and ill-fitting evaluative criteria—not the general ideal of accountability. The challenge was not that their organizations were being scrutinized but that the specific evaluations to which they were subject often obscured organizational realities and interfered with nonprofits' longstanding efforts to be accountable to the communities they served. Indeed, what accountability meant to the soup kitchen and what it meant to the kids' club were two very different things, and standard-issue assessments reflected neither. Both vexed by obligatory oversight and committed to its underlying ideal, nonprofits began repurposing newly imposed practices as

means through which they could pursue accountability in locally meaningful terms. Expanding on these general observations, I present my findings in three sections organized around the dissonance of obligations, the resonance of ideals, and the transformation of practice.

THE DISSONANCE OF ACCOUNTABILITY OBLIGATIONS

By 2005, the shifting landscape of nonprofit oversight had become clear. Reflecting on the shift, the director of a San Francisco–based international school explained, “People are much more critical of the product they’re buying in the nonprofit environment. We feel we have to prove [ourselves]. . . . Now people say ‘Don’t just tell us what tuition is. Tell us what we’re getting for it.’ . . . They’re asking for more accountability.” Nonprofits felt hobbled by demands—from regulators, raters, and funders—to disclose information through reports, evaluations, site visits, and other assessments. Recent years had seen the addition of “so many different layers of scrutiny,” explained the head of a rehabilitation center, that “it’ll boggle your mind.” Beyond bewilderment, nonprofits experienced contradictions between the specific character of their work and the abstract standards by which their work was judged. Obligations distracted from activities, were discrepant with goals, and distorted public understandings of charitable work.

Distractions

In 2005, complying with external demands strained resources and impeded nonprofits’ abilities to pursue long-standing goals. For one eldercare organization, beholden to “six different program audits” from city, state, federal, and foundation partners, the time spent complying with assessments left them feeling “besieged” by “unrealistic” demands. One funder insisted they use an intake assessment that had “grown to nine pages,” provide a client photo, and assess each client’s “level of frailty.” “It distracts from our mission,” the director lamented, noting that this was only one of several intake assessments his organization was now required to perform.

Others were less reserved in expressing frustration. A choir director called obligatory reporting a “burden and a nuisance,” and the director of a children’s safety organization called it “a big waste of paper [that] takes resources away from our services.” According to a housing assistance program director, “more time goes into the reporting and less time to actually working with people.” Describing the absurdity, an after-school club director detailed how evaluations interfered with daily work:

We have to train our staff to “get results.” You’ve got thirty-five kids enrolled in basketball. You’re gonna start off by putting their names in Excel and teach

Organizational Supererogation

them all how to take their pulse. Then you're gonna play your basketball game. . . . And you're gonna do that every day for nine weeks. . . . [Evaluations] force you to take part of your activity time and include measurement time. . . . But at the same time, we gotta let them play some basketball.

For some, conflicting performance criteria further complicated the situation. One evaluation might measure pulses, another might measure 100-yard dash times, and another might require measures of academic performance. As the head of a homeless services organization complained, "Funders will say, 'You didn't meet our latest criteria,' and we're saying, 'Well, two years ago those *were* your criteria.'" Noting the irony of evaluations designed to eliminate inefficiencies, the head of a yoga ashram quipped, "There's a lot of waste happening there."

Additional workloads brought additional costs. A youth employment center director described one funder requiring data collection but refusing to pay for it, leaving the organization with a \$19,000 shortfall. Another respondent worried he would have to "eliminate three or four positions" to defray the costs of complying with demands.

It is important to note the character of these frustrations. One director explained that, despite the notoriety of recent scandals, "Most nonprofits—I'm sure almost all of them—are totally honest and will indeed use [funding] the way they think best. . . . Reporting requirements just impede the process."

Discrepancies

Many nonprofit leaders were frustrated that external assessment criteria were poorly matched to their organization's work. In 2005, among the 65% of respondents whose organizations were required to disclose information to external parties, 74% observed a major discrepancy between their own definitions of success and those imposed from outside. Abstract units, classifications, and standardized measurements routinely failed to reflect local knowledge.

For some, external performance standards were comically inappropriate to their work. One respondent from a family counseling organization noted that, when she first saw the data collected on her program, "I was on the floor laughing," sarcastically adding, "Back to the drawing board!" Arts organizations fumed at funder-imposed definitions of success. The head of a Latino performance troupe emphasized that "oftentimes a failure for us is actually a good thing. . . . If we did a play and it flopped, we learn and then develop the right kind of play." Asked about funders requiring outcome evaluations of a youth choir's performances, a director cheekily replied: "What outcome? The show was wonderful."

External demands for quantification proved especially challenging for organizations providing in-person services to beneficiaries. According to a

school director, the drive toward “more scientific and more quantitative” outcome criteria had forced teachers to make definitive statements they “don’t always feel comfortable” making about students. The head of an LGBTQ community center in San Francisco explained that those who “want more statistics . . . don’t really care how well you’re serving them, they just care *that* you’re serving them.” Many directors lamented that donors wanted to see “proof” or something “that proves they are making a difference.” But “proof” was a matter of perspective. For organizations, proof was grounded in concrete experiences and face-to-face interactions. For funders and evaluators, proof was grounded in abstraction and impersonal comparability. Consequently, many organizations found that distilling complex outcomes into bare numbers was an exercise in obfuscation, not elucidation. The director of a family therapy organization elaborated:

We could, for example, say we have trained fifty parents and tell our funders that the fifty parents had their behavior changed, but we would really like to say . . . “the families could have a family meal without it erupting into a big argument.” . . . Lots of [the required assessments] are simply frustrating and a lot of baloney. . . . They are not measuring anything besides the level of frustration that you can put a client through having them sit there for an hour doing this silly stuff. . . . I hope it is going to get better, and I think it is a worthwhile thing . . . but it needs to get closer to the actual activity.

Facing discrepancies between reporting requirements and goals, organizations found ways to reinsert the essence and meaning of their work into the reports they were required to file. Discussing how quantification was only a “small part of the story,” the director of a youth employment center shared, “Whenever I do quantitative data, I always write and explain that number. To me, without the explanation, that number is meaningless. To them, the explanation is meaningless.” Noting that outcome data failed to capture the true nature of her organization, the leader of an outdoor adventures program suggested a better way to assess program effectiveness: “meeting the people whose lives we’ve impacted, hearing their stories, seeing the smiles on their faces, and seeing the glows in their eyes.”

Distortions

Many organizations worried that colorless and reductive representations would distort public understandings of their work. Respondents pointed to financial rating agencies such as Charity Navigator as a major driver of such distortions. As one accountant remarked, “How can you have a measuring stick that is equally viable for the YMCA as it is for a three hundred-bed hospital?”

Respondents were especially irked by use of overhead ratios. The “bizarre” classifications resulting from these metrics led many nonprofits to

discredit (although rarely ignore) raters' verdicts. The head of a Boys and Girls club explained, "I'm constantly stuck not being able to pay people because I need to keep my percentage of overhead—quote, unquote—at a certain level."

Discussing how financial ratings led community members to draw inaccurate conclusions about the quality of an organization's work, a Contra Costa-based housing director lamented that "it's really easy to make yourself look good when you're not doing the work and when you're not really being ethical." Even though her organization was in the midst of taking over another organization's programs ("because they've made such a mess of them"), public ratings did little to distinguish the two nonprofits: "If you go on Charity Navigator, they look okay. They look as good as we do." Ironically, raters' attempts to assess accountability from afar ended up obscuring organizational differences that, to respondents, were plainly visible on the ground.

Beyond financial ratios, respondents worried that overconfident quantitative analyses and the patina of scientific objectivity had ultimately distorted public understandings of their work. The director of a counseling center explained:

Let's say you have a kid who is rating very low because their behavior is awful. [Evaluators] will be saying 'Why are you spending resources on this kid? Your program must suck . . . you haven't made a difference in their life.' And they base [their view] on the numbers and not the fact that his dad beat his mom and his brother shot himself. . . . They'll conclude another organization's numbers are better. But that organization deals with kids whose parents actually talk to them.

Instead of assessing nonprofits from afar, respondents proposed that the public should "come and see" for themselves. As the head of a safety training organization opined, "Our world is going into numbers and trying to do things online that ought to be done in person," adding, "It's not that we're not willing to be accountable. But the way that we're being forced to do this, it's not working."

The new slate of accountability demands marked the depersonalization of relations between organizations and their audiences and the reallocation of public trust from nonprofits to intermediaries claiming to certify nonprofit integrity (see Power 1997; Espeland and Vannebo 2007). As respondents made clear in the early 2000s, abstract and impersonal verification protocols were, at best, poor reflections and, at worst, inaccurate distortions of their organizations' activities and values.

THE RESONANCE OF ACCOUNTABILITY IDEALS

Although external obligations clashed with goals and impeded work, organizations did not question *whether* they should be held accountable. Instead, as

evidenced by efforts to complement impersonal quantitative reports with personal narratives and entreat outsiders to “come see what we do,” the question to nonprofits was *how* they should be held accountable. In devising an answer to this question, nonprofits differentiated two general types of accountability: the thick accountability practiced through relationships with the community and the people they served and the thin accountability practiced by fulfilling newfound obligations. Understanding the former helps to explain the effects of the latter.

Accountability as a Deep-Rooted Value

Before the deluge of oversight demands in the early 2000s, nonprofits often practiced accountability through direct communications and face-to-face interactions. Such activities, respondents explained, were rooted in long-standing commitments to openness and responsiveness to the people they served and the broader communities in which they were embedded. According to the director of an ecumenical housing organization, “If we’re not doing what we said we were going to do, we should be called on the carpet.” To her, being a good nonprofit meant being “transparent about strengths and shortcomings” and seeking to “correct” course when community needs changed.

Indeed, many nonprofits in the sample were founded by community-focused activists and collectives with abiding beliefs in the causes they pursued. Articles of incorporation, bylaws, and mission statements chronicle these founding commitments, dedicating organizations to “responding to community needs,” “representing community interests,” being “accessible to all,” and “stimulating public discussion and debate.” “Accountability,” as the director of a family services center remarked, “is in our belief system.”

In 2005, reflecting on past approaches to accountability, respondents described practices involving direct feedback, in-person interactions with clients and community members, and diffuse word-of-mouth communications. Such practices, respondents explained, allowed beneficiaries, donors, collaborators, and community members to assess nonprofits and, at the same time, kept nonprofits responsive to community needs. A soup kitchen director described her intimate approach to accountability in which she dined with clients: “I ask them ‘How’s the meal?’ . . . ‘Is the food okay?’ and I immediately give feedback to the cook: ‘Too salty’ . . . ‘The rice was a little hard.’” The head of a Latino radio station described “making our airwaves available” by hosting gatherings in the community and having conversations with listeners: “We want to know what hurts and what they need.” Using what they learned from these interactions, they would tailor their programming by offering airtime to groups that “provide different types of services” that could “heal those pains.”

Organizational Supererogation

Alongside informal and ad hoc approaches, nonprofits also practiced accountability through structural means. One approach was to host town hall meetings in which organizations would solicit input and community members could voice concerns. Another approach relied on overlapping memberships and interorganizational coalitions. Such networks facilitated information flow and allowed organizations to, as one director put it, provide “checks and balances” on each other’s work. Others felt that if beneficiaries and community members were to “have a say,” they should “serve on the board.” A domestic violence shelter set aside one-third of its board seats for beneficiaries. An employment center’s board was entirely composed of past beneficiaries. Several disabilities-focused organizations reserved board seats for clients, noting it would be inconsistent to promote the self-determination of disabled persons without also making space for their voices at the top of the organization. The director of a Romanian aid group described how his organization, in order to hold itself accountable to beneficiary needs, established a subsidiary in Romania, staffed it with recipients, and authorized it to oversee the parent organization’s decisions.

Explaining how the broader public assessed accountability sight unseen, respondents often referenced the importance of “word of mouth.” Referrals, in their view, were extensions of the trust fostered through in-person interactions. According to the director of a drug rehabilitation center, the transitivity of trust ultimately improved outcomes for beneficiaries. If referring agencies saw the center as “credible and reliable,” then the clients they referred to the center tended to more easily develop the “trusting relationship” necessary for successful treatment. Such trust required the center’s staff to be straightforward with referrers, “even when we mess up.” Similarly, the director of a K–8 after-school program explained how “word of mouth” kept her organization “on top of things and doing its best work.” In 2005, many respondents recognized that, when evaluating the accountability of otherwise unknown nonprofits, people turned to trusted friends, caseworkers, and organizations for an opinion.

Resonant Ideals and Reexamined Practices

Whatever nonprofits’ misgivings about their new accountability obligations, they were remarkably compliant with mandatory disclosures and other external demands. Between 2005 and 2015, 97% of nonprofits reported top paid salaries on their 990s and 86% satisfied every filing requirement with the California Attorney General’s Office (as required under the Nonprofit Integrity Act). According to the head of a career readiness program, “We’re doing everything the way it needs to be done . . . you’ve got to run properly in the eyes of the state [and] you’ve got to run properly in the eyes of the IRS.” Likewise, nonprofits were keen to stay within the bounds of financial rating criteria

as well. In 2015, nonprofits recorded an average overhead ratio of 26%—a decrease of 4% since 2005 and “good” according to Charity Navigator’s criteria. Even so, accountants worried their clients were being “overly conservative” and “punishing themselves” by not reallocating expenses in ways that would further reduce overhead ratios. But as the director of a homeless services center explained, such maneuvering—however legitimate in the eyes of accountants—would undermine cherished values: “We won’t try and move our data around to make it look like it fits. That would be a moral issue. . . . It’s an imperative thing that we be transparent. [We’re] not trying just to go after a dollar because it’s there.”

Whereas face-to-face interactions and word-of-mouth communications helped build a “bridge of trust” (as a swim team director put it) between nonprofits and the broader public, the arrival of more standardized and quantified forms of accountability promised another way across. An East Bay soup kitchen director observed that, unlike people who “have known us for years,” a new crop of community members “want to know how we rank from an outside source.”

The experience led nonprofits to reexamine how they practiced accountability. In 2005, the director of a San Jose community health center explained that complying with ill-conceived performance criteria had prompted his organization to “develop an evaluation model” of its own: “We provide a great service, but we need to document it. We need to show the benefit of our programs.” Proper evaluation, he added, would be “different than what [others are] asking for.” The director of a relationship counseling center explained that being subjected to poorly designed evaluations had persuaded her organization to better “articulate our goals.” “We have not changed our goals,” she clarified, “but those evaluations made us actually write them up.” Similarly, the head of an HIV support community described how external evaluations prompted internal discussions around “how do *we* document caring?” and “how do *we* evaluate how well we’re providing emotional support [besides saying] ‘That’s one unit of service?’” Some organizations sought to answer these questions by enrolling volunteers and community members to “develop and refine measures of success.” These activities were imperative, explained a mental health service director, because external assessments lacked “validity and reliability.” Accordingly, his organization began developing its own performance measurement tools.

Creating their own measures allowed nonprofits to overcome the inadequacies of externally imposed evaluative criteria. Frustrated by demands for short-term results, the head of a nonprofit advocating for special needs children explained her organization’s efforts to devise criteria based on a longer time horizon: “You’re going to see some impact now, [but] if you

want your socks blown off, it's going to be twenty years." To her, external evaluations were a step in the right direction, but they needed refinement:

Any tools that we can put in the hands of the people so that they have understandings of nonhyped facts to [guide] where they spend their money, the better. Is it a perfect system? Of course not. But it's a step. . . . Are they able to measure exactly what we want to measure? No. . . . Garbage in, garbage out. [But] I think the more tools, even if they're imperfect tools . . . the better. It's still better than hype. We always look at how we can do a better job of reporting what's *actually* happening.

Oversight obligations—however dissonant they were in practice—prompted organizations to reconsider how they pursued long-held commitments to accountability. Had nonprofits resisted both the particular obligations *and* underlying ideals of compulsory oversight, the practices and values of accountability might have been “rendered experientially unusable” (McDonnell et al. 2017, p. 5). Ceremonial compliance would have been the likely result. But nonprofits identified common ground between their values and the ideal of accountability instantiated in external obligations. The resonance of accountability helped make dissonant practices useful; the dissonance of such practices helped cast nonprofit values in new light.

We have already seen examples of this generative dynamic: organizations called for better data, included community members in developing evaluative tools, and inserted qualification into quantitative reports. Pursuing the ideal of accountability required that nonprofits repurpose external verification practices, using them in ways better suited to their organizations' actual circumstances. Without this transformation, accountability would remain, as the director of a boys' choir put it, “just totally spitless, no juice.”

THE TRANSFORMATION OF ACCOUNTABILITY PRACTICES

Even as reporting requirements and external evaluative criteria frustrated organizations, they provided a resource through which nonprofits could pursue locally meaningful conceptions of accountability. In the following, I first explore the expansion of voluntary disclosures on 990s as a window into the broader transformation of accountability that manifested across websites, blogs, reports, and other public documents. Then, examining these multiple forms of organizational communication, I detail two general themes in nonprofits' efforts to thicken otherwise thin approaches to accountability. The first theme centers on the extensive narratives nonprofits used to give voice to beneficiaries and staff, provide the public with insight into decision-making, and foster opportunities for input. The second theme centers on nonprofits' efforts to create bespoke performance standards and secure the help of analysts to assess their work in these terms. As I show, by

2016, the heterogeneous forms of disclosure and evaluative criteria to which nonprofits held themselves accountable both outstripped external obligations and demonstrated nonprofits' distinctive understandings of accountability.

Old Forms, New Purposes

Between 2002 and 2015, nonprofits increasingly used their 990 filings not merely as standard-issue tax forms but also as a medium through which they could enact characteristic understandings of accountability. This change is evident in the growing amount of text and detail organizations provided in the accomplishments section of the form. Historically, these statements were either perfunctory or nonexistent. "Back in the bad old days," one director explained in 2015, "the view [of the 990 was] 'the less said, the better' because it's just time down a rat hole." For example, in 2002, the following text comprised the entirety of an after-school club's statement: "Participatory Recreation; Social Development; Formal Education."

In 2015, however, the same organization's statement included the following text:

This year . . . we served 2,696 youth, ages 6–18 . . . 93% were from low-income households, 68% Hispanic; 16% Asian; 5% Caucasian; 4% multi-racial; 4% African American. . . . We provided 180 days of after-school care, serving 1,806 members. Results of our annual assessment indicate continued impact on the youth we serve . . . 80% of parents reported improvements . . . 94% of members are on-track academically . . . 100% of seniors graduated on time . . . attending the following colleges: CSU East Bay, UCSB, USC . . . We are thrilled to announce that [a student] . . . made history by winning competitions . . . at the state, Pacific region and national levels. This is merely the first step in her life's journey as she pursues her undergraduate degree at USC with a plan to attend law school.

The statement from which the above is excerpted occupies 42 lines of text and incorporates 51 instances of quantification. Such disclosures, the club's director explained, were intended to foster connections between the organization and the broader community while helping to stimulate "a thoughtful conversation about how much work it actually takes to change the cycle of poverty." Through extensive and detailed disclosures like those found on 990s, organizations could make their work "visible" to the broader community and illuminate the complex nature of the issues they sought to address. The director of one organization whose 990 included 120 lines of text and 30 instances of quantification explained, "As popular as we are with the people we touch, there's this huge percentage of the population that has no flipping clue what we do."

When asked why extensive narratives had become a common feature of nonprofit disclosures, accountants explained that the practice was a

widespread reaction to outsiders “taking the numbers and crunching them in ways [nonprofits] don’t like.” Nonprofits began seeing annual filings as “a chance to really describe what they are up to.” Accordingly, accountants left it to nonprofits to prepare their own accomplishments sections. One accountant explained, “This is where we say, ‘this is *your* tax return. It should be a reflection of what *you* want to have on it.’” Exactly what nonprofits write in these sections, she added, “really depends on the people of the organization.” Statements were often prepared collaboratively by staff in charge of various programs and services, organizational leaders who aggregated staff accounts, and board members who provided approval.

Because nonprofits are required to file 990s, the expansion of detailed disclosure on these forms provides a window into the broader transformation of nonprofit accountability. Figure 1 shows that the median number of lines of the accomplishments text increased from five in 2002 to 12 in 2015. Figure 2 shows that the text has become more quantitative. In 2002, the median instances of quantification per line of text was zero. In 2015, the median had grown substantially: 8% of all lines included quantification.¹¹

To be sure, over the entire period studied, official IRS instructions recommended the use of “statistical information” where possible (IRS 2016). Yet, accountants dismissed the idea that this recommendation explained growing disclosure. According to accountants: “The IRS couldn’t care less about that schedule,” “Never in my practice have I ever heard of them questioning what’s on the accomplishments,” and, “This is just for the general public, not for the IRS to enforce.” Emphasizing their point, they referred to additional IRS guidance: “Be clear, concise, and complete in your description.” But nonprofits refused to heed these instructions. Accountants explained that statements “can go on for ten, fifteen pages” and include “giant, giant narratives” and “pages and pages of Schedule O.”

Indeed, accountants shared that the explosion in voluntary disclosure had become a liability: “Sometimes we’ve advised them to cut things down,” one accountant explained. Another felt her professional obligation was to “help them reduce”: “I have advised on some occasions to be less transparent. . . . I had one organization that put an entire transcript of all their meetings, and there’s a lot of stuff that was being said that could be taken out of context. . . . I said ‘You probably should be summarizing some of the things instead of putting entire details.’ . . . There’s stuff in there that could have been a legal issue for them.” Despite advising clients that they “don’t need to do that” and the public “doesn’t need to know that level of detail,” accountants reluctantly

¹¹ Supplementary analyses fail to find statistically significant support for alternative explanations of changes in 990 reporting including resource dependencies, impression management concerns, and practice diffusion through accountants, consultants, and peers. Analyses are available from the author.

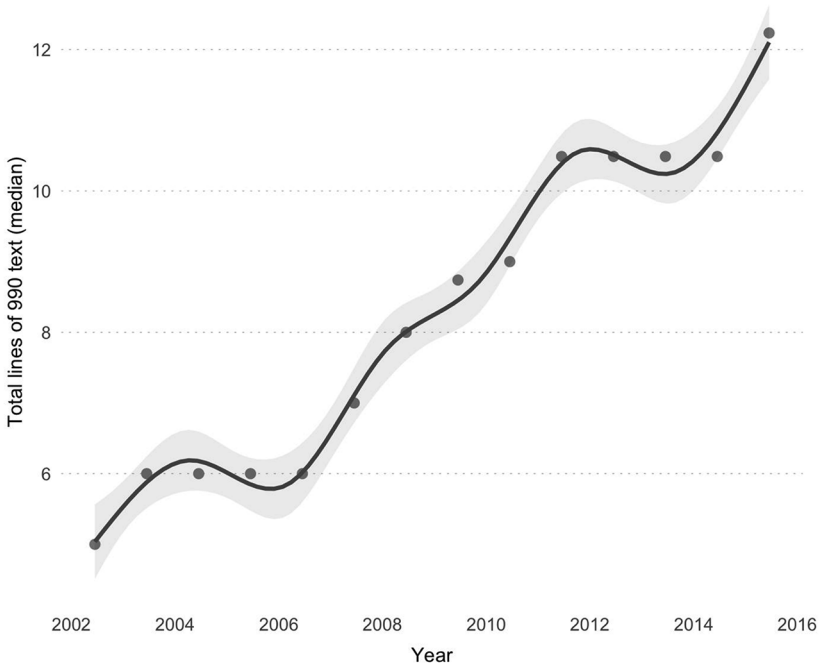


FIG. 1.—Growth in median number of lines of accomplishments text, 2002–15

accepted that, for nonprofits, radical disclosure is “their thing” and “they just [feel] they have to do that.” One accountant characterized nonprofits’ attitudes toward the 990 in the following way: “We’re doing the right thing because it’s the right thing, not because we’re going to get caught [if we don’t].”

“The right thing,” as suggested by the above-quoted after-school club’s comprehensive accomplishments statement, involved detailing the past year’s events, reporting in terms of bespoke evaluative standards, conducting community and beneficiary surveys, and presenting the results. Long frustrated by reductive quantification, the organization also included client stories that warmly illustrated the personal nature of its work. Asked about such mixed modes of disclosure, the director explained, “We have all these stories to tell but then we also know all these percentages.” Stories provided depth; numbers added breadth.

Storytelling

Although 990s reflect broad transformations in nonprofit accountability practices, they are but one of many channels through which organizations

Organizational Supererogation

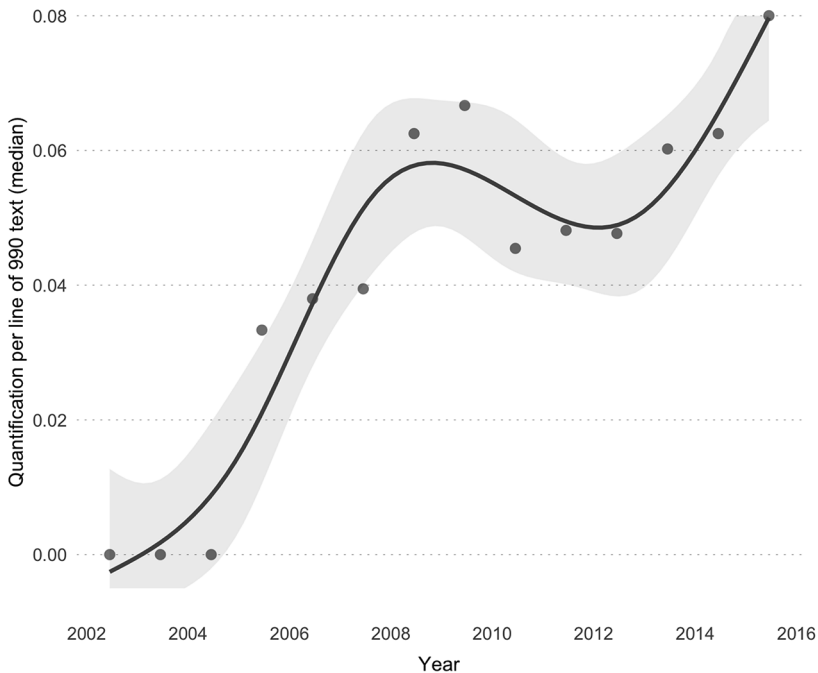


FIG. 2.—Growth in median instances of quantification per line of text, 2002–15. The dip in growth, 2010–12, is a product of the ratio (total instances of quantification divided by total lines of text) represented in the figure. Even as total instances of quantification remained relatively consistent throughout this period, the total amount of text continued to grow.

reported on their work. In fact, the many successes, failures, opportunities, and challenges narrated on these forms suggest another theme in nonprofit accountability: storytelling. By 2016, many nonprofits were sharing extensive information about their activities and amplifying beneficiaries' stories through websites, social media, blogs, newsletters, and 990s. In fact, the once defiant act of wedging personal narratives into quantitative reports became a form of disclosure in its own right. In 2015, the director of a large housing nonprofit in Alameda County explained:

Our lives have taken on this whole return on investment notion. You have to prove a value . . . with these artificial constructs. I count whatever I can and tell the story. . . . You have to really start showing people meaningful examples of what the work does. . . . I can show you thousands of kids that have benefited from this. . . . I look at those kids and I think “Holy shit, this really changed these kids' worlds.” . . . We've been trying to tell this story of our residents and why our work matters.

To tell this story, the housing organization sought to amplify the voices of staff and residents via numerous interviews, essays, photographs, and videos it began posting to its website around 2010. Maintained by staff, the website also aggregated and shared public reviews of its properties, posted otherwise private audits, and offered a live energy-use ticker to let the public observe in real time the organization's environmental commitments.

Such disclosures served multiple purposes. Some organizations, such as a sexual assault counseling center and a homeless services organization, found that providing extensive details about their work offered the broader community "opportunities to encounter" complex issues that otherwise received scant attention. For others, such as a prisoner reentry program and a traveling choir, amplifying the voices of past and current participants gave prospective participants a means of assessing the organization from afar. Indeed, by calling attention to beneficiaries' stories, nonprofits were able to demonstrate meaningful aspects of their organization that were otherwise obscured by external assessments. For example, a refugee resettlement organization demonstrated the knock-on effects of its programs by reporting that clients had begun "teaching other newcomers and helping them navigate their new communities." Similarly, an organization that helped single mothers navigate arcane public assistance requirements reported on its clients' collective efforts to change state welfare policies.

Although storytelling positioned nonprofits to depict themselves in a positive light, many used their websites to voluntarily disclose deficiencies and missteps as well—a radical take on disclosure that "challenge[s] conventional bureaucratic approaches to organizing" (Weber 1978; Turco 2016, p. 165). A Contra Costa housing association described its "deeply disappointing" inability to house low-income families in an aborted development project, attributing the failure to its own poor planning. A theater discussed its shortcomings in handling sexual harassment complaints brought against one of its staff. A school described its failure to curb a lice outbreak. And an organization offering swim therapy reported how poor upkeep of its pool had sent staff scrambling for alternative locations to hold sessions.

Nonprofits also posted detailed meeting minutes online, providing outsiders with an unvarnished perspective on how the organization made decisions. Over a series of meetings, a drama group analyzed results from the community survey it had recently conducted ("people find the older male staff condescending toward women"), debated the pros and cons of raising ticket prices, and grappled with their failure to stage more racially diverse performances ("people have posted [on social media] about the whiteness of casts"; "we can't just sit back saying 'We invited everyone but they didn't show'"; "Sound of Music . . . stress that any race can play these parts"; "do color-blind casting").

Organizational Supererogation

Asked about disclosing less flattering information on his organization's website, a youth center director explained, "We put numbers on there that aren't one hundred percent. A lot of our impact numbers are well below that and we still share them. . . . I consider [this evidence of] a project in progress. . . . We show our room for improvement. We don't select only the good stuff." Remarking on the prevalence of this attitude among nonprofits, one accountant described clients that "put everything they can possibly think of on their website [including a] list of 'here's all our mistakes that we've made over the years, and this is what we learned from them.'"

Nonprofits emphasized that disclosures through various media were part of a two-way dialogue between themselves and the communities they served. For some, frequent e-newsletters opened the door to "feedback from recipients," which, in turn, allowed the organization to "tailor" itself to evolving community needs. Aiming to blur the organizational boundaries presupposed by external audits and inspire more active community discourse about their charitable cause, some organizations invited members, volunteers, and beneficiaries to write blog posts and contribute to social media pages. Such activities allowed nonprofits to substantiate long-standing commitments to porousness, responsiveness, and accountability.

Bespoke Performance Standards and Self-Evaluation

The growing quantification on 990 filings indicates another trend in nonprofit disclosure: self-evaluation. In the spirit of responsiveness, many nonprofits routinized the cycle of disclosure and dialogue, often by conducting surveys of the people and communities they served. As one director explained, "We do biannual surveys to determine the impact of our services for parents, and those responses shape the continuing design of our programs." Such feedback, the head of an organization offering software training workshops explained, helps with "discontinuing programs that don't sufficiently meet the needs of the community at large."

Surveys and feedback also helped nonprofits devise evaluative standards fitted to their particular goals and programs. For example, an organization helping Latinas to launch food co-ops reported drawing on "lessons from years of developing co-ops in low-income communities" to create several "standards and benchmarks" against which to measure the effectiveness of its work. By evaluating themselves against these criteria, staff sought to better serve their community and offer insights to other nonprofits looking to do the same.

As opposed to the ill-fitting, off-the-rack assessments imposed by funders, raters, and other evaluators, many nonprofits collaborated with researchers to carefully tailor assessments to the particulars of their work. In 2015, the director of a family health care organization explained: "[We're] undergoing

a big study right now to prove the effects of [our organization]. . . . It is getting published in peer-reviewed journals, so we've got some hard data to back up the things we've always known in our hearts." By 2016, 52% of sampled nonprofits had either hired researchers to evaluate their programs or expanded a current staff role to include program evaluation. For example, the director of an eldercare organization had grown frustrated that funders and state agencies refused to evaluate his organization's programs. For years, he was repeatedly told, "We're not going to do seniors" because eldercare program effectiveness was too difficult to measure. Yet the director felt his organization owed an explanation to people who "want to know what difference [the programs] make." Under the mantra of "show me the numbers," he hired a full-time researcher from the University of California, San Francisco, to establish performance standards, rigorously assess program effectiveness, and "be a pain in the ass to everybody." By imposing rigorous research methods on its own programs, the organization embraced an accountability regime more demanding than anything external evaluators required.

A large housing organization also hired a researcher from UC San Francisco to do a longitudinal study on its 20-year-old portfolio of properties. "They can actually go back and look at the people still living there and see what happened to their kids," the director explained. Although she was excited to rigorously demonstrate that "the home really matters," she was also keenly aware of the potential downside: "Everybody wants you to prove you're worthy, so hopefully [the study] doesn't prove that we're not worthy. That's the worry, right? Sometimes it's better not to know . . . but you've got to be okay with that." She believed the alternative—burying negative results—was "not super enlightened." If "you get feedback that says [you have] some real issues . . . then you should be grappling with those." Thus, much like storytelling, voluntarily soliciting feedback and opening the organization to rigorous evaluation were not intended to present the organization in a good light. Rather, they were intended to present the organization in the correct light. Importantly, the correct light was distinct to each organization.

The Reproduction of Heterogeneity

Through 990s, websites, blogs, evaluations, and other media, nonprofits began widely disseminating information that was previously available only through face-to-face interactions, direct community oversight, and word of mouth. These disclosures extended far beyond what nonprofits were required to report and revealed the heterogeneity of their missions and circumstances. Heterogeneity manifested both in the diverse combinations of storytelling, self-evaluation, and other approaches to accountability, as

Organizational Supererogation

well as in the plurality of custom evaluative criteria to which nonprofits sought to be held accountable.¹² Examining nonprofits' diverse uses of quantification on their 2015 990s illustrates this point:

A family crisis center reported "6,408 nights of shelter for 96 women and 53 children provided at a 20-bed safe house for women and children fleeing domestic violence."

A homeless services organization reported the number of turkeys and chickens it distributed to families at Thanksgiving and Christmas: 215.

A professional association concerned about the lack of "equal representation for women in the field" distributed 27 scholarships to women.

A mentoring group reported the number of hours staff spent with each first-generation college student: 75–100, depending on circumstances.

A food pantry reported that staff enjoyed 2,516 meals with beneficiaries, that "49% of the food distributed was fresh fruits and vegetables," and that, among the 7,762 people they served, "43% were farmworkers and 35% were children."

A little league reported that "375 girls and boys received baseball uniforms."

Nonprofits also sought to demonstrate effectiveness. An organization helping single mothers attend college reported that "of the parents that graduated, 30% found jobs paying between \$36,000 and \$65,000 per year . . . a family supporting wage." An organization teaching English to recent immigrants shared results from its "social and linguistic isolation survey," proudly reporting "a significant increase in the ability to make appointments, fill out forms, and read instructions."

Not all news was positive, however. Noting mixed results, a reentry program reported that "recidivism is halved for inmates who participate" but there were no statistically significant changes in "feelings of social competence and emotional control." And an organization that links clients with local employers reported that only nine of its 24 business partners hired anyone from its programs.

¹² Combining data from surveys and content analyses, I examined the prevalence of and interrelations between seven general types of accountability practices: narratives on 990s, quantification on 990s, soliciting public feedback, conducting self-evaluations, publicizing internal information, sharing external assessments, and inviting in-person interaction and oversight. I identified 40 unique combinations of these practices. The most common combination, practiced by 18.8% of the sample, included all seven types of accountability. Of the remaining 39 combinations, 36 were performed by less than 3.7% of the sample, with 24 combinations being unique to single organizations. Altogether, 78% of nonprofits pursued at least four general types of accountability, with only 9% pursuing two or fewer types.

Demonstrating responsiveness to community demand, organizations disclosed changes in their ability to serve beneficiaries. A preschool added a second classroom to “differentiate instruction and better fit needs.” Expansions at a women’s treatment facility meant “10 people can now be added.” A health care organization shared that adding eight employees ensured staff would be “awake and available to support families at any hour.”

Some numbers—and, moreover, the narratives accompanying them—spoke to the deeply meaningful nature of organizations’ work. Working to foster independence among intellectually disabled adults, one organization reported that 316 clients were now able to “make their own goals, choices, and decisions.” An ashram reported the number of yogis able to find “spiritual solace.” And an outdoor adventures program that in 2005 wished evaluators could see the smiles on participants’ faces, reported in 2015 that, for 223 participants, a recent trip “was the first time they experienced the magic of snow.”

Even when nonprofits used the same units of analysis, the numbers they reported carried strikingly different valences. This is especially evident when considering, for example, the number of repeat visits per unique beneficiary. A health clinic reported an average of 5.2 repeat visits, a musical theater reported an average of three repeat visits, and a substance abuse treatment center reported an average of 17.1 repeat visits. For the clinic, repeat visits indicated failure; for the theater, they indicated success; for the treatment center, they were a matter of course. Acceptance rates, occupancy rates, accessibility, time spent, and other metrics all revealed a similarly multivalent character.

By enumerating various aspects of their work and using numbers to articulate previously ineffable qualities and moral commitments, nonprofits repurposed quantification as an expression of distinctiveness. This observation contrasts sharply with expectations that evaluative obligations foster conformity between organizations and external criteria of worth (Espeland and Sauder 2016). Indeed, the introduction of oversight obligations was designed to locate all nonprofits along narrow dimensions of accountability: higher overhead is bad; higher throughput is good. To be sure, nonprofits diligently maintained overhead and reported client throughput. But compliance alone was insufficient. It also mattered who these clients were and if they bonded with staff, were treated with dignity, were able to support their families, or had found solace. Through extensive disclosures generally—and quantification specifically—nonprofits reasserted categorical boundaries that evaluative obligations threatened to erode (see Fourcade 2016).

DISCUSSION

In the early 2000s, the rising tide of external oversight marked a sea change in public expectations of nonprofits and nonprofit obligations to the public.

Regulators expanded disclosure requirements, rating agencies assigned scores for financial accountability, and funders made authoritative pronouncements on how well nonprofits were achieving donor-determined outcomes. Participating in this new regime of social ordering was hardly optional, especially given the legal, financial, and social sanctions attached to noncompliance. While nonprofits bristled at the assumptions and activities associated with these obligations, they ultimately identified with the broader ideal of accountability that these demands represented. Rather than minimize external scrutiny through ceremonial compliance or subordinate internal goals to external criteria of worth, nonprofits went above and beyond external demands and practiced accountability in ways that reflected their distinctive goals, circumstances, and understandings of what it meant to be accountable.

To be sure, sociologists have long known that regulations, reporting protocols, and evaluations are powerful instruments that capture attention, shape public expectations, and provoke organizational responses. Yet, by observing the transformation of accountability in the nonprofit sector over the last two decades, this study reveals limitations in prevailing accounts of how such responses unfold.

Contrary to decoupling, nonprofits neither ceremonially conformed to institutional prescriptions nor concealed the raw realities of their work from public view. Strikingly, they invited more oversight and public scrutiny than outside evaluators ever required. They devised evaluative criteria tailored to the particulars of their work, published unvarnished accounts of their decision-making processes, and established new channels for public feedback. Many amplified the voices of staff and beneficiaries whose work and personal experiences had been obscured by ratings and other reductive metrics. Beyond maintaining face through institutionally legitimated “rituals of confidence and good faith” (Meyer and Rowan 1977, p. 360), nonprofits innovated their own rituals of accountability that more accurately reflected their activities. After all, the misalignment between internal and external accounts was one of the very issues they sought to overcome.

Considering a more strategic variant of decoupling, one might see nonprofits’ abundant disclosures as attempts to “instrumentally manipulate” public approval and neutralize reputational threats (Suchman 1995, p. 572). We see such maneuvering, for example, when attorneys disclose extraneous information to opposing counsel in order to “bury adverse evidence under mounds of obfuscating evidentiary debris” (Brazil 1978, p. 1313) or when companies discharge salvos of prosocial claims to defend their public image (McDonnell and King 2013).

But nonprofit disclosures were neither calculated attempts to mollify public concerns nor efforts to obscure organizational realities from public view. As Porter (2012, pp. 594–95) has observed, deceitful organizations

try to “keep the numbers boring,” often using orderly statistical reports to conceal the messiness behind the scenes. By contrast, nonprofits tried to bring the numbers to life. Rather than “select on the good stuff” or move numbers around to “make it look like it fits,” nonprofits sought to portray themselves as they really were. This meant that, alongside stories of graduations, successful workforce reentries, and community celebrations, nonprofits shared *mea culpas* about lice outbreaks, deteriorating facilities, and racial insensitivities. Even accountants—tasked with certifying the truthfulness of organizational claims—worried some of their clients were being too honest in their disclosures.

From another perspective, nonprofits’ widespread embrace of quantification might evince disciplining. After all, many proponents of the early 2000s accountability regime espoused discipline as an explicit goal and hoped that evaluative demands would bring the unruly sector to heel. And critics, fearing exactly this outcome, worried that narrow evaluative frameworks would sap the sector’s eclectic soul. Yet neither of these futures materialized as envisioned.

While nonprofits’ encounters with quantitative oversight certainly transformed how they practiced accountability, they also provoked the innovative reproduction of organizational idiosyncrasies. Nonprofits adapted evaluations to their unique circumstances. They tailored outcome criteria to fit distinctive goals and used surveys, rigorous analyses, and quantitative reporting to demonstrate meaningful aspects of their work in numeric terms. In this sense, quantification became a sort of “contact language” through which diverse understandings of accountability could be communicated—a feat of harmonization without homogenization (Galison 1997, p. 783). Even as quantification became commonplace, the specific things nonprofits counted, how they enumerated them, and what these numbers signified were as heterogeneous as the missions they pursued and the communities they served. *Contra* disciplining, quantification did not erode meaningful boundaries and social relations; it became the medium through which such boundaries and relations were preserved and expressed anew (see Zelizer 1994; Murray 2010).

In sum, attempting to account for the transformation of nonprofit accountability through prevailing theories provides only partial explanations: structural elaboration without decoupling, alternative assertions of worth without deflecting, and reactivity without disciplining. As we have seen, nonprofits were frustrated by the impersonal, abstract, and quantitative demands imposed on them in the early 2000s. These demands distracted from core tasks, were discrepant with how nonprofits appraised their own work, and distorted public understandings of organizational integrity. Despite their frustrations, however, nonprofits identified with the broader ideal of accountability these evaluative obligations were supposed to represent. Indeed,

accountability—in terms of accessibility, oversight, and responsiveness—had long featured in nonprofit activities, albeit in more personal and qualitative ways. Accordingly, nonprofits responded to oversight obligations with neither ceremonial compliance nor comprehensive conformity. While they were diligent in satisfying external demands, they also repurposed aspects of these obligations—such as quantification and routine reporting—using them as means to express the distinctive nature of their work. Through these practices, nonprofits went above and beyond external accountability obligations and pursued the broader ideal of accountability in locally meaningful ways.

This process of supererogation offers a new sociological rendering of how organizations navigate tensions between their values and the demands of their institutional environments. Below, I elaborate on this process by discussing its implications for economic and organizational sociology, considering the settings and conditions under which it is likely to occur, and providing guidance for future research.

CONTRIBUTIONS AND IMPLICATIONS FOR FUTURE RESEARCH

This article makes several contributions to economic and organizational sociology. First, by attending to the dual nature of institutional demands—both as obligations and ideals—supererogation helps overcome a dramatic institutionalist trope treating organizational values as inflexibly at odds with those handed down from the environment. It is important to understand the complex and contradictory ways organizations make sense of external demands. Indeed, organizations may have misgivings about particular obligations while nevertheless embracing the broader ideals these obligations represent. Understanding how organizations navigate such conflicted situations opens analytical possibilities beyond compartmentalization and subversion. This does not invert institutionalism in its entirety but rather illuminates how encounters between internal values and external expectations can produce entirely new activities. There is no question environments weigh heavily on organizations, but we must appreciate how that weight is felt locally to understand how organizations respond to its burden.

Along these lines, future research should examine how internal organizational factors—such as available resources and the plurality of value commitments—condition how organizations make sense of new obligations and their underlying ideals. How might financial reserves, staff capacity, or community support influence the presence or character of supererogatory responses? And how do organizations make trade-offs between potentially contradictory organizational values? For instance, the pursuit of radical transparency among nonprofits occasionally came into conflict with other goals, such as protecting beneficiary privacy. Understanding how organizations resolve these moral tensions—especially when resources are scarce or there is internal

disagreement over values—should offer deeper insights on supererogation's inner workings and illuminate the contingencies through which it occurs.

Second, supererogation demonstrates how efforts to abstract, flatten, and commensurate qualitatively different goals and activities do not necessarily result in interorganizational convergence. This idea is broadly familiar to economic sociologists who argue that monetary exchange need not profane the meaningful qualities of social relations or that being thrust into the global marketplace need not erase local conceptions of authenticity (Zelizer 1994; Wherry 2008). Along these lines, supererogation helps us understand how institutional pressures, rather than eroding meaningful distinctions between organizations, may actually be redeployed in ways that mark and communicate these differences.

The implication here depends on the level of analysis. Considered at the level of the single organization, supererogation explains how stifling demands may provoke new expressions of distinctiveness. Considered at the level of the organizational field, supererogation offers a novel mechanism through which isomorphic pressures catalyze the reproduction of interorganizational heterogeneity (see Beckert 2010; Bromley et al. 2012).

Because supererogation points to a type of organizational change that is difficult to capture through conventional institutionalist methodologies, it presents a challenge to how we conceptualize heterogeneity in sociological research. Analyzing the adoption of formal policies and structures without attending to adjacent activities (which may be highly idiosyncratic) risks the mistaken conclusion that convergence is afoot. The challenge, of course, is that nonstandard practices are, by their very nature, hard to examine with data that impose standardization. Alongside ethnographies and historical approaches, the present study offers one possible model for addressing this methodological challenge.

Whatever the empirical approach, future research should examine the sources of heterogeneity and the processes by which it is reproduced. This article has emphasized the reproduction of fieldwide heterogeneity as a cumulative product of organizational supererogation. But other processes are possible. Heterogeneity may result from *whether* organizations supererogate—as opposed to decouple or become disciplined—in response to shared regulatory pressures.¹³

One might also consider how heterogeneity plays out in relatively homogeneous organizational fields. As the nonprofit case illustrates, supererogation facilitates the *reproduction* of underlying differences in contexts already

¹³ Future research should try to understand why some organizations supererogate while others do not. One possible explanation is that nonsupererogating organizations feel their own conception of a given ideal is adequately captured in external obligations. In such situations, we might expect high resonance but insufficient dissonance to inspire novel practices that exceed external demands.

characterized by heterogeneity. But what about contexts where there is little heterogeneity to speak of? In such settings, we might expect relatively homogenous organizational responses. But organizations are not automata. The different pathways through which ostensibly homogenous organizations make sense of new demands might provoke divergent courses of action. New obligations might differentially foreground internal frictions, mobilize external stakeholders, call attention to dormant values, or trigger reckonings with organizational histories. Conditioned through these channels, supererogation may actually *cultivate* heterogeneity in otherwise homogenous contexts.

Third, supererogation holds implications for sociological accounts of regulation and institutional change. Consider recent developments in how regulators, raters, and funders approach evaluation and accountability in the nonprofit sector. Some raters have recently decried the “overhead myth,” distanced themselves from the financial metrics they previously promoted, and overhauled rating criteria in an attempt to better reflect nonprofits’ evaluative concerns. Foundations have also changed their tune. Some early proponents of donor-centric evaluations now counsel funders to “give [grantees] the benefit of the doubt . . . write the check and get out of the way” (Harvey 2016). And several foundations have rewritten their grant-making policies to ensure that evaluation criteria reflect “the needs and priorities of grantees” (James Irvine Foundation 2020, p. 6). While these developments warrant additional research—and there are certainly many factors at play—one interpretation is that supererogation hampered efforts to establish more unified, top-down criteria for nonprofit accountability.

Along these lines, future research should examine supererogation’s fieldwide effects and regulatory implications, especially in comparison to analogous processes like decoupling and disciplining. For instance, it is conceivable that nonprofits’ improvisational and idiosyncratic responses to regulatory demands will eventually become institutionalized as the way accountability ought to be done.¹⁴ Furthermore, regulators may adapt future demands in light of organizations’ supererogatory actions. The long-run implications of such adaptations, however, are unclear. New regulations may reveal an expanded conception of the ideal in question. But they may also grow increasingly elaborate, making compliance an even more onerous task.

Alternatively, supererogation may indicate regulatory failure. The profusion of idiosyncratic criteria may delegitimize standard-setting bodies, undermine public faith in the intermediaries tasked with governing organizational

¹⁴ Indeed, recent years have seen the proliferation of nonprofit storytelling consultants, and Charity Navigator now urges nonprofits to use their 990s to “provide better context for what you truly do” (Doyle 2019).

action, and foster easily exploited regulatory ambiguities. In the nonprofit sector, where pluralism has long been celebrated as essential to democratic vibrancy, this may be relatively un concerning. But in other domains—for example, mortgage lending, building safety, and drug efficacy—the collapse of common evaluative standards could be catastrophic. Indeed, what looks like autonomy from one perspective may look like anarchy from the perspective of those seeking to impose order.

Supererogation beyond Nonprofits

As the above discussion suggests, it is important to consider supererogation not as unique to the nonprofit sector but as a generalizable process that can shed light on phenomena across numerous domains. But where might one find this process in action? Beyond examining situations where organizational behaviors plainly outstrip the standards to which they are held, researchers might also identify situations where obligations are salient and the ideals they represent align with the values organizations espouse or pursue in practice.

To illustrate, the political turmoil and public health crises of recent years have brought many examples of supererogation to the surface. Consider, for example, how some Title IX offices, dismayed by Trump administration directives on campus sexual violence, have sought to surpass the regulatory “floor” and “reach a higher ‘ceiling’ above and beyond compliance” (National Association of Student Personnel Administrators 2020, p. 2). Or consider how, during the COVID-19 pandemic, some employers have come to regard workplace safety regulations as insufficient and have taken measures far in excess of legal or industry standards to protect workers. For instance, some restaurants, bars, and music venues continue to require masking and social distancing well after ordinances mandating these practices have lapsed. Although there are many ways to make sense of these examples, a supererogatory account would emphasize the fraught interplay between regulatory pressures and organizational values as the catalyst for organizations’ above-and-beyond activities.

Supererogation offers a new lens on familiar research topics as well. Consider, for example, the rich body of literature on organizational responses to antidiscrimination laws and bureaucratic diversity initiatives. Supererogation might explain why some organizations take additional steps toward meaningfully addressing workplace inequities while others do the bare minimum to pass muster (e.g., Kalev, Dobbin, and Kelly 2006). Additionally, we might see supererogation at work in the historical sociology of money, credit, and finance. For instance, Polillo (2011) demonstrates how bankers in the late 1800s transcended newly established federal lending standards by innovating

distinctive financial instruments imbued with locally meaningful conceptions of creditworthiness.¹⁵

It is important to note that supererogation can produce objectionable outcomes as well. Consider, for example, how police precincts may respond to the imposition of department-wide statistical monitoring. They may resist the top brass's assertions of control while nevertheless identifying with an ideal of aggressive policing enshrined in the new requirements. In the abstract, supererogation would operate much like it did among nonprofits: unwelcome directives and resonant ideals resulting in action that goes above and beyond what is required. In the concrete, however, the outcomes are markedly different. Whereas nonprofit supererogation results in new forms of community responsiveness, police supererogation results in new forms of violence against communities of color. The point here is that, unlike the philosophical conception of supererogation (where going above and beyond is morally praiseworthy), the sociological conception is not limited to any normative valence. Supererogation may help to address chronic inequities. But it may exacerbate them as well.

Supererogation need not be limited to the realm of organizational action either. Insofar as people are subject to regulations, standards, evaluations, and ratings, one might consider whether or how supererogation unfolds at the individual level. We might see this, for example, in the rules that govern our own research. Many sociologists likely believe that IRB obligations are well intended but nevertheless find the requirements irksome and inadequate for protecting research subjects. No doubt there are some scholars who go beyond merely complying with IRB stipulations to pursue the ideal of ethical research as they understand it in the context of their work. More generally, we might expect to find supererogation in settings where professional judgment is burdened with one-size-fits-all regulations. Furthermore, one might find shades of supererogation across numerous settings including contractual relationships (Why do some parties go above and beyond the expectations agreed to on paper?), systems of rating individual trustworthiness (How have people responded to the social credit system in China?), and social movements (How do activists repurpose repressive measures or token concessions as tools for advancing their cause?).

Across these examples—both organizational and individual—the common thread is this: when there are salient obligations and there are people that identify with the ideals these obligations represent, supererogation is possible. Even in the absence of actions that go above and beyond external demands,

¹⁵ Historical cases notwithstanding, the intersection of two trends in recent decades—the expansion of formal oversight (Power 1997) and empowered assertions of individual and organizational autonomy (Meyer and Jepperson 2000)—suggest that supererogation may be a characteristically modern phenomenon.

supererogation should sensitize us to an important duality. The tasks people are required to perform (however begrudgingly) represent broader aspirations with which they may identify. Further exploring this complexity—how people are simultaneously attracted to and repelled by the demands of their social environment—promises new insights for economic and organizational sociology and for other sociological subfields as well.

CONCLUSION

One of the fundamental puzzles of organizational sociology is “why there is such startling homogeneity of organizational forms and practices” (DiMaggio and Powell 1983, p. 148). The question posed here is why, given the array of formidable, homogenizing forces pressing upon organizations, do we continue to see such vibrant heterogeneity? Contrary to the expectations of contemporary organizational theory, the organizations I followed neither superficially nor substantively conformed to the intrusions of their environment. Faced with extraordinary pressures to project a polished public image, nonprofits performed rituals of self-purification and demonstrated fallibility. While they certainly complied with external demands, they also went above and beyond these demands in pursuit of organizationally distinctive conceptions of accountability. Through the process of supererogation, the stultifying slide rules expected to induce acquiescence and conformity instead provoked innovations and assertions of individuality.

According to the French historian and philosopher Michel de Certeau (1984, p. xiv), if the “grid of discipline is everywhere becoming clearer and more extensive,” it is necessary to understand how people and organizations transform these pressures in ways that allow them to pursue “their own interests and their own rules.” Put simply: without rules there is no play. Today, all social entities—bodies, families, organizations, and states—are subject to rigorous evaluation, quantification, and comparison. Although scholars have long viewed rationalization and calculation as socially corrosive, it is important to recognize that iron cages, gray flannel suits, and homogenization are not the inevitable result. These monochromatic effects are eminently contingent. The challenge for sociologists is to understand how, despite the enormous power such pressures wield over social life, people and organizations redeploy these forces in ways that facilitate their distinctive pursuits.

REFERENCES

- Barley, Stephen R. 1986. “Technology as an Occasion for Structuring: Evidence from Observations of CT Scanners and the Social Order of Radiology Departments.” *Administrative Science Quarterly* 31:78–108.
- Barman, Emily. 2016. *Caring Capitalism*. New York: Cambridge University Press.

Organizational Supererogation

- Barnard, Alex V. 2016. "Making the City 'Second Nature': Freegan 'Dumpster Divers' and the Materiality of Morality." *American Journal of Sociology* 121 (4): 1017–50.
- Beckert, Jens. 2010. "Institutional Isomorphism Revisited: Convergence and Divergence in Institutional Change." *Sociological Theory* 28 (2): 150–66.
- Benjamin, Lehn. 2012. "Nonprofit Organizations and Outcome Measurement: From Tracking Program Activities to Focusing on Frontline Work." *American Journal of Evaluation* 33 (3): 431–47.
- Bourdieu, Pierre. 1990. *The Logic of Practice*. Stanford, Calif.: Stanford University Press.
- Brandtner, Christof. 2017. "Putting the World in Orders: Plurality in Organizational Evaluation." *Sociological Theory* 35 (3): 200–227.
- Brazil, Wayne D. 1978. "The Adversary Character of Civil Discovery: A Critique and Proposals for Change." *Vanderbilt Law Review* 31 (6): 1296–348.
- Brest, Paul. 2020. "The Outcomes Movement in Philanthropy and the Nonprofit Sector." Pp. 381–408 in *Nonprofit Sector*, 3rd ed. Edited by Walter W. Powell and Patricia Bromley. Stanford, Calif.: Stanford University Press.
- Bromley, Patricia, Hokyung Hwang, and Walter W. Powell. 2012. "Decoupling Revisited: Common Pressures, Divergent Strategies in the U.S. Nonprofit Sector." *M@n@gement* 15 (5): 469–501.
- Bromley, Patricia, and Walter W. Powell. 2012. "From Smoke and Mirrors to Walking the Talk: Decoupling in the Contemporary World." *Academy of Management Annals* 6 (1): 483–530.
- Brunsson, Nils. 1993. "Ideas and Actions: Justification and Hypocrisy as Alternatives to Control." *Accounting, Organizations and Society* 18 (6): 489–506.
- Carman, Joanne. 2007. "Evaluation Practice among Community-Based Organizations: Research into the Reality." *American Journal of Evaluation* 28 (1): 60–75.
- Center for Effective Philanthropy. 2002. "Indicators of Effectiveness." <https://cep.org/portfolio/indicators-of-effectiveness-understanding-and-improving-foundation-performance-2/>.
- Charity Navigator. 2004. "How Do We Rate Charities?" <https://web.archive.org/web/20050313024224/http://www.charitynavigator.org/index.cfm/bay/content.view/catid/2/cpid/35.htm>.
- Charity Watch. 2005. "Criteria." <https://web.archive.org/web/20050301085321/http://www.charitywatch.org/criteria.html>.
- Christensen, Lars Thøger, and Joep Cornelissen. 2015. "Organizational Transparency as Myth and Metaphor." *European Journal of Social Theory* 18 (2): 132–49.
- Christin, Angèle. 2018. "Counting Clicks: Quantification and Variation in Web Journalism in the United States and France." *American Journal of Sociology* 123 (5): 1382–415.
- Cyert, Richard M., and James G. March. 1963. *A Behavioral Theory of the Firm*. Englewood Cliffs, N.J.: Prentice-Hall.
- de Certeau, Michel. 1984. *The Practice of Everyday Life*. Berkeley: University of California Press.
- DiMaggio, Paul J. 2002. "Measuring the Impact of the Nonprofit Sector on Society Is Probably Impossible but Possibly Useful." Pp. 249–72 in *Measuring the Impact of the Nonprofit Sector*, edited by P. Flynn and V. Hodgkinson. New York: Kluwer.
- DiMaggio, Paul J., and Walter W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review* 48 (2): 147–60.
- Dobbin, Frank, and Erin L. Kelly. 2007. "How to Stop Harassment: Professional Construction of Legal Compliance in Organizations." *American Journal of Sociology* 112 (4): 1203–43.
- Doyle, Kevin. 2019. "Telling Your Story Using the 990." Charity Navigator Blog. <https://blog.charitynavigator.org/2019/08/telling-your-story-using-990.html>.
- Ebrahim, Alnoor. 2005. "Accountability Myopia: Losing Sight of Organizational Learning." *Nonprofit and Voluntary Sector Quarterly* 34 (1): 56–87.

- Edelman, Lauren B. 1992. "Legal Ambiguity and Symbolic Structures: Organizational Mediation of Civil Rights Law." *American Journal of Sociology* 97 (6): 1531–76.
- Espeland, Wendy N., and Michael Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." *American Journal of Sociology* 113 (1): 1–40.
- . 2016. *Engines of Anxiety*. New York: Russell Sage Foundation.
- Espeland, Wendy Nelson, and Berit Irene Vannebo. 2007. "Accountability, Quantification, and Law." *Annual Review of Law and Social Science* 3 (1): 21–43.
- Feinberg, Joel. 1961. "Supererogation and Rules." *Ethics* 71 (4): 276–88.
- Feldman, Martha S., and James G. March. 1981. "Information in Organizations as Signal and Symbol." *Administrative Science Quarterly* 26 (2): 171–86.
- Fleishman, Joel L. 2001. "Public Trust in Not-for-Profit Organizations and the Need for Regulatory Reform." Pp. 172–97 in *Philanthropy and the Nonprofit Sector in a Changing America*, edited by Charles T. Clotfelter and Thomas Ehrlich. Bloomington: Indiana University Press.
- Foucault, Michel. 1977. *Discipline and Punish*. New York: Vintage.
- Fourcade, Marion. 2011. "Cents and Sensibility: Economic Valuation and the Nature of Nature." *American Journal of Sociology* 116 (6): 1721–77.
- . 2016. "Ordinalization." *Sociological Theory* 34 (3): 175–95.
- Galison, Peter. 1997. *Image and Logic*. Chicago: University of Chicago Press.
- Giddens, Anthony. 1984. *The Constitution of Society*. Berkeley: University of California Press.
- Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. New York: Doubleday.
- Hallett, Tim. 2010. "The Myth Incarnate: Recoupling Processes, Turmoil, and Inhabited Institutions in an Urban Elementary School." *American Sociological Review* 75 (1): 52–74.
- Hammack, David. 1995. "Accountability and Nonprofit Organizations: A Historical Perspective." *Nonprofit Management and Leadership* 6 (2): 127–39.
- Harvey, Hal. 2016. "Why I Regret Pushing Strategic Philanthropy." *Chronicle of Philanthropy*. <https://www.philanthropy.com/article/why-i-regret-pushing-strategic-philanthropy/>.
- Hernández, Georgiana, and Mary Visher. 2001. "Creating a Culture of Inquiry." Report. James Irvine Foundation, Los Angeles.
- Heyd, David. 2019. "Supererogation." In *Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta. <https://plato.stanford.edu/entries/supererogation>.
- Horvath, Aaron, and Walter W. Powell. 2020. "Seeing Like a Philanthropist: From the Business of Benevolence to the Benevolence of Business." Pp. 81–122 in *Nonprofit Sector*, 3rd ed. Edited by Walter W. Powell and Patricia Bromley. Stanford, Calif.: Stanford University Press.
- House Committee on Energy and Commerce. 2001. *Charitable Contributions for September 11: Protecting against Fraud, Waste, and Abuse*. November 6. H.R. Rep. 107-67.
- Hwang, Hokyū, and Walter W. Powell. 2009. "The Rationalization of Charity: The Influences of Professionalism in the Nonprofit Sector." *Administrative Science Quarterly* 54:268–98.
- IRS (Internal Revenue Service). 2016. "2015 Instructions for Form 990 Return of Organization Exempt from Income Tax." <https://www.irs.gov/pub/irs-prior/1990-2015.pdf>.
- James Irvine Foundation. 2020. "Impact Assessment and Learning Framework." Report. James Irvine Foundation, Los Angeles.
- Jerolmack, Colin, and Shamus Khan. 2014. "Talk Is Cheap: Ethnography and the Attitudinal Fallacy." *Sociological Methods and Research* 43 (2): 178–209.
- Kalev, Alexandra, Frank Dobbin, and Erin Kelly. 2006. "Best Practices or Best Guesses? Assessing the Efficacy of Corporate Affirmative Action and Diversity Policies." *American Sociological Review* 71 (4): 589–617.
- Kiviat, Barbara. 2019. "The Moral Limits of Predictive Practices: The Case of Credit-Based Insurance Scores." *American Sociological Review* 84 (6): 1134–58.

Organizational Supererogation

- Lampkin, Linda M., Mary K. Winkler, Janelle Kerlin, Harry P. Hatry, et al. 2006. "Building a Common Outcome Framework to Measure Nonprofit Performance." Report. Urban Institute, Washington, D.C.
- Lévi-Strauss, Claude. 1987. *Introduction to the Work of Marcel Mauss*. Translated by F. Baker. London: Routledge & Kegan Paul.
- Levine, Jeremy R. 2017. "The Paradox of Community Power: Cultural Processes and Elite Authority in Participatory Governance." *Social Forces* 95 (3): 1155–79.
- MacIndoe, Heather, and Emily Barman. 2012. "How Organizational Stakeholders Shape Performance Measurement in Nonprofits: Exploring a Multidimensional Measure." *Nonprofit and Voluntary Sector Quarterly* 42 (4): 716–38.
- McDonnell, Mary-Hunter, and Brayden King. 2013. "Keeping Up Appearances: Reputational Threat and Impression Management after Social Movement Boycotts." *Administrative Science Quarterly* 58 (3): 387–419.
- McDonnell, Terence E. 2010. "Cultural Objects as Objects: Materiality, Urban Space, and the Interpretation of AIDS Campaigns in Accra, Ghana." *American Journal of Sociology* 115 (6): 1800–1852.
- McDonnell, Terence E., Christopher A. Bail, and Iddo Tavory. 2017. "A Theory of Resonance." *Sociological Theory* 35 (1): 1–14.
- McMullan, Bernard, and Patricia Patrizi. 2001. "Evaluation in Foundations: The Unrealized Potential." Report. Kellogg Foundation Evaluation Unit, Battle Creek, Mich.
- Mellema, Gregory. 1991. *Beyond the Call of Duty*. Albany: State University of New York Press.
- Mennicken, Andrea, and Wendy N. Espeland. 2019. "What's New with Numbers? Sociological Approaches to the Study of Quantification." *Annual Review of Sociology* 45 (1): 1–23.
- Meyer, John W., and Ronald L. Jepperson. 2000. "The 'Actors' of Modern Society: The Cultural Construction of Social Agency." *Sociological Theory* 18 (1): 100–120.
- Meyer, John W., and Brian Rowan. 1977. "Institutionalized Organizations: Formal Structure as Myth and Ceremony." *American Journal of Sociology* 83 (2): 340–63.
- Miller, Peter, and Michael Power. 2013. "Accounting, Organizing, and Economizing: Connecting Accounting Research and Organization Theory." *Academy of Management Annals* 7 (1): 557–605.
- Mohr, John W. 1994. "Soldiers, Mothers, Tramps and Others: Discourse Roles in the 1907 New York City Charity Directory." *Poetics* 22 (4): 327–57.
- Mosley, Jennifer E., Nicole P. Marwell, and Marci Ybarra. 2019. "How the 'What Works' Movement Is Failing Human Service Organizations, and What Social Work Can Do to Fix It." *Human Service Organizations: Management, Leadership and Governance* 43 (4): 326–35.
- Muller, Jerry Z. 2018. *The Tyranny of Metrics*. Princeton, N.J.: Princeton University Press.
- Murray, Fiona. 2010. "The Oncomouse That Roared: Hybrid Exchange Strategies as a Source of Distinction at the Boundary of Overlapping Institutions." *American Journal of Sociology* 116 (2): 341–88.
- Murray, Vic. 2001. "The State of Evaluation Tools and Systems for Nonprofit Organizations." *New Directions for Philanthropic Fundraising* 31:39–50.
- National Association of Student Personnel Administrators. 2020. "Title IX Guide: Going above and beyond Compliance." <https://naspa.org/book/beyond-compliance>.
- Padgett, John F., and Walter W. Powell. 2012. *The Emergence of Organizations and Markets*. Princeton, N.J.: Princeton University Press.
- Panel on the Nonprofit Sector. 2005. *Strengthening Transparency, Governance, Accountability, of Charitable Organizations*. Washington, D.C.: Panel on the Nonprofit Sector.
- Polillo, Simone. 2011. "Money, Moral Authority, and the Politics of Creditworthiness." *American Sociological Review* 76 (3): 437–64.

American Journal of Sociology

- Porter, Theodore. 2012. "Funny Numbers." *Culture Unbound: Journal of Current Cultural Research* 4:585–98.
- Power, Michael. 1997. *The Audit Society*. Oxford: Oxford University Press.
- Quinn, Sarah. 2008. "The Transformation of Morals in Markets: Death, Benefits, and the Exchange of Life Insurance Policies." *American Journal of Sociology* 114 (3): 738–80.
- Reich, Adam D. 2014. "Contradictions in the Commodification of Hospital Care." *American Journal of Sociology* 119 (6): 1576–628.
- Reiser, Dana B. 2005. "There Ought to Be a Law: The Disclosure Focus of Recent Legislative Proposals for Nonprofit Reform." *Chicago-Kent Law Review* 80:559–612.
- Rose, Nikolas, and Peter Miller. 1992. "Political Power beyond the State: Problematics of Government." *British Journal of Sociology* 43 (2): 173–205.
- Sauder, Michael, and Wendy N. Espeland. 2009. "The Discipline of Rankings: Tight Coupling and Organizational Change." *American Sociological Review* 74 (1): 63–82.
- Schneiberg, Marc, and Tim Bartley. 2008. "Organizations, Regulation, and Economic Behavior: Regulatory Dynamics and Forms from the Nineteenth to Twenty-First Century." *Annual Review of Law and Social Science* 4 (1): 31–61.
- Selznick, Philip. 1957. *Leadership in Administration*. New York: Harper & Row.
- Sewell, William H., Jr. 1992. "A Theory of Structure: Duality, Agency, and Transformation." *American Journal of Sociology* 98 (1): 1–29.
- Sharkey, Amanda J., and Patricia Bromley. 2015. "Can Ratings Have Indirect Effects? Evidence from the Organizational Response to Peers' Environmental Ratings." *American Sociological Review* 80 (1): 63–91.
- Soskis, Benjamin, and Stanley N. Katz. 2016. "Looking Back at 50 Years of Philanthropy." Report. William and Flora Hewlett Foundation, Menlo Park, Calif.
- Stark, David. 2009. *The Sense of Dissonance*. Princeton, N.J.: Princeton University Press.
- Strathern, Marilyn. 2000. *Audit Cultures*. New York: Routledge.
- Suchman, Mark. 1995. "Managing Legitimacy: Strategic and Institutional Approaches." *Academy of Management Review* 20 (3): 571–610.
- Swords, Peter. 1998. "The Form 990 as an Accountability Tool for 501(c)(3) Nonprofits." *Tax Lawyer* 51 (3): 571–618.
- Timmermans, Stefan, and Steven Epstein. 2010. "A World of Standards but Not a Standard World: Toward a Sociology of Standards and Standardization." *Annual Review of Sociology* 36:69–89.
- Turco, Catherine. 2012. "Difficult Decoupling: Employee Resistance to the Commercialization of Personal Settings." *American Journal of Sociology* 118 (2): 380–419.
- . 2016. *The Conversational Firm*. New York: Columbia University Press.
- Urmson, J. O. 1958. "Saints and Heroes." Pp. 198–216 in *Essays in Moral Philosophy*, edited by A. I. Melden. Seattle: University of Washington Press.
- Weber, Max. 1978. *Economy and Society*, edited by Guenther Roth and Claus Wittich. Berkeley: University of California Press.
- Wherry, Frederick F. 2008. *Global Markets and Local Crafts*. Baltimore: Johns Hopkins University Press.
- White, Harrison C. 2008. *Identity and Control*. Princeton, N.J.: Princeton University Press.
- Wing, Kennard, and Mark Hager. 2004. "Getting What We Pay For." Report. Urban Institute, Washington, D.C.
- Wise Giving Alliance. 2003. "Standards for Charity Accountability." <https://web.archive.org/web/20050614074951/http://www.give.org/standards/newcbbstds.asp>.
- Zelizer, Viviana A. 1978. "Human Values and the Market: The Case of Life Insurance and Death in 19th-Century America." *American Journal of Sociology* 84 (3): 591–610.
- . 1994. *The Social Meaning of Money*. Princeton, N.J.: Princeton University Press.