trends in higher education

MDPI

Article

Institutional Contradictions and the Reproduction of Service Devaluation in Higher Education

Pinar Ozturk [1], Lori E. Koelsch [2], Rachael Miller Neilan [3,*] and Misook Heo [3]

[1] Department of Accounting, Information Systems and Technology, and Supply Chain Management, Palumbo-Donahue School of Business, Duquesne University, Pittsburgh, PA 15282, USA; ozturkp@duq.edu
[2] Department of Psychology, McAnulty College of Liberal Arts, Duquesne University, Pittsburgh, PA 15282, USA; koelschl@duq.edu
[3] Department of Mathematics and Computer Science, School of Science and Engineering, Duquesne University, Pittsburgh, PA 15282, USA; heom@duq.edu
* Correspondence: millerneilanr@duq.edu

Abstract

Universities depend on faculty service to support teaching, research, and everyday institutional functioning, yet their evaluation systems devalue this work through policy and practice. Although the prioritization of teaching and research over service is often deliberate, it creates a paradox in which institutions rely on work that their reward systems give faculty limited incentive to perform. Drawing on interviews with nine former members of a university-level promotion and tenure committee, this study examines how evaluators navigate contradictory institutional demands that require service while providing no career incentive for service excellence. The analysis reveals four mechanisms through which this paradox is reproduced: (1) handbook requirements create binds between institutional needs and evaluation criteria; (2) evaluation procedures marginalize service through time allocation and sequencing; (3) the absence of institutional standards, which leads committee members to rely on personal judgment; and (4) well-intentioned adaptations reinforce devaluation and uneven distribution of service. These findings demonstrate how implementing an established evaluative hierarchy can generate consequences beyond those specified in formal policy, affecting institutions' abilities to sustain the work on which they depend.

Keywords: faculty service; promotion and tenure; faculty evaluation; institutional contradictions; organizational paradox

Check for updates

Academic Editors: Panayiotis Antoniou and Maria Eliophotou Menon

Received: 24 September 2026
Revised: 18 October 2026
Accepted: 25 October 2026
Published: 2 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Although universities depend on faculty service for everyday functioning, this work is systematically devalued by structures used to assess faculty performance. Committees must be staffed, curricula developed, students mentored, and institutions governed, but these essential activities rarely carry weight comparable to teaching and research in promotion and tenure (P&T) systems [1–3]. Although service is formally recognized as one of the "three pillars" of faculty work, institutional reward systems consistently privilege research productivity and, to a lesser extent, teaching effectiveness [4,5]. This hierarchy is well established and often deliberate. Yet the teaching and research that institutions prioritize also depend on service, including peer review, editorial work, faculty hiring, and the P&T review process itself. Faculty therefore have limited incentives to engage in service beyond what advancement requires, even as institutions depend on these contributions. This tension

between universities' dependence on faculty service and the limited weight it carries in advancement creates a central contradiction in academic life [6].

Prior research has documented unequal service distribution across gender, race, and rank, showing that women and faculty of color disproportionately perform institutional "housework" that carries limited career rewards [7–9]. Yet most research on faculty service focuses on those who perform service rather than on those assessing it. As a result, we know less about how evaluators apply service criteria in practice and what consequences emerge when service is assigned limited weight in advancement without clear standards for evaluating its quality and impact. This gap is important because institutional contradictions do not persist through formal policies alone; they are enacted through the routines, interpretations, and judgments of actors charged with applying those policies.

This study addresses that gap by examining how P&T committee members enact and reproduce institutional contradictions surrounding service. We draw on semi-structured interviews with nine former members of a university-level P&T committee at a Catholic research university. P&T committee members are critical institutional actors because they translate handbook language, procedural routines, informal norms, and institutional values into evaluative judgments. Their accounts therefore provide evidence of how service devaluation becomes embedded in evaluation practice, rather than simply how it is experienced by faculty who perform service.

The institution's explicit commitment to service provides a particularly useful context for examining how service is evaluated in practice. At this Catholic university, service is emphasized as a core institutional value, creating potential tensions between institutional values and evaluative structures through which faculty work is recognized and rewarded [10,11]. This mission makes the tension especially visible, although the dependence on service is revealed to be common across higher education. We ask the following: how are the institutional contradictions surrounding service enacted and reproduced through the evaluative practices of P&T committee members, and what consequences emerge from these practices?

By analyzing these evaluative practices, we find four linked practices through which service becomes underspecified, procedurally marginalized, informally standardized, and reproduced as a devalued category of faculty work. Theoretically, this paper extends organizational paradox scholarship by showing how contradictions in academic evaluation are sustained not only through formal policies, but through the routine evaluative practices by which these policies are enacted [12,13]. Empirically, it provides an analysis grounded in the perspectives of P&T committee members positioned at the intersection of policy, culture, and practice. Practically, it offers insights for institutions seeking to clarify service criteria, support more consistent P&T evaluation and align faculty reward structures with espoused commitments to service, equity, and community engagement.

2. Literature Review

The systematic devaluation of faculty service reflects a core contradiction in higher education: institutions rely on service to function, yet reward systems rarely treat it as comparable to teaching or scholarship. To explain how this pattern persists, we first review scholarship documenting service as both essential and undervalued; then, we draw on organizational theory to conceptualize this devaluation as an institutional contradiction; finally, we locate its reproduction in the evaluative practices of P&T committee members.

2.1. Service as Necessary but Devalued

By the mid-twentieth century, teaching, scholarship, and service had become widely institutionalized as the three nominal pillars of academic careers [1,4]. Within this framework, service represents one of the core obligations of academic citizenship, encompassing

activities such as peer review, editorial work, committee service, institutional governance, and broader contributions to students, the institution, the profession, and the community [3,14]. Yet the formal inclusion of service alongside teaching and scholarship has not produced equal evaluative standing. In practice, reward systems consistently privilege research productivity, positioning service as necessary to institutional functioning but marginal to individual advancement [5].

Service is not only lower in this evaluative hierarchy but also inadequately specified. Compared to teaching and scholarship, expectations for service are often described in broad or ambiguous terms [15], leaving committees to interpret vague standards without clear criteria of quality or impact. Although reform efforts have sought to strengthen the role of service within faculty evaluation, its devaluation has remained remarkably persistent [16,17].

The costs of this arrangement are unevenly distributed. Female faculty perform more service on average than their male colleagues, particularly in student support and institutional caretaking [7,18], and faculty of color are frequently recruited for diversity-related work that can impede research productivity [8,19]. Much of this is "hidden service": institutional maintenance work that is typically absent from CVs and often unrewarded despite being foundational to institutional functioning [9,20]. Because such service often involves care and relationship-building that defies easy quantification, it is particularly vulnerable to devaluation [21]. Heavy service burdens are connected to early career dissatisfaction [22], intent to leave [23], and lower job satisfaction.

At the same time, faculty are not simply passive in the face of these incentives. Some continue to engage in service despite knowing it is undervalued, treating it as a professional commitment rather than strategic calculation. Faculty of color may engage in race-related service as a form of critical agency that resists and redefines the structures constraining them [24], and female faculty perform intellectual and boundary-spanning work that both serves the institution and advances their own scholarly aims [25,26]. This agency matters because it shows that reward structures do not fully determine faculty behavior, and that the persistence of service despite its devaluation is itself something a study of evaluation must explain.

These dynamics also vary by type of service, with internal, institution-facing service recognized differently from external, professional service, though both are typically undervalued relative to research [9,27]. Taken together, this literature documents that service is essential yet devalued and that its burden falls unequally. What it explains less well is why the devaluation persists even where its unfairness is widely recognized. These patterns suggest that service devaluation is not simply a discrete evaluative problem but also an institutional contradiction.

2.2. Service Devaluation as Institutional Contradiction

Understanding how service can be simultaneously required and devalued calls for a framework attentive to competing demands within organizations. The institutional logics perspective holds that organizations often operate under multiple, potentially incompatible value systems and evaluation criteria that create tensions requiring ongoing management [11]. In faculty evaluation, these logics include commitments to research excellence, teaching quality, and meaningful service, even when only some of these commitments are structurally rewarded. As universities have increasingly emphasized research productivity and external funding, evaluation systems (such as P&T and annual reviews) have adapted to reward measurable scholarly outputs while struggling to articulate parallel standards for teaching and service [2]. This has created a form of mission drift where

rhetoric about balanced evaluation diverges from actual reward structures that prioritize research achievement.

Neo-institutional theory helps explain how such divergence can persist. Organizations often adopt formal structures for symbolic legitimacy rather than functional alignment, creating gaps between policy and practice [10]. This loose coupling allows institutions to maintain legitimacy while accommodating competing pressures through informal adaptations. Formal commitments to service can thus coexist with reward practices that discount it, without the contradiction ever being openly resolved.

Evaluation systems can also reproduce contradictions through the incentives and indicators they privilege. Campbell's Law [28] suggests that "the more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor" (p. 49). This principle is especially relevant for understanding how faculty evaluation systems may undermine the very activities they purport to measure and reward.

Studies of performance measurement have documented evaluation systems creating perverse incentives. In K-12 education, high-stakes testing has been shown to narrow curricula, encourage teaching to the test, and widen achievement gaps rather than improve learning [29]. Similarly, hospital quality metrics have been shown to incentivize avoiding difficult cases rather than improving care quality [30]. More broadly, when evaluation criteria fail to capture the full complexity of desired activities, individuals adapt by focusing effort on measurable aspects while neglecting unmeasured but important dimensions of performance [31]. Yet this explains only part of the pattern. Faculty frequently continue to perform service despite its limited rewards, as the agency literature [24–26] makes clear, so incentive distortion alone cannot account for how devaluation is sustained. The contradiction persists not simply because individuals respond to incentives, but because it is embedded in the structures and practices through which evaluation occurs.

Scholarship on organizational paradoxes offers a way to hold these elements together. Smith and Lewis [13] define organizational paradoxes as "contradictory yet interrelated elements that exist simultaneously and persist over time" (p. 382) and characterize "performing paradoxes" as tensions between competing performance demands requiring ongoing management rather than resolution. Service and the reward systems that discount it are interrelated in this sense because the teaching and scholarship those systems prioritize depend on service activities, including the curriculum development, peer review, and institutional governance that support academic work. From this view, universities can depend on faculty service while structurally discouraging service excellence—not as a temporary misalignment, but as an embedded paradox requiring continuous navigation. Such paradoxes are constituted through ongoing practice. Putnam et al. [12] argue that seemingly neutral organizational procedures, such as evaluation protocols, discussion sequences, and time allocations, can actively create and maintain paradoxical tensions, while attempts to reconcile them often result in symbolic compliance that preserves underlying tensions [32].

The concept of institutional work further emphasizes how individual actors actively maintain institutional arrangements even while recognizing their problems through everyday practices of interpretation, application, and legitimation [33]. In faculty evaluation, this means that service devaluation is not sustained by policy language alone; it is reproduced through the practices of those who apply and legitimate evaluation criteria. In promotion and tenure, this work falls to committee members.

2.3. P&T Committee Members as Policy Enactors

Formal policies are not simply implemented; they are enacted by the actors responsible for translating policy language into situated judgments [34]. In P&T, committee members

occupy this enactment role. They translate handbook language, procedural routines, and informal norms into judgments about what counts as evidence of faculty contribution. Their accounts therefore provide evidence of how formal policy becomes evaluative practice within disciplinary culture of academic judgment [35].

This enactment role becomes especially consequential where service criteria are thinly specified. When handbooks offer little guidance on what service quality or impact looks like, committee members are left to supply the missing standards, and their judgments can turn an ambiguous formal commitment into a concrete determination of merit. In these conditions, the institutional contradiction is not resolved so much as worked out case by case, through the evaluative practices committee members use to assess, weigh, and legitimate service.

Despite extensive documentation of service devaluation and its consequences, existing research has emphasized faculty experiences as service providers rather than how evaluators navigate the contradictions within service evaluation systems. While P&T committees often struggle with service evaluation [36], limited research has examined how committee members enact institutional requirements that simultaneously depend on and devalue faculty service. This gap matters because P&T committee members do not simply observe these contradictions; they participate in translating that contradiction into evaluative judgments, procedural routines, informal standards, advice to faculty, and local distribution of service work. Examining their accounts therefore allows us to analyze how service devaluation is reproduced through routine practices of evaluation rather than experienced only as an outcome.

3. Materials and Methods

This study employed a qualitative interview design to examine how P&T committee members enact and reproduce institutional contradictions in service evaluation.

3.1. Study Context

The research was conducted at a mid-sized Catholic research university in the northeastern United States. The university's mission emphasizes service to the community, nation, and world as a core institutional value and explicitly encourages a spirit of service among faculty, staff, and students. Despite this mission's emphasis, faculty evaluation follows a traditional three-pillar framework of teaching, scholarship, and service while maintaining P&T requirements that privilege teaching and scholarship over service. P&T criteria are specified in the faculty handbook, and candidates are evaluated holistically through a portfolio of evidence rather than through a point or scoring system.

The university's P&T processes follow standard academic timelines. Candidates are reviewed successively at the department and college or school levels before their cases reach a university-level committee, whose recommendation precedes the final decision by the president. This university-level committee is chaired by the provost and includes at least one full-time tenured faculty member from each college or school. Members are elected by the tenured and tenure-track faculty of their units and serve staggered three-year terms, with no more than two consecutive terms; academic administrators such as deans and department chairs are ordinarily ineligible. Because the committee includes representatives from each unit, its members routinely evaluate candidates from disciplines outside their own.

Although the Catholic mission and emphasis on service make this institutional context distinctive, it provides a theoretically informative case for examining tensions between institutional values and evaluative practices. If service is devalued even at an institution that explicitly elevates service in its mission and identity, this setting provides a useful case for examining how broader tensions in academic evaluation can persist despite local com-

mitments to service. The single-institution design enables close examination of evaluation practices that may extend beyond this case.

3.2. Participants

We purposively sampled members of the university-level P&T committee because they are the actors who put institutional policy into practice. This committee applies shared institutional criteria to candidates from across the university, making it the point where handbook language, procedural routines, and institutional norms become concrete judgments about faculty cases. Its members are therefore well positioned to reveal how the contradictions surrounding service are enacted and reproduced. We focused on this committee rather than department or school-level committees because it is the final, cross-disciplinary stage of review before the president's decision.

Former committee members were identified through institutional records of committee membership. We invited 19 individuals who had served on the committee within the five years preceding data collection but were not currently serving, and nine agreed to participate in a voluntary interview about their experiences evaluating faculty cases. Interviewing former rather than sitting members allowed participants to speak candidly without concern about ongoing committee dynamics. Participants' service length varied, with some having served multiple terms under different provosts, while others served shorter, more recent terms.

Participants represented diverse disciplinary backgrounds across the university's major colleges and schools. Collectively, they represented six of the university's nine academic units, providing perspectives from a broad range of disciplinary and professional contexts. This count excludes a newly established academic unit that had not yet been represented on P&T committees. This diversity was essential for understanding how service evaluation challenges manifest across different academic contexts and disciplinary cultures. All participants were tenured associate or full professors, consistent with committee eligibility requirements.

3.3. Data Collection

We conducted individual semi-structured interviews between 30 October 2024 and 19 November 2024. A semi-structured interview guide was developed to explore several key areas related to service evaluation in P&T decisions. Interviews lasted 45 to 90 min. To ensure comprehensive data collection and reduce bias, two research team members facilitated each interview, with one exception due to scheduling.

Alongside the interviews, we reviewed relevant institutional documents, including the faculty handbook, to contextualize participants' accounts and to examine the formal language through which service was represented in P&T evaluation. We treated these documents as context for interpreting the interviews rather than as a separate corpus for formal document analysis.

3.4. Data Analysis

We analyzed the interviews using Braun and Clarke's [37] six-phase approach to thematic analysis, treating it as a recursive rather than linear process. We began by familiarizing ourselves with the data. Interviews were transcribed using Whisper (20240930 release) [38], checked for accuracy, deidentified, and read closely before coding. In generating initial codes, two team members independently coded the transcripts, working from deductive codes drawn from our framework of institutional contradiction and policy enactment and from the interview guide, while adding inductive codes as they arose. They reconciled their

coding after each transcript and, as a check for consistency, calculated Cohen's kappa [39] on the final transcript ($\kappa = 0.764$) before consolidating a shared codebook.

We then collated codes into candidate themes across participants and reviewed these against both individual extracts and the full dataset. In defining and naming themes, we refined each in light of the study's conceptual framework, developing them into the four findings reported below. Producing the report involved selecting illustrative extracts and relating the analysis to our research question and the literature.

3.5. Rigor and Ethics

The research team included faculty with P&T experience as both candidates and evaluators. This insider knowledge facilitated rapport-building with participants and enhanced our understanding of complex institutional dynamics described in interviews. At the same time, our own experiences and social positions within academia may have shaped which aspects of service, evaluation, and inequity we noticed most readily. We addressed this through the collaborative structure of the analysis. Because two team members coded independently and reconciled their readings transcript by transcript, divergent interpretations were surfaced and tested against the data rather than settled by a single reader's assumptions. An internal evaluator (an experienced researcher external to the research team but within the institution) independently reviewed the interview protocol, analytic decisions, and findings, providing methodological rigor and an external check on our interpretations.

This study was approved by the university's Institutional Review Board. All participants provided informed consent and were assured that their participation was voluntary and they could withdraw at any time. Given the sensitive nature of discussing institutional evaluation processes, we protected participant confidentiality through secure data storage, transcript deidentification, and the use of de-identified participant codes in place of names in all research outputs.

4. Results

Interviews with nine former university P&T committee members show how the contradiction surrounding service is enacted and reproduced through the ordinary work of evaluation. Committee members did not describe disregarding service or deciding that it did not matter. They described valuing service, yet through the routine practices of reviewing cases, they reproduced its devaluation. The four findings below show how the evaluation of service sustains the contradiction between institutional dependence on service and its limited weight in advancement.

First, service is formally recognized as necessary but left evaluatively underspecified, named as a domain of review without clear criteria for what counts as quality or excellence. Second, evaluation procedures systematically marginalized service through time allocation, discussion sequencing, and limited preparation, making service less consequential than teaching and scholarship. Third, in the absence of clear evaluation criteria, committee members filled the gap with informal standards and cautious discretion, creating subjective and inconsistent evaluation processes dependent on individual judgment. Fourth, protective practices and uneven service demands across units reproduce service devaluation socially, concentrating the work on a smaller set of faculty while signaling that service is necessary for the institution but risky for individual advancement. Together, these practices form a mutually reinforcing set that allows service to remain formally recognized while leaving its underlying devaluation intact.

4.1. Finding 1: Service as Necessary but Evaluatively Underspecified

The first way service devaluation was reproduced was through the mismatch between service's formal inclusion in P&T review and its limited evaluative specification. Committee members described service as necessary to institutional functioning, yet the formal criteria positioned it as less consequential to advancement than teaching and scholarship. Service was therefore not absent from evaluation; rather, it was required without being clearly defined as a domain in which candidates could demonstrate excellence or advance their cases.

This asymmetry was embedded in the university's faculty handbook. For tenure and promotion to associate professor, candidates must demonstrate excellence in at least teaching or scholarship, effectiveness in the other, and only effectiveness in service. For promotion to full professor, requirements become more demanding for teaching and scholarship—excellence is required in both—while service continues to be held only to the standard of effectiveness. One participant explained, "The evaluation is stated in the handbook, says you have to be at least excellent in one of these two, effective in the other, and at least effective in service" (P8). The relative weight of service was therefore clear. What remained unclear was what constituted effective service.

Participants recognized this as a fundamental asymmetry and several noted that service excellence provides no evaluative advantage. One committee member stated, "You don't get any brownie points for being awesome at service. And that's actually in the handbook" (P4). Another commented, "Excellent in service, that's not going to get you through the door" (P6). One participant questioned the logic of even evaluating service excellence: "Should we even be evaluating excellent service?... Should we just evaluate people as effective versus ineffective? Because it doesn't get you anything" (P8). These accounts show that service was not simply overlooked by individual committee members; the formal criteria themselves made service difficult to treat as a career-advancing category. Committee members could recognize exceptional service, but the evaluation system gave them limited grounds for making that recognition consequential.

The limits of this structure became especially visible when service did not align neatly with stronger performance in teaching and scholarship. One committee member described how this structure left them without guidance in cases where weak service conflicted with strong teaching and scholarship:

If that person hit the criteria for excellent in teaching and effective in scholarship, is ineffective in service enough to keep them back? And that led to a very difficult discussion... It wasn't people pounding their fists on the table. It was more like, what do we do? How do we manage this? They were just trying to solve the problem and figure out how to move ahead when the system hadn't anticipated this situation (P6).

The difficulty here was not disagreement over whether service mattered; it was uncertainty about how service should matter when the formal criteria did not specify its evaluative force.

The handbook also provides a limited basis for distinguishing levels of service quality. As one participant noted, "Excellence in service is never an expectation. It's just simply being effective, whatever effective means" (P3). Although the handbook provides examples of service activities, it offers no guidance for distinguishing what makes service effective, excellent, or insufficient, so service is formally named but weakly translated into evaluative standards.

Similar ambiguity emerged around expectations for different forms of service. Participants occasionally distinguished between departmental, institutional, professional, and community forms of service, particularly when discussing expectations associated with different academic ranks. While they generally expected broader forms of service from more

senior faculty, they lacked clear criteria for determining when contributions within those domains were sufficient, significant, or exemplary. This under-specification is what allowed service to remain institutionally necessary while carrying limited weight in advancement. It entered the P&T process as something candidates had to demonstrate, but not as something that could reliably strengthen their case.

4.2. Finding 2: The Procedural Marginalization of Service

If the formal criteria left service underspecified, committee procedures translated its lower formal standing into marginalization during review. Three routines contributed to this work: the order in which domains were discussed, the time allotted to each, and the preparation members received for the task. Each resembled a neutral matter of administration, but together they made service formally present yet less visible, less discussable, and less consequential than teaching and scholarship.

Committee members consistently reported reviewing each case in the order of scholarship, teaching, and then service. This ordering proved both practical and symbolic, as one participant observed, "service was usually the final thing discussed" and "I had a sense that the scholarship and teaching conversations that preceded service were the determining factors" (P3). By the time discussion reached service, the case was often effectively settled. One participant noted that "by the time you got to service... you knew if they were going to pass" (P7), and another described service as constrained by judgments already formed:

We always consider service last, at least when I was on the committee. We always considered it last as far as the candidate was concerned. And then ultimately you make the overall vote, but... it has to be in line with how you voted for the other three (P8).

Positioned last, service could confirm a favorable case or flag a concern about a marginal one, but it rarely altered a recommendation. The sequence itself, a seemingly practical choice, made service consequential only at the margins.

Service is not only discussed last, but it is also discussed least. Participants consistently reported that service received substantially less discussion time than teaching or scholarship. One committee member provided specific estimates as follows:

If we would evaluate a candidate and spend about two hours doing it, right? Generally, we would spend... maybe, 40 to 45 min on teaching, 40 to 45 min on scholarship, and maybe 20 min on service (P8).

Another member estimated that service comprised only "25% of the discussion unless there was disagreement" (P9), while others described service discussions as "definitely shorter in the time period" (P4) and noted spending very little time talking about service even for the full professors. These proportions were consistent enough that participants treated them as normal, and longer service discussions arose only as exceptions; as one noted, "longer conversations typically happened if there was concern that there was not enough service" (P3). The pattern conveyed what one participant heard as follows: "a definite air of service being perhaps the least important of the things" (P5).

Compounding these patterns, committee members were minimally prepared to evaluate service. Most described learning the role informally, on the job, rather than through structured guidance. One participant captured this experience as "we were thrown into the fire. It was like, 'Hey, welcome to the committee' " (P9) and another reported that "no training was provided" (P5), with guidance limited to the mechanics of accessing files rather than criteria for judgment. Because service criteria were already the least specified of the three domains, this absence of preparation fell hardest on service, leaving members to evaluate it with the least institutional support.

None of these procedures announced a decision to devalue service. Each looked like a reasonable way to manage a heavy workload. Yet in combination they routinized

service's secondary status, so that even committee members who valued service in principle enacted its marginalization through the ordinary conduct of review. This devaluation was reproduced less through what committee members believed about service than through how they were procedurally arranged to evaluate it.

4.3. Finding 3: Informal Standards and Cautious Discretion in Service Evaluation

Where the formal criteria were lacking, committee members had to supply standards of their own. Because the handbook set no threshold for how much service was expected or what constituted excellence, members constructed informal expectations case by case. One participant described the gap as follows:

There was clearly an expectation to see more service for cases that were requesting promotion and tenure. But again... it was really never clearly articulated what that threshold is... How much is a reasonable amount for associate professor? What's a reasonable amount for full professor (P3)?

In the absence of an answer, definition defaulted to the candidate. One committee member explained the following:

Service was defined—and this is true for everything—as how you make the case for it. If you tell us that such and such is service, we will either agree with you and count it as service, or we will disagree with you and disregard what you offered (P4).

This placed the burden of definition on candidates rather than on the institution. Committee members recognized service largely by where candidates filed it, so that: "if it wasn't teaching and it wasn't research... it tended to be looked at as service" (P1), and "So, if you're doing something astonishing in service that we hadn't thought of, that doesn't make it not service. That just makes us dumb" (P4).

While this flexibility might seem generous, it systematically shifts definitional authority from the institution to individual evaluators and candidates. The result is that service evaluation depends not on institutional standards but on candidates' rhetorical abilities and evaluators' personal receptiveness.

Without stable standards, committee members were left to rely on candidate framing and their own improvised judgments. One participant observed the following:

If you [the candidate] tell us what the service expectations are, we know what they are. If you don't tell us either a rubric or a target, we'll make something up. The funniest thing I said, I said that I'm not telepathic. Here's the secret. Well, nobody else in the committee was either (P4).

These improvised standards made service evaluation inconsistent and heavily dependent on who was in the room. To manage ambiguity, committee members developed an informal practice of deferring to disciplinary expertise that further individualized evaluation decisions. One participant explained the following: "we will tend to defer to the person from the relevant school about what typical service is or is not in their field" (P4). While this strategy provided context, it also meant that service evaluation depended on the particular expertise and perspectives of whichever committee members happened to represent candidates' relevant disciplines. Beyond disciplinary deference, members drew on their own philosophies and experiences. Disagreements were often "attributed to people's philosophies of service around the table" (P1), and members acknowledged the subjectivity openly: "A lot of this is subjective. So we share our opinions. And there isn't always a, you know, there definitely is not always unanimity in how we view things" (P8). One participant named the absence of a shared standard plainly: "I didn't sense that we had a kind of a rubric in our work. So I don't feel that we all viewed it the same way" (P3).

The discretion committee members exercised is not itself unusual, since evaluating any faculty domain involves judgment within institutional parameters [40]. What is notable

is that, in some accounts, discretion operated cautiously rather than expansively. This cautious discretion took the form of default skepticism in which service counted most clearly when a candidate made an unusually strong case or had a champion in the room. As one participant observed, candidates' success could hinge on whether they "have an advocate on the committee" because otherwise "you [the candidate] really have to make sure things are a slam dunk" (P1). Where teaching and scholarship had clearer criteria to meet, service had to be argued through uncertainty that the criteria did little to resolve.

None of this required committee members to devalue service deliberately. They filled a vacuum the criteria had left, and they did so in good faith. But the standards they improvised were uneven, and the caution with which they applied them meant that service, already lowest in the formal hierarchy and last in the procedure, now also carried a heightened burden of proof. In exercising the discretion the system handed to them, committee members reproduced service's devaluation rather than correcting it.

4.4. Finding 4: Protective Socialization and the Uneven Distribution of Service

The final way service devaluation was reproduced was through the adaptive strategies committee members and departments used to manage its consequences. Participants did not describe devaluing service because they thought it unimportant. Instead, they described efforts to protect individual faculty from the risks created by a system in which service was necessary but weakly rewarded. These responses were often reasonable and well-intentioned. Yet in aggregate, they conveyed the message that service should be treated as career-risky, redistributing service burdens unevenly, and leaving the underlying evaluative structure intact.

This protective logic was clearest in how participants described advising other faculty. Committee members described advising faculty to limit service involvement despite institutional dependence on faculty service. One committee member explained the following: "Telling them that a certain amount of service is expected and required is good, telling them to spend the bulk of their week on service I don't think is good... Because that's not how the evaluation goes" (P8). Committee members frequently worried about faculty doing "too much service," with one noting that excessive service might be "a disservice to that candidate" (P8).

Departments acted on the same reasoning, shielding junior faculty from service responsibilities that the institution requires. As one noted, "The department was shielding you from doing too much service so that you could concentrate on getting tenure" (P7). Service was not dismissed as meaningless; it was treated as necessary but risky if it displaced work more likely to count towards P&T. Protecting candidates within the existing rules, this advice also reproduced the message that service should be minimized during the very stage when evaluation mattered most, socializing faculty to see service as institutionally required but individually risky.

Protecting some faculty from service meant the work still had to be carried out by others. One participant described the following consequence: "you're going to have 20% of the people doing 80% of the work. And you're just hoping that somehow it all evens up in the end" (P1). Shielding junior faculty did not reduce the institution's need for committee work, governance, student support, or accreditation-related labor; it redistributed that labor to faculty who were more available, more willing, more senior, more likely to be asked, or less able to refuse.

Committee members recognized that these burdens fell unevenly. Several observed that the same people performed service repeatedly, with one voicing frustration at "seeing all the same people in these meetings" (P9), and several identified gendered dimensions. One participant noted the following: "women are more likely to do more... and there's

probably a lot of invisible [service] that's not documented" (P1). Another described how a high-value leadership role went repeatedly to one person whose identity differed from the department majority, asking the following: "why is this one person constantly in charge of this when most of the department is this type of person?" They acknowledged that "not everyone is getting a fair shake at some of these service activities" (P9).

Service expectations also varied across academic units. One participant explained that "in the [school name], I think service was valued more, because they really have a heavy service load" (P2), while another noted that accreditation requirements in certain schools created uneven burdens (P8). These disparities were significant enough that "some academic units had to be told to get it in gear that their dossiers are not going to make it" (P9). The result was an uneven institutional landscape in which service was expected everywhere but not performed, documented, or valued in the same way across the campus. Because these expectations were set locally rather than institutionally, the contradiction was not merely passed from mentor to mentee; it was built into how service was organized across the institution.

Committee members had limited capacity to address these patterns, because their work was organized around individual cases. They could evaluate the dossier before them and draw on department letters or school representatives for context, but they lacked the institutional memory to see patterns across departments or over time. As one explained:

It's not like we see every case and we remember every case from before out of a department or a program or an academic unit. And so you can have really nuanced, but nefarious type of things going on and not realize they're happening because of who had access to get some of those service activities and those who don't (P9).

These individual accommodations allowed committee members to navigate individual cases, but could not address the systemic inequities they recognized as problematic.

They were also wary of reforms that might make service evaluation more mechanical. One worried that "if we start nickeling and diming everything, then people start to try to do it mechanically, and then want to get credit for it. And it's actually hurting" (P1). Another explicitly rejected such approaches, "Don't make it a rubric. Good Lord. That is my biggest fear is people will start ranking them and make them some weird rubric" (P9). These concerns reflected a legitimate worry that quantifying service could distort its meaning or reward performative participation. Yet resisting more systematic approaches also left the existing ambiguity in place.

The result was a form of well-intentioned reproduction. By protecting junior faculty, absorbing inequities informally, and resisting reforms that might create new distortions, committee members and departments made an imperfect system workable without changing the conditions that made service risky, uneven, and weakly rewarded. Service devaluation was reproduced not because committee members disregarded service, but because the solutions available to them kept the contradiction manageable rather than resolving it.

Table 1 summarizes the four mechanisms identified in the findings, the evaluator practices associated with each mechanism, and their institutional outcomes.

Table 1. Mechanisms reproducing service devaluation in promotion and tenure evaluation.

Mechanism	Evaluator Practice	Institutional Outcome
Evaluative underspecification	Evaluating service without clear standards	Exceptional service receives limited evaluative recognition
Procedural marginalization	Discussing service last; allocating less time to it	Service becomes secondary in deliberations

Table 1. *Cont.*

Mechanism	Evaluator Practice	Institutional Outcome
Informal standards and cautious discretion	Relying on personal judgment	Inconsistent evaluation across candidates
Protective socialization and uneven distribution	Advising junior faculty to limit service involvement	Reproduction of service devaluation and workload inequities

5. Discussion

This study shows how contradictions in faculty service evaluation persist through the routine work of the committees charged with carrying it out. Across the four findings, service is formally required but evaluatively underspecified, marginalized by the procedures of review, filled in through informal standards and cautious discretion, and redistributed and passed on through protective advice. What links them is a single dynamic: the contradiction is reproduced not through opposition to service but through evaluative practices that make service difficult to recognize as career-advancing evidence. Committee members who valued service still reproduced its devaluation, because the system gave them weak criteria, marginalizing routines, and few tools to do otherwise.

Together these practices form a mutually reinforcing set that sustains contradictory demands while appearing to function effectively. Committee members consistently recognized the problems inherent in service evaluation yet relied on individual strategies that kept the contradictory system workable. In aggregate, these accommodations plausibly reduced the pressure for structural change, leaving core tensions intact while reinforcing the devaluation of service. This helps explain why service devaluation persists even when faculty and evaluators openly acknowledge its inequitable effects.

These findings also suggest that the consequences of prioritizing teaching and scholarship over service extend beyond the formal evaluation of individual faculty. While the hierarchy itself may be intentional, the practices that emerge from it can reinforce uneven service burdens and discourage faculty from investing in work that is essential to institutional functioning. This creates a broader organizational problem in which the same reward structures designed to prioritize teaching and research may undermine the collective service contributions needed to sustain them. Thus, the consequences of service devaluation are not limited to individual career advancement but also concern the institution's ability to maintain the activities on which it depends.

5.1. Theoretical Implications

Our findings extend organizational paradox scholarship by specifying how contradictory institutional arrangements persist over time. Specifically, we show that paradoxes can be reproduced through ordinary evaluative practices that appear neutral, practical, or protective rather than through overt resistance to institutional values. While Smith and Lewis [13] identify performing paradoxes as tensions between competing performance demands that require ongoing management, our study shows how such management can occur through accumulated individual adaptations that sustain rather than resolve the paradox. P&T committee members did not simply navigate the tension between institutional dependence on service and service devaluation; they were actively, if unknowingly, reproducing it through evaluation practices.

This reproduction was constituted through communicative and procedural routines. Consistent with Putnam et al. [12], committee deliberations enacted the paradox through sequences of discussion, systematic disparities in time allocation, and the substitution of

individual judgment for institutional standards. These seemingly neutral administrative procedures appeared practical and orderly, yet they shaped what became visible, discussable, and consequential in evaluation, thereby sustaining evaluative priorities that conflicted with the institution's stated commitments to service. Recognizing the pattern did little to loosen it, which shows how routine practices become institutionalized and resistant to individual correction.

The dynamics also extend Lawrence and Suddaby's [33] account of institutional work, the ongoing efforts through which actors create, maintain, and disrupt institutional arrangements. Committee members were clearly engaged in institutional maintenance work, sustaining contradictory logics rather than resolving them and making the tension between institutional need and career incentive manageable case by case. Their work illustrates the limits of individual agency inside contradictory structures. Aware of the misalignment and concerned about fairness, they could not, through their own choices, alter the arrangements that produced it.

These limits reposition the question of agency by shifting attention from service as labor performed by faculty to service as a category produced through evaluation. The faculty agency literature has largely examined the agency of those who perform service, showing how faculty, and especially women and faculty of color, use service to resist and redefine the structures that constrain them [24,26]. Our evaluators exercised agency from a different position. They recognized the structural conditions shaping how faculty respond to service, yet their own agency was bound by the evaluation system they administered, so their adaptive responses reproduced the very conditions they named as problematic. Agency, in this account, is not only a source of resistance but also a vehicle of reproduction. The same capacity that lets faculty push against devaluation lets evaluators, differently situated, sustain it.

The finding on cautious discretion carries a related implication for how discretion is understood in faculty evaluation. O'Meara [40] treats faculty discretion as a lever that can advance full participation when it is checked and structured, but that can reproduce inequity when it is not. Our study documents the second case for service. Handed wide discretion but no standards, committee members exercised it cautiously, raising rather than lowering the bar for service to count as persuasive evidence, and using their latitude in ways that reproduced devaluation rather than countered it. Expanding evaluator discretion is therefore not, by itself, a remedy for service's marginalization. Discretion advances equity only when the criteria and procedures that structure it are built to support service rather than to sideline it.

5.2. Empirical Contributions

Empirically, this study shifts attention from the distribution and experience of service to the evaluation practices through which service becomes career-relevant or career-limiting. Prior research has shown that service is unevenly distributed and often devalued, especially for women, faculty of color, junior faculty, and those in service-intensive roles. By examining committee members who apply P&T criteria, our findings show how the limited weight of service is enacted, and in some respects extended, inside the review process. The four findings specify how service is formally included but weakly specified, procedurally marginalized through discussion order and time allocation, evaluated through informal and inconsistent standards, and redistributed through protective socialization. Together they offer an account of service devaluation, locating it not in individual bias [41] or policy design flaws [2] alone, but in the linked practices of evaluation, which helps explain why interventions aimed at individual knowledge or attitudes have had limited effect [16,17].

Centering on university-level P&T committee members provides a distinctive vantage point. These evaluators sit where institutional policy, disciplinary variation, procedural routine, and career consequence converge, and their accounts bring to the surface tensions less visible in the studies of those who perform service: the difficulty of defining service excellence, the procedural ways service becomes secondary, the reliance on informal judgment when criteria are weak, and the limits of a case-by-case review for addressing systemic inequity. Seen from this position, service devaluation is not only experienced by candidates but enacted by the institutional actors charged with judging their work.

Finally, the study shows how an institution can reproduce the devaluation of service even while rhetorically elevating it. Service is central to the university's stated mission and identity, yet that emphasis did not translate into clear evaluative criteria, sustained procedural attention, or consistent recognition of service as career-advancing evidence. This suggests that institutional commitments to service require more than symbolic affirmation; they must be embedded in the criteria and routines through which faculty work is evaluated.

5.3. Policy and Practice Implications

The findings suggest that addressing service devaluation requires more than encouraging committees to value service more. Committee members in this study often value service, but they worked within criteria and procedures that made service difficult to recognize as career-advancing evidence. Institutions seeking to address this contradiction should therefore focus on the structures through which service is defined, documented, discussed, distributed, and rewarded.

A central implication is the need for clearer pathways for recognizing service excellence. As long as service remains a threshold requirement while teaching and scholarship define advancement, evaluators will treat service as secondary even when institutions depend on it. This might involve allowing service excellence to substitute for excellence in other areas, creating service-focused promotion tracks, or developing clearer distinctions among routine participation, meaningful contribution, leadership, impact, and institutionally significant service.

Clear standards, however, raise a concern that participants voiced directly. Some resisted the idea of a service rubric, worrying that itemizing service would lead faculty to pursue it mechanically for credit. This concern reflects Campbell's Law [28]: the more an indicator is used to make decisions, the more it can distort the activity it is meant to measure. Counting committee memberships or hours could reward visible participation over meaningful contribution and push the care and relationship-building work that makes service valuable further out of view. Yet the same participants also described the costs of having unclear standards. Without shared criteria, committee members supplied their own, making evaluations dependent on individual judgment and potentially influenced by whether a candidate had an advocate on the committee.

The choice, then, is not between a scoring rubric and unclear criteria. O'Meara [40] cautions that replacing judgment with supposedly objective tools can hide bias rather than remove it, since faculty can build their own assumptions into the criteria they create. Bias is more likely when criteria are vague and decisions are rushed [40], and our findings suggest that the combination of underspecified service criteria and limited discussion time may create similar conditions. The alternative is to structure discretion rather than eliminate it [40].

Institutions could establish shared standards that focus on the quality, significance, and impact of service contributions rather than the quantity of activities performed. These standards should allow different forms of evidence to demonstrate meaningful service. For example, chairing a faculty search committee, organizing a community outreach initiative,

and performing essential but less visible institutional service may demonstrate value in different ways. Requiring candidates to meet uniform numerical thresholds or demonstrate the same types of outcomes could disadvantage contributions that are difficult to quantify. Evaluators should instead assess these contributions in context, using shared criteria while explaining the evidence supporting their judgments. Such an approach would not eliminate the risk of strategic behavior, since even qualitative criteria can become targets. Institutions should therefore periodically review whether their standards are encouraging meaningful contributions or simply rewarding activities that are easier to document. In this way, clearer standards can support more consistent evaluation without reducing service to a set of measurable outputs.

Evaluation procedures also need attention because service can be marginalized even when it is formally included in review. The findings show that discussion order, time allocation, and limited preparation can make service secondary even when committee members value it in principle. Committees could adopt procedures that require explicit discussion of service quality and impact, provide guidance on how service evidence should be weighed, and train members in evaluative criteria rather than only in the mechanics of accessing files. Beyond training, committees may also need calibration practices in which members discuss sample service records, compare interpretations of service quality and impact, and surface disciplinary assumptions before evaluating actual cases. Allowing candidates to designate which area should be discussed first acknowledges their expertise in framing their own contributions and prevents the automatic devaluation that occurs when service is always considered last.

Service must also be made more visible and consistently documented. While candidates submit service narratives, the absence of structured guidance allows invisible work to remain hidden and creates inconsistent documentation quality. Institutions should provide clear narrative guidance that prompts candidates to address specific dimensions of their service: nature and scope of activities, time commitments, leadership roles, institutional impact, contributions to equity and community building, and forms of care or maintenance work that often remain invisible. These structured prompts would surface work that typically goes undocumented without reducing service to checklists or numeric metrics. Committee training should reinforce how to recognize substantive service within these narratives, emphasizing that impact and context matter more than quantity.

Reform must also reach the departmental and disciplinary settings where service expectations are communicated and distributed. Much of what reproduces service devaluation happens below the university level, in how departments advise faculty and distribute service. Protective mentoring may help junior faculty survive the existing system, but it also teaches them to treat service as career-risky and can concentrate essential work among senior faculty, women faculty, faculty of color, or others more likely to be asked. Departments should therefore pair mentoring about sustainable service with transparent workload practices that distribute the necessary service more equitably. Because service expectations also vary widely across units, institutions should develop shared but field-sensitive expectations so that faculty in service-intensive units are not disadvantaged relative to those in units where little service is expected. Institutions may also need mechanisms for tracking service participation and access to high-value service roles across units and over time, since a case-by-case review makes patterned inequities difficult for committees to see.

Reform of service evaluation must ultimately integrate broader institutional reward systems. Evaluation practices that discourage service engagement are incompatible with institutions' reliance on service for governance, accreditation, curriculum, student support, and community engagement, particularly when essential work is concentrated among a small subset of faculty. Meaningful reform requires coordinated change across departments,

schools, and university-wide policies so that service is not only rhetorically valued but structurally recognizable, equitably distributed, and consequential in faculty advancement.

5.4. Broader Significance for Higher Education

The contradictions revealed in this study extend beyond P&T processes to broader questions about sustainability and governance in higher education. Faculty service supports everything from curriculum development and student mentoring to accreditation and institutional governance, yet evaluation systems often discourage excellence in these essential activities. When institutions rely on faculty service but provide no career incentive for it, they create conditions that push toward either service failure as faculty rationally minimize service engagement, or faculty exhaustion as conscientious faculty continue providing essential service without adequate recognition. Both outcomes threaten institutional capacity and faculty well-being.

These implications are particularly relevant given current debates about academic workload equity and the disproportionate service burdens carried by women faculty and faculty of color. The contradictions identified here help explain why inequities persist despite widespread awareness and good intentions. When evaluation systems are fundamentally contradictory, individual efforts to address inequitable service distribution cannot succeed because the underlying structure continues to devalue the very work that creates equity problems. Given documented gender and racial disparities in service participation, the current evaluation structure may further contribute to inequities in faculty advancement [5,42]. The implication is not simply that institutions should value service more, but that they must make service evaluable, so that its quality, impact, and institutional importance can be recognized rather than assumed.

Other institutions facing contradictions between stated values and reward structures may show comparable patterns of well-intentioned reproduction. The account offered here, of a contradiction sustained through the ordinary practices of those who administer it, may help explain how institutional paradoxes persist through individual adaptation more broadly, and may inform organizational change efforts across higher education.

Extending beyond the immediate data, the deepest of these contradictions may be professional as much as institutional. Even where a single university resolves to value service, the standards against which its faculty are judged are set partly beyond it, within disciplinary evaluation cultures that shape what counts as academic excellence [35]. Promotion and tenure evaluations lean heavily on external reviewers, whose letters are written within disciplinary cultures that equate scholarly standing with research productivity. Within that prestige economy, a common objection to rewarding service is that doing so would dilute the research standard, so that elevating service is heard as lowering the bar. This locates service devaluation in the reward culture of the academic profession, not only in any institution's handbook, and it suggests that local reform can only go so far. Grappling with the contradiction at its root would require disciplines, scholarly societies, and the external-review system to reconsider what counts as valuable academic work, so that recognizing service is understood not as a threat to scholarly standing but as part of a fuller account of what faculty contribute.

5.5. Limitations and Future Research

This study has several limitations that suggest directions for future research. First, the research was conducted at a single university with a service-focused mission. This context provides a theoretically informative case for examining contradictions between stated values and evaluation practices, but the findings may not transfer directly to institutions with different missions, organizational structures, or faculty evaluation systems. Future research

should examine service evaluation at other institutional types to explore how differing missions shape the reproduction of service devaluation.

Our sample consisted of nine former P&T committee members, all tenured associate or full professors. These participants were well-positioned to speak to institutional contradictions and cross-disciplinary decision-making, but they represent only one vantage point in a broader evaluation ecosystem. Future research should incorporate perspectives from candidates, department chairs, deans, and administrators to better understand how contradictions are experienced and reproduced across roles and levels of decision-making.

Additionally, we did not collect demographic data from participants. With nine former committee members from a single identifiable university-level committee, even limited demographic breakdowns could have increased the risk of participant identification and conflicted with our commitments to confidentiality. This limits our ability to examine how evaluators' own social identities may have shaped their perspectives on service, equity, and evaluation.

The study draws on interview data rather than direct observation of committee deliberations. Interviews enable retrospective reflection on dilemmas that may be difficult to articulate in real time, but they do not capture observable interactions. Future research could triangulate interview-based findings with ethnographic observation of committee meetings, content analysis of evaluative narratives, or systematic coding of P&T dossiers to examine how service is discussed, weighted, and documented in practice.

6. Conclusions

This study examined how universities depend on faculty service while giving it limited weight in promotion and tenure. Drawing on interviews with former university-level P&T committee members, we show that this contradiction is reproduced through the ordinary work of evaluation. Service is formally required but evaluatively underspecified, procedurally marginalized, interpreted through informal standards and cautious discretion, and redistributed through protective advice and uneven service demands across units.

Across these findings, service devaluation is not the product of indifferent or ill-intentioned evaluators. It is sustained through the routine practices by which committed evaluators judge faculty work within systems, which give them limited tools for recognizing service as evidence of career advancement. This study extends organizational paradox and institutional work scholarship by showing how contradictions persist through evaluation practices rather than being resolved through formal policy, structural reform, or individual goodwill. Unless institutions make service more evaluable and consequential in advancement, they will continue to depend on work that their own reward systems render secondary.

Author Contributions: Conceptualization, P.O., L.E.K., R.M.N. and M.H.; methodology, P.O. and L.E.K.; validation, P.O. and L.E.K.; formal analysis, P.O. and L.E.K.; investigation, P.O., L.E.K., R.M.N. and M.H.; data curation, P.O. and M.H.; writing—original draft preparation, P.O. and L.E.K.; writing—review and editing, P.O., L.E.K., R.M.N. and M.H.; project administration, R.M.N.; funding acquisition, P.O., L.E.K., R.M.N. and M.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research is based upon work partially supported by the National Science Foundation ADVANCE program under Award No. 2426556. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and was reviewed by the Institutional Review Board of Duquesne University.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Participants provided written consent before their interview, including separate consent to audio recording.

Data Availability Statement: The data presented in this study are not publicly available due to privacy and ethical restrictions. The interview transcripts contain information that could identify participants, who were drawn from a small population of faculty serving on a single institution's promotion and tenure committee, and participants were assured that their data would not be shared with other researchers. Requests regarding the data should be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Boyer, E.L. *Scholarship Reconsidered: Priorities of the Professoriate*; Carnegie Foundation for the Advancement of Teaching: Princeton, NJ, USA, 1990.
2. Fairweather, J.S. The mythologies of faculty productivity: Implications for institutional policy and decision making. *J. High. Educ.* **2002**, *73*, 26–48. [CrossRef]
3. Macfarlane, B. *The Academic Citizen: The Virtue of Service in University Life*; Routledge: London, UK, 2007.
4. Finkelstein, M.J. *The American Academic Profession: A Synthesis of Social Scientific Inquiry Since World War II*; Ohio State University Press: Columbus, OH, USA, 1984.
5. Misra, J.; Lundquist, J.H.; Holmes, E.; Agiomavritis, S. The ivory ceiling of service work. *Academe* **2011**, *97*, 22–26.
6. Jaeger, A.J.; Thornton, C.H. Neither honor nor compensation. *Educ. Policy* **2006**, *20*, 345–366. [CrossRef]
7. Guarino, C.M.; Borden, V.M.H. Faculty service loads and gender: Are women taking care of the academic family? *Res. High. Educ.* **2017**, *58*, 672–694. [CrossRef]
8. Jimenez, M.F.; Laverty, T.M.; Bombaci, S.P.; Wilkins, K.; Bennett, D.E.; Pejchar, L. Underrepresented faculty play a disproportionate role in advancing diversity and inclusion. *Nat. Ecol. Evol.* **2019**, *3*, 1030–1033. [CrossRef] [PubMed]
9. O'Meara, K.; Kuvaeva, A.; Nyunt, G.; Waugaman, C.; Jackson, R. Asked more often: Gender differences in faculty workload in research universities and the work interactions that shape them. *Am. Educ. Res. J.* **2017**, *54*, 1154–1186. [CrossRef]
10. Meyer, J.W.; Rowan, B. Institutionalized organizations: Formal structure as myth and ceremony. *Am. J. Sociol.* **1977**, *83*, 340–363. [CrossRef] [PubMed]
11. Thornton, P.H.; Ocasio, W. Institutional logics. In *The SAGE Handbook of Organizational Institutionalism*; Greenwood, R., Oliver, C., Sahlin, K., Suddaby, R., Eds.; SAGE: Los Angeles, CA, USA, 2008; pp. 99–129. [CrossRef]
12. Putnam, L.L.; Fairhurst, G.T.; Banghart, S. Contradictions, dialectics, and paradoxes in organizations: A constitutive approach. *Acad. Manag. Ann.* **2016**, *10*, 65–171. [CrossRef]
13. Smith, W.K.; Lewis, M.W. Toward a theory of paradox: A dynamic equilibrium model of organizing. *Acad. Manag. Rev.* **2011**, *36*, 381–403. [CrossRef]
14. Driscoll, A.; Lynton, E.A. *Making Outreach Visible: A Guide to Documenting Professional Service and Outreach*; Routledge: London, UK, 1999.
15. O'Meara, K. Uncovering the values in faculty evaluation of service as scholarship. *Rev. High. Educ.* **2002**, *26*, 57–80. [CrossRef]
16. O'Meara, K.; Templeton, L.; Culpepper, D.; White-Lewis, D.K. *Equity-Minded Reform of Faculty Evaluation: A Call to Action*; American Council on Education: Washington, DC, USA, 2022.
17. Miller Neilan, R.; Koelsch, L.E.; Heo, M.; Ozturk, P.; Kalarchian, M.A.; Clossen, A.S. Hidden faculty service in U.S. higher education: A literature review. *Educ. Sci.* **2026**, *16*, 352. [CrossRef]
18. Docka-Filipek, D.; Draper, C.; Snow, J.; Stone, L.B. 'Professor moms' & 'hidden service' in pandemic times: Students report women faculty more supportive & accommodating amid U.S. COVID crisis onset. *Innov. High. Educ.* **2023**, *48*, 787–811. [CrossRef] [PubMed]
19. Misra, J.; Kuvaeva, A.; O'Meara, K.; Culpepper, D.K.; Jaeger, A. Gendered and racialized perceptions of faculty workloads. *Gend. Soc.* **2021**, *35*, 358–394. [CrossRef]
20. Heijstra, T.M.; Steinthorsdóttir, F.S.; Einarsdóttir, T. Academic career making and the double-edged role of academic housework. *Gend. Educ.* **2017**, *29*, 764–780. [CrossRef]
21. Gordon, H.R.; Willink, K.; Hunter, K. Invisible labor and the associate professor: Identity and workload inequity. *J. Divers. High. Educ.* **2024**, *17*, 285–296. [CrossRef]
22. Reybold, L.E. Surrendering the dream: Early career conflict and faculty dissatisfaction thresholds. *J. Career Dev.* **2005**, *32*, 107–121. [CrossRef]

23. Johnsrud, L.K.; Rosser, V.J. Faculty members' morale and their intention to leave: A multilevel explanation. *J. High. Educ.* **2002**, *73*, 518–542. [CrossRef]
24. Baez, B. Race-related service and faculty of color: Conceptualizing critical agency in academe. *High. Educ.* **2000**, *39*, 363–391. [CrossRef]
25. Gonzales, L.D. Framing faculty agency inside striving universities. *J. High. Educ.* **2014**, *85*, 193–218. [CrossRef]
26. Gonzales, L.D. Subverting and minding boundaries: The intellectual work of women. *J. High. Educ.* **2018**, *89*, 677–701. [CrossRef]
27. Pyke, K. Service and gender inequity among faculty. *PS Polit. Sci. Polit.* **2011**, *44*, 85–87. [CrossRef]
28. Campbell, D.T. *Assessing the Impact of Planned Social Change (Occasional Paper Series No. 8)*; Public Affairs Center, Dartmouth College: Hanover, NH, USA, 1976.
29. Koretz, D. *Measuring Up: What Educational Testing Really Tells Us*; Harvard University Press: Cambridge, MA, USA, 2008.
30. Dranove, D.; Kessler, D.; McClellan, M.; Satterthwaite, M. Is more information better? The effects of "report cards" on health care providers. *J. Polit. Econ.* **2003**, *111*, 555–588. [CrossRef]
31. Bevan, G.; Hood, C. What's measured is what matters: Targets and gaming in the English public health care system. *Public Adm.* **2006**, *84*, 517–538. [CrossRef]
32. Pache, A.C.; Santos, F. Inside the hybrid organization: Selective coupling as a response to competing institutional logics. *Acad. Manag. J.* **2013**, *56*, 972–1001. [CrossRef]
33. Lawrence, T.B.; Suddaby, R. Institutions and institutional work. In *The SAGE Handbook of Organization Studies*, 2nd ed.; Clegg, S.R., Hardy, C., Lawrence, T.B., Nord, W.R., Eds.; SAGE: Los Angeles, CA, USA, 2006; pp. 215–254. [CrossRef]
34. Braun, A.; Maguire, M.; Ball, S.J. Policy enactments in the UK secondary school: Examining policy, practice and school positioning. *J. Educ. Policy* **2010**, *25*, 547–560. [CrossRef]
35. Lamont, M. *How Professors Think: Inside the Curious World of Academic Judgment*; Harvard University Press: Cambridge, MA, USA, 2009.
36. Mamiseishvili, K.; Miller, M.T.; Lee, D. Beyond teaching and research: Faculty perceptions of service roles at research universities. *Innov. High. Educ.* **2016**, *41*, 273–285. [CrossRef]
37. Braun, V.; Clarke, V. Using thematic analysis in psychology. *Qual. Res. Psychol.* **2006**, *3*, 77–101. [CrossRef]
38. Radford, A.; Kim, J.W.; Xu, T.; Brockman, G.; McLeavey, C.; Sutskever, I. Robust speech recognition via large-scale weak supervision. In Proceedings of the 40th International Conference on Machine Learning, Honolulu, HI, USA, 23–29 July 2023; pp. 28492–28518.
39. McHugh, M.L. Interrater reliability: The kappa statistic. *Biochem. Med.* **2012**, *22*, 276–282. [CrossRef]
40. O'Meara, K. Leveraging, checking, and structuring faculty discretion to advance full participation. *Rev. High. Educ.* **2021**, *44*, 555–585. [CrossRef]
41. Hanasono, L.K.; Broido, E.M.; Yacobucci, M.M.; Root, K.V.; Peña, S.; O'Neil, D.A. Secret service: Revealing gender biases in the visibility and value of faculty service. *J. Divers. High. Educ.* **2019**, *12*, 85–98. [CrossRef]
42. Domingo, C.R.; Gerber, N.C.; Harris, D.; Mamo, L.; Pasion, S.G.; Rebanal, R.D.; Rosser, S.V. More service or more advancement: Institutional barriers to academic success for women and women of color faculty at a large public comprehensive minority-serving state university. *J. Divers. High. Educ.* **2022**, *15*, 365–379. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.