Eliciting Domain Expertise in the Absence of Formal Authority: The Case of AI Developers and Domain Experts in a Large Firm

Abstract

Professionals need to interact with domain experts within organizations to elicit their expertise for executing strategic projects and accomplishing work. However, these professionals often do not have formal authority over domain experts, leading to several challenges in eliciting expertise. *When and how are professionals able (or unable) to elicit expertise from domain experts over whom they have no formal authority?* We examine this research question by drawing on four years of qualitative field work conducted at a large multinational fashion company, Weave (a pseudonym). We compare two AI development projects—involving one successful and one unsuccessful attempt to elicit domain expertise—within Weave that required interactions between AI developers and domain experts (e.g., supply chain allocators, store managers, retail finance managers). We unpack the interplay between task and organizational structures in enabling (or constraining) the effectiveness of AI developers in eliciting domain expertise. In particular, we show that in situations that are characterized by jurisdictional clarity (versus ambiguity), task centrality (versus peripherality), and task enactment homogeneity (versus heterogeneity), AI developers were more effective in accessing domain experts and eliciting their expertise. Building on these findings, we develop a model outlining how the interplay between task and organizational structure shapes both the *legibility* of domain experts as well as the *concentrated* nature of domain expertise, and its consequences for the effective (or ineffective) elicitation of domain expertise.

Luca Vendraminelli
Postdoctoral Fellow, HAI & DEL
Stanford University
lvendra@stanford.edu

Devesh Narayanan
Ph.D. Student
Management Science & Engineering
Stanford University
deveshn@stanford.edu

Arvind Karunakaran
Assistant Professor
Management Science & Engineering
Faculty Affiliate, HAI
Stanford University
arvindka@stanford.edu

Equal Authorship

Professionals within firms need to interact with domain experts—actors with substantive knowledge, experience, or access to unique information in a specific domain—to elicit their expertise for executing strategic projects and accomplishing organizational goals. For instance, to implement sustainability initiatives aimed at aligning a firm's procurement processes with environmental targets, sustainability officers need to elicit domain expertise about current procurement practices from procurement managers and logistics/supply chain experts within the firm (Augustine, 2021). To address issues of diversity, equity, and inclusion, DEI professionals need to elicit domain expertise about current hiring practices from human resources professionals and hiring managers (Weeks et al., 2023). To implement social policies in hospitals to expand medical coverage to traditionally disadvantaged groups, legal professionals need to elicit domain expertise about medical diagnostic, eligibility, and treatment protocols from medical professionals and hospital administrators (Kellogg, 2014). In all such cases, professionals must first identify and gain access to those in their organization who possess the relevant expertise they are seeking, and, in turn, interact with them and build the relational bases to elicit their expertise (Helmstädt, Koljonen, & Elmholdt, 2024)

However, identifying and gaining access to domain experts—let alone eliciting their expertise—is ridden with challenges. In many such cases, professionals have no *formal authority*—here understood as a type of legitimate power that is formalized and vested in official positions and roles through bureaucratic arrangements (Blau, 1968; Aghion & Tirole, 1997)—over those from whom they seek to elicit domain expertise. In large, multidivisional organizations, domain experts are typically embedded in complex hierarchical arrangements and insular departmental structures—such that they may not be formally subordinate to professionals seeking to elicit their expertise, and therefore need not feel obligated to give their time, let alone cooperate or collaborate with those professionals. Indeed, several news media reports allude to the challenges faced by professionals in eliciting expertise from domain experts (Levine, 2023; Palumbo & Edelman, 2023; Somers, 2024). In such cases, professionals who attempt to elicit expertise from domain experts cannot rely on direct commands, sanctions, or coercion.

Prior research, while not directly examining this question, has more generally explored the tactics professionals can use to "get things done" from other actors within their organization (e.g., (Dutton & Ashford, 1993; Huising, 2015; DiBenigno & Kellogg, 2014; Karunakaran, 2022). However, this line of research cannot adequately account for cases where professionals, in the *absence* of formal authority, need to elicit expertise from other actors who are part of a different authority order within their organization, such as domain experts from a different department/function, and/or belonging to a distinct professional/occupational group. Given that such domain experts have expert authority (Wrong, 1979) based on their specialized knowledge and experience, and are not subordinate to the professionals who are attempting to seek and elicit their expertise, tactics that involve direct commands and sanctions are not possible to enact in these situations; even if enacted, such tactics are likely to be ineffective and even backfire (Basbug et al., 2023). Likewise, tactics that involve escalation (e.g., escalating the non-cooperative behavior of domain experts to their managers) are also ineffective in eliciting cooperation, in part because domain experts might consider such tactics as unfair actions from someone who is "punching above their weight" or as "ratting out" to the higher-ups (Karunakaran, 2022). Moreover, beyond these issues related to cooperation, there are also other challenges faced by professionals seeking to interact with and elicit expertise from domain experts. For one, these professionals need to navigate further complexities arising from domain experts being unable to articulate their tacit expertise (Polanyi, 1966), or being reluctant to have their expertise codified for fear of being replaced.

Viewed together, we need to better understand *when and how professionals are able (or unable) to elicit expertise from domain experts over whom they have no formal authority*. Through a qualitative field study drawing on four years of fieldwork at a large multinational fashion company (Weave, a pseudonym), we examine this research question. Specifically, we leverage the case of AI developers and domain experts at this company as a "strategic research site" (Merton, 1987) to examine when and how AI developers were able to effectively elicit expertise from domain experts (e.g., supply chain allocators, retail store managers, finance experts) over whom the AI developers did not have formal authority. Similar to most internal AI developers in large organizations, the AI developers at Weave were part of a separate group tasked with building tools for optimizing various internal organizational processes, while being functionally separated from the domain experts who had experience and knowledge about the very processes the AI developers were trying to optimize. This placed AI developers in a difficult bind: to build useful AI tools, they needed to access and interact with domain expertise to better understand how these work processes were carried out on the ground. However, the AI developers at Weave did not have formal authority to compel the relevant domain experts to meet with them and share their expertise. As we followed this team of AI developers longitudinally for over four years in their attempts to elicit domain expertise from their peers, we observed considerable variation in their effectiveness.

In this paper, we compare two AI development projects that involved the same set of AI developers but produced starkly different outcomes with respect to eliciting domain expertise (and the success of the respective AI tools developed, more broadly) to examine when and how professionals are able (or unable) to elicit expertise from domain experts over whom they have no formal authority. Our findings suggest an interplay between task structures—here understood as the composition and assembly of tasks within a given job role (Cohen, 2013; Feldberg, 2022; Wilmers, 2020)—and organizational structures—here understood as the structuring of activities within an organization into different functions and jurisdictions, and the lines of authority within and across these functions and jurisdictions (Pugh et al., 1968; Scott, 1975)—in enabling (or constraining) the effectiveness of AI developers in eliciting domain expertise. In particular, we show that in situations that are characterized by *jurisdictional clarity* (versus ambiguity), *task centrality* (versus peripherality), and *task enactment homogeneity* (versus heterogeneity), AI developers were more (versus less) effective in accessing domain experts and eliciting their expertise.

Building on these findings, we develop a model unpacking how and when professionals may be effective or ineffective in eliciting domain expertise in the absence of formal authority over domain experts. We highlight how the interplay between task and organizational structure shapes both the *legibility* (versus illegibility) of domain experts as well as the *concentrated* (versus dispersed) nature of domain expertise, and its consequences for the effective or ineffective elicitation of domain expertise. In so doing, we contribute to the literature on cross-occupational collaboration for eliciting domain expertise, and to the field of work and occupations, more generally. We also contribute to advancing our understanding of technology and organizations, focusing on why several in-house AI development projects within organizations fail, highlighting the challenges of cross-occupational collaboration between AI developers and domain experts across authority hierarchies, and how issues emerging "upstream" during the development of AI technologies may shape their "downstream" adoption and use in organizational contexts.

Theoretical Background

Cross-Occupational Collaboration to Elicit Domain Expertise in the Absence of Formal Authority

Prior research on cross-occupational collaboration has examined how professionals can collaborate with members of another professional/occupational group who occupy positions of relatively greater status or authority (Barley, 1986; DiBenigno, 2018; DiBenigno & Kellogg, 2014; Huising, 2015). This research may be grouped into two broad streams.

One stream focuses on *demographic* aspects of cross-occupational collaboration across social hierarchies (DiBenigno & Kellogg, 2014; Koppman, Bechky, and Cohen, 2022). For instance, DiBenigno & Kellogg (2014) show that "cross-cutting demographics"—i.e., when demographic group membership is uncorrelated with occupational membership—allows low-status occupational members to draw on shared demographic characteristics with their high-status counterparts, which in turn enables them to build rapport with and elicit cooperation from these actors. Similarly, Koppman et al. (2022), in their qualitative study of "creatives" (e.g., copywriters, designers) and "suits" (e.g., account executives, strategists) in advertising agencies, discuss how women creatives and account executives in horizontally segregated occupations enact essentialist gender roles and stereotypes, a practice that the authors refer to as "gender ordering," to overcome conflicts and facilitate collaboration with their male peers.

A second stream focuses on *relational* aspects of cross-occupational collaboration across social hierarchies (Bourgoin et al., 2020; Huising, 2014, 2015, Karunakaran, 2022). For instance, Huising (2015) discusses how lower-status professionals who perform "scut work"—i.e., "physically, socially, or morally difficult or dirty work" (p. 267)—can interact closely with, acquire knowledge about, and ultimately build the relational bases to elicit cooperation from higher-status professionals. Karunakaran (2022) discusses how lower-status professionals who adopt a "peer publicizing" relational style—i.e., by calling out the non-compliant behavior of higher-status actors to their immediate peers—can elicit compliance and cooperation from them. Bourgoin et al. (2020) discuss how professionals, such as management consultants, “perform” authority by tactically leveraging or downplaying relations with collaborators, to elicit their cooperation.

While these streams of research have considerably advanced our understanding of the dynamics of cross-occupational collaboration across social hierarchies, the challenges faced by professionals without formal authority in initiating collaborations and eliciting domain expertise from their lateral counterparts are distinct and relatively underexplored. Consider an organization’s sustainability officers who are tasked with aligning a firm’s procurement process with sustainability objectives. They need to collaborate with and elicit domain expertise about current procurement practices from procurement managers and logistics/supply chain analysts in the company. As is typically the case in large, multidivisional organizations, sustainability officers and procurement managers are likely to be lateral peers in different divisions/functions but at roughly similar hierarchical levels within their organization, with distinct (and often, non-overlapping) lines of authority. Within such bureaucratic arrangements, while sustainability officers need not necessarily exhibit deference as if they were dealing with their supervisors, neither can they issue commands and sanctions to procurement managers as if they were dealing with their subordinates. Even with a broader organizational mandate to align the firm’s procurement process with sustainability objectives, sustainability officers may find it challenging even to obtain time commitments from procurement managers for eliciting their domain expertise regarding the firm’s procurement processes, let alone collaborate with them to redesign those processes (Karunakaran & Etzion, 2024; see also Soderstrom & Weber, 2020; Sandhu & Kulik, 2019). Existing discussions of cross-occupational collaboration seem insufficient for understanding how such collaborations to elicit domain expertise might be facilitated in the absence of formal authority.

Moreover, prior analyses of cross-occupational collaboration across social hierarchies have tended to stop at a point where members of a lower-status occupation achieve cooperation from a higher-status occupation (DiBenigno, 2018; Huising, 2015; Karunakaran, 2022). However, in many cases, the goals of such efforts extend far beyond merely achieving cooperation, and the demands of such overarching goals can pose additional challenges. This is especially the case when one such goal is the *elicitation of domain expertise*. In large part, as argued above, this is because while occupational groups seeking to elicit domain expertise are not necessarily subordinate to the domain experts who possess this expertise, they often do not have formal authority over these experts—and as such, may find it difficult to persuade these experts to meet with them and provide them with domain expertise. Furthermore, in cases where domain experts perceive that their alters (for instance, technology developers) are trying to codify and formalize their expertise, they may be motivationally invested in withholding or obfuscating their expertise for fear of losing power or being replaced in their jobs (Forsythe, 1993; Nelson & Irwin, 2014; Koljonen & Chan, 2024). Finally, the domain experts' knowledge can often be *tacit* and *embodied,* and, thus difficult to articulate (Polanyi, 1966). In such cases characterized by 'Polanyi's Paradox' (cf. Autor, 2014) —i.e., when domain experts are unable to articulate tacit expertise to their alters—mere cooperation from these experts is unlikely to be sufficient for eliciting their domain expertise.

Viewed together, the elicitation of domain expertise in the absence of formal authority to do so represents a particularly interesting—and relatively underexamined—case of cross-occupational collaboration: where professionals typically do not have formal authority over those who possess domain expertise, and where domain experts may not always be able or willing to provide this expertise.

Eliciting Domain Expertise for Technology Development

Processes of technology development present a paradigmatic case of cross-occupational collaboration for eliciting domain expertise in the absence of formal authority. While technology developers have technical skills and knowledge to build complex tools, they typically do not have domain expertise about the specific problems that these tools are being designed to address. This expertise resides with domain experts—i.e., those who experience, work on, and are affected by these problems on the ground. To ensure that developers build tools that are beneficial and useful to those who use them on the ground, developers are often given the dictum "*follow domain experts*"—i.e., to identify a key set of potential users with domain expertise, and "shadow" them to understand their practices, needs, and preferences (cf. Margolin, 1997; Wilson, 1981; Hassenzahl & Tractinsky, 2006). However, in large, multidivisional organizations, domain experts are often embedded in departmental structures that are functionally separated from technology development teams. As such, in many cases, developers have no formal authority over domain experts in their organization. Technology development processes therefore regularly entail the elicitation of expertise in the absence of formal authority, where developers need to identify and gain access to relevant domain experts across their organization, and convince them to share their domain expertise.

The dynamics of such collaborations between technology developers and domain experts have been previously explored in the information systems (IS) and human-computer interaction (HCI) literature. Early IS studies of software development processes, specifically within the subdomain of “requirements elicitation,” have focused on how software developers might effectively learn the needs of their intended users, and in turn, incorporate these needs into their designed software solutions (Brooks & Bullet, 1987; Byrd et al., 1992; Boland & Tenkasi, 1995; Levina & Vaast, 2005). Within this broad subdomain, studies have focused on evaluating various techniques for software engineers to prompt, observe, interview, or survey domain experts (cf. Zowghi & Coulin, 2005; Goguen & Linde, 1993; Browne & Rogich, 2001), emphasizing the importance of cultivating empathy with domain experts, and triangulating domain expertise by flexibly incorporating different methods (e.g., observations, interviews, field visits, prototyping, etc.) Studies have also developed models and criteria for software engineers to select optimal techniques for their specific development projects (Chakraborty et al., 2010; Hickey & Davis, 2004; Tiwari et al., 2012), highlighting how optimal elicitation techniques depend on prior collaborative arrangements, the level of trust between developers and domain experts, and the degree of standardization/formalization of expertise within a given domain, among other factors (see also Wiesenfeld et al., 2022).

More recent works have focused on the communication breakdowns that arise when technology developers—who typically often use statistical, technical language—interact with domain experts—who also use their own specific jargon related to problems in their domain (Mao et al., 2019; Piorkowski, 202; see also Nielsen, Elmholdt, & Noesgaard, 2024). For instance, Piorkowski et al. (2021) discuss how collaborations between technology developers and domain expertise can come to be “lost in translation” due to communication gaps between them, and explore how developers can leverage shared mental models with domain experts to bridge these communication gaps. Relatedly, Mao et al. (2019) discuss how technology developers can communicate more effectively with domain experts when they seek to establish “process common ground”—i.e., agreement over broad processes, shared goals, and rules of engagement—rather than “content common ground”—i.e., shared understanding about the problem domain and possible solutions. And, Sosa-Hidalgo et al., (2024) show how developers tactically increase or decrease the complicatedness of their solutions when sharing their work with domain experts, and discuss how this practice might serve developers in convincing domain experts to cooperate with them.

However, while this literature has advanced various ‘best practices’ for technology developers to effectively collaborate with domain experts, some important shortcomings remain. First, prior work has tended to individualize problems arising during the elicitation of domain expertise: i.e., ineffective attempts to elicit expertise have tended to be attributed to individual developers selecting suboptimal techniques, and/or being unable to effectively carry out their chosen techniques. Relatedly, domain experts have been conceptualized as an abstract, empty category—quite simply, as those people who happen to know a lot about the problem that developers are trying to solve. Little attention has been paid to the *specific occupational/professional group*s the developers belong to, the professional norms and interests they represent, and how their collaborations are structured by the organizational contexts (e.g., functions and jurisdictions) within which both groups are embedded. Such occupational/organizational factors are likely to significantly shape collaborative efforts between developers and domain experts (cf. Chan & Hedden, 2023; Beane & Anthony, 2023; Koljonen & Chan, 2024; Lifshitz-Assaf, 2018; Evans, 2021; Bruns, 2013).

These shortcomings are starting to be addressed by a nascent stream of organizational scholarship that has sought to pay greater attention to occupational and organizational aspects of collaborations between developers and domain experts. For instance, van den Broek et al. (2021) examine the development of an AI system for hiring job candidates at a large organization, and discuss how developers and domain experts (here, HR professionals) go through “mutual learning cycles” during which the two groups develop shared representations through processes of negotiation and reflection on each other’s practices, and in turn, arrive at a new “hybrid practice” for combining AI and domain expertise. Truelove & Kellogg (2016) show how “moderate” developers and technologists (i.e., those who hold incrementalist—rather than radical—views about their occupation’s role in driving technological change) facilitate effective collaborations with “moderate” domain experts (e.g., marketers), who each strengthen their bargaining position by contrasting with more “radical” members of their respective occupations. And, Stice-Lusvardi et al., (2023) discuss how developers (specifically, data scientists) conceded to requests from domain experts to pursue illegitimate data science practices (e.g., using sloppy code to manage data, ‘peeking’ at experimental results before data was fully collected, etc.), to maintain sustained cooperation from these domain experts.

In seeing technology development processes as instances of cross-occupational collaboration, these studies have considerably advanced our understanding of how and when technology developers can be more (versus less) effective in eliciting expertise from domain experts. However, with a few exceptions (e.g., Stice-Lusvardi et al., 2023), prior work has largely treated the organizational context as a background condition within which collaboration between technology developers and domain experts takes place. Questions of non-overlapping lines of authority between developers and domain experts—which emerge when one considers how the interactions between occupational groups are structured by the organizational context within which these groups are embedded—continue to be relatively underemphasized in prior work. Specifically, little attention has been paid to the ways in which both developers and domain experts are embedded within different units in the organization, characterized by complex lines of authority, jurisdictions, and task structures, and how these complexities, in turn, structure the cross-occupational collaborative processes between the two groups. Such considerations occupy one of the central aims of this research, helping us to further refine our focus on when and how professionals are able (or unable) to elicit expertise during technology development from domain experts over whom they have no formal authority.

Methods

Research Setting

We draw on data collected during a four-year field study at Weave (a pseudonym)—a large, New York-based multinational fashion company—between October 2019 and December 2023. In 2018, in response to widespread enthusiasm about the potential of data science and AI to transform work processes, Weave began a concerted effort to invest time and resources in scaling the use of data and AI across its business operations. As part of these efforts, the company’s Chief Operating Officer hired a Chief of Data Science and equipped him with an ambitious budget to build Weave’s first team of AI developers.

In October 2019, around the time we began our observations, the team comprised of ten AI developers with expertise in data science and machine learning, and evolved to more than fifty AI developers by the end of our study in December 2023. Weave’s AI development team was structured as an independent unit—reporting directly to the COO, with no functional overlap with the company's 'traditional' divisions (e.g., marketing, sales, retail operations). Similar to other large organizations structuring their AI initiatives, Weave had structured its AI development team to reflect its strategic priorities, such that the developers would not have to compete with other units or divisions for resources. However, this arrangement also left developers with no formal authority over the domain experts from whom they needed to elicit expertise. Domain experts were embedded in divisions and departments with their own lines of authority, and the developers played no role. The AI developers had no formal authority over the domain experts and thus were in no position to issue commands or sanctions, or otherwise coerce domain experts to "get their time," let alone get their support and cooperation to elicit domain expertise.

Data Collection

The data employed in this paper draws on a larger project of Weave's AI development initiatives and its organizational and strategic impacts. Between October 2019 and December 2023, the first author collected data about the overall evolution of Weave's high-level strategies and priorities related to AI through regular meetings with the Chief of Data Science, as well as data about specific AI development projects through participant observation and interviews with members of the AI development team.

Data on Weave's portfolio of AI strategic initiatives and projects was collected through monthly steering meetings with the Chief of Data Science, which occasionally also included other senior members of the AI development team. During these meetings (approximately 1 hour each), the Chief of Data Science would discuss the high-level strategic goals of the AI development team, what senior leadership expected of this team, and the progress of the various AI projects they were working on. This was often done while reviewing recent slide decks that had been presented to the C-level suite to update them on the status of the firm's investments in AI. Sometimes, during these conversations, the first author would probe and ask questions about certain projects—or ideas for future projects—that seemed particularly compelling, in part because they represented substantive attempts to augment the work of Weave's employees using AI. In such cases, the first author would ask for permission to talk to and observe the AI developers as they tried to bring these projects to fruition.

Throughout our study period, we followed the actions of the team of AI developers as they went about building various AI tools to optimize internal processes and solve organizational problems. In rare cases, the AI development team was given specific and well-defined problems to solve, in which case they could jump right into developing AI tools. In most cases, however, their project briefs were broad and unspecific: typically describing a general operational problem to solve, or a performance metric to optimize. In such cases, in order to determine what kind of AI tool they should build, developers needed to identify, gain access to, and elicit expertise from key domain experts in the organization. As we followed the developers in their attempts to navigate Weave's organizational hierarchies to elicit domain expertise from finance managers, store designers, product designers, marketing managers, and merchandisers, to name a few, we observed considerable variation in their effectiveness across different projects. The data collected and analyzed in this paper specifically focuses on comparing two such projects.

Project 1. In the first project, conducted from May 2020 to May 2022, the developers were tasked with building an AI tool to improve the ways in which products were distributed across the supply chain—minimizing the risk of product obsolescence, and optimally matching products to customer demand across different store locations. In Weave's supply chain, approximately 5,000 products were distributed each week through a network of approximately 180 stores across North America, where potential customers might have the opportunity to purchase them. However, these distributions were not always optimal. Certain stores saw over-allocations (i.e., when shipped quantities exceeded customer demand), resulting in overstocked inventories, while other stores saw under-allocations (i.e., when shipped quantities were lower than customer demand), resulting in stockouts and losses in potential revenue. Thus, Weave's executives hoped that the developers could build an AI tool to help those in charge of the company's supply chain to consistently produce optimal allocation decisions.

Project 2. In the second project, conducted from July 2021 to December 2023, the *same set* of AI developers were tasked with building an AI tool to optimize "retail productivity"—an organizational metric defined as the ratio of sales-to-payroll costs—for Weave's retail stores. On any given day, each store would see a certain amount of 'traffic'—potential customers who came into the store to browse products. Sales would be generated when this "footfall" traffic was converted to revenue—i.e., when these potential customers bought a product and completed their purchases. To facilitate this conversion, store managers had to schedule shifts of sales associates to guide customers through product selection, “runners” to retrieve the correct product sizes from the warehouse, and cashiers to facilitate sales transactions. However, overstaffing on off-peak days or locations often resulted in unnecessary cost spikes without a corresponding increase in sales, whereas understaffing in popular store locations and/or during peak periods resulted in lost sales opportunities. Thus, Weave’s executives hoped that the developers would build an AI tool that would help each store optimally plan their workforce scheduling needs in ways that maximized their revenues with the least possible payroll costs.

Rationale. We selected the above two projects as “strategic research sites” (Merton, 1987) to examine our research question for the following reasons. In both projects, the same set of AI developers were involved in the development process. They started with similar levels of uncertainty in that they lacked domain expertise about the processes (i.e., allocation and workforce scheduling), including the roles, responsibilities, and current work practices of the people involved, as well as the tools that were already being used. In both projects, aligned with prior research on requirements gathering, the developers employed similar strategies and ‘best practices’ to capture domain expertise: starting with identifying domain experts, interviewing and observing them at their work, and finally, parsing this information to scope out a specific problem that could be solved using AI-based tools. Even more importantly, the case where developers were able to effectively elicit domain expertise (Project 1) was *temporally prior* to the one where they were ineffective in doing so (Project 2)—such that in Project 2, developers were pursuing the same strategies for identifying, accessing, and interacting with domain experts that they had already effectively deployed in their previous successful project. Whereas in Project 1, the AI developers were able to effectively elicit domain expertise, which resulted in the development of a well-received AI tool that was widely rolled out and used across the organization, in Project 2, the developers were ineffective in eliciting domain expertise. After a long period of experimentation, they were asked by Weave’s senior management to cease working on their AI tool.

For each of these selected projects, the first author would embed himself as part of the AI development team throughout the project duration, and closely follow how the AI tool or solution was designed, developed, and deployed. Data collection for each project followed a similar process. The initial step was to schedule a meeting with the internal team of AI developers assigned to that specific project, to discuss in detail the project's goals and timelines, relevant domain experts that had been identified (if any), and the work that had been accomplished thus far. This allowed us to be up to date with the information that the developers already possessed, and to evaluate whether an AI solution had already been identified or still needed to be conceptualized. By this point, with the explicit sponsorship of the Chief of Data Science, the first author would be seen by the developers as part of their team, albeit as an 'external' observer. After these initial meetings, the first author would have recurring weekly meetings with the AI developers to get updates on project progress, and specifically on what conversations they were planning to have with key domain experts.

During these meetings, he would capture the open questions that AI developers had at various stages of the project, as well as the plans that developers were pursuing to address these questions. The first author was also invited to brainstorming sessions where tentative solutions for the project were conceptualized. He also had access to the developers' internal communication platforms (e.g. Slack channels created for the project)—where they shared communications, updates, or material developed during the project (e.g. notes, diagrams, reports, or documents). Whenever possible, he would also ask to attend meetings they held with domain experts (e.g., site visits to Weave's retail stores, interviews, 'shadowing' sessions, etc.), or to schedule meetings with these domain experts directly to conduct private interviews (i.e., without the presence of other AI developers). The primary goal of these observations and interviews was to understand the practices that the AI developers used when they were trying to elicit domain expertise, and to map out key factors that enabled or hindered them in eliciting this expertise. Table 1 summarizes the data collected about the two specific AI projects analyzed in this paper.

==== Insert Table 1 Here ====

Data Analysis

Our data analysis was iterative and proceeded along three stages (Charmaz, 2006; Glaser & Strauss, 2017). In the first stage, we open-coded our field notes, interview transcripts, and archival documents, focusing on the different kinds of information the AI developers were receiving through their interactions with domain experts in each project, how they were interpreting and using this information, and how they felt about the overall status of each project. At this stage, we discerned that the developers had obtained a clearer understanding of the supply chain processes in Project 1 relative to the workforce scheduling processes in Project 2. Codes associated with field notes from internal meetings between AI developers in Project 1 indicated that developers felt relatively more confident about their ability to identify a useful tool for supply chain allocation (e.g., “smooth sailing”, “satisfied”) relative to a tool for workforce scheduling in Project 2 (e.g., “feeling lost”, “frustrated”, “confused”).

In the second stage, we then took a within-project perspective, producing a project case history and chronological reconstruction of the main events and timeline for each project. At this stage, we traced the different interactions that took place between the developers and domain experts, focusing on how the developers’ understanding was evolving (or not) over time, and how well (or poorly) they were able to narrow down on a specific AI tool to build for each project. In each case, we first mapped the various tasks involved in making decisions about product allocation or workforce scheduling, and the actors in charge of performing these tasks. For each task, we also noted the various pre-existing tools that domain experts and workers on-the-ground were using to carry out their work. Finally, for each project, we traced the various interactions between developers and domain experts over time—i.e., who the developers were speaking to at different stages of the project, what the stated purpose of these conversations was, and how useful (or not) the developers found these interactions. Throughout this process, we wrote descriptive memos describing the challenges (or lack thereof) faced by developers in identifying, gaining access to, and learning from relevant domain experts.

At this stage, we were able to identify that developers faced certain barriers in identifying and gaining access to domain experts in Project 2 that they did not experience in Project 1. For instance, our interaction timelines indicated that in Project 2, relative to Project 1, developers took longer both to identify *which* domain experts to talk to, as well as to establish *connections* (e.g., to set up meetings) with their target domain experts. Relatedly, in Project 2, we observed that developers repeatedly moved back and forth in interacting with different groups of putative domain experts throughout the project, typically after receiving conflicting information from one group that they needed to double-check with another. However, in Project 1, once developers had identified a specific group of experts, their subsequent interactions remained largely focused on members within this group. We also discerned that interactions between developers and domain experts were generally more informationally rich and useful for developers in Project 1 than in Project 2. Our analysis of Project 1 made clear that successive interactions between developers and domain experts resulted in developers cultivating a richer understanding of allocation processes, and in turn, narrowing down the scope of their intended AI solution. In Project 2, however, this did not seem to be the case: developers remained uncertain about *how* workforce scheduling decisions were made on the ground, and about what sort of AI tool they should develop, even after multiple rounds of interacting with different domain experts. These two sets of issues—i.e., (a) issues related to identifying and gaining access to domain experts, and (b) issues related to the elicitation of domain expertise and narrowing down on a problem definition—became a key focus for our subsequent analysis.

Finally, in our third stage of analysis, we carried out theoretical coding of our fieldnotes and transcripts, with a focus on explicating the mechanisms underlying the effectiveness (or ineffectiveness) in developers' attempts to identify, gain access to, and elicit expertise from domain experts across the two projects. Here, we performed a comparative analysis of the data and codes of each project, abstracting our previous analyses to generate higher-order analytical categories. For instance, we abstracted codes such as "making allocation decisions is central to the occupational identity of allocators" and "allocators perform few other tasks besides making allocation decisions" to the higher-order analytical category, "task centrality". Similarly, we abstracted codes such as "maximizing sales [rather than workforce scheduling] is central to the job of store managers", and "store managers perform a wide variety of different tasks [besides workforce scheduling]" to the higher-order analytical category, "task peripherality". At this stage, we observed that our higher-order analytical categories all pertained to aspects related to the *task structure* of domain experts' jobs, and/or the *organizational structures* within which both developers and domain experts were embedded.

This interplay between task and organizational structures became an important anchor for our subsequent theorizing. Throughout this phase, we also engaged closely with relevant literature on cross-occupational collaboration (DiBenigno, 2018; Huising, 2015; Karunakaran, 2022; Stice-Lusvardi et al., 2023; Truelove & Kellogg, 2016), expertise in and around organizational contexts (e.g., Eyal, 2013; Collins & Evans, 2002; Heimstädt et al., 2024; Monteiro, 2024; Almandoz & Tilcsik, 2016; Pakarinen & Huising, 2023), and technology development (Piorkowski et al., 2021; Stice-Lusvardi et al., 2023; van den Broek et al., 2021). We found Eyal's (2013) analytic distinction between 'experts' and 'expertise' especially generative for our analysis. Subsequent rounds of analysis focused on unpacking the ways in which the interplay between task and organizational structure rendered the *legibility* of domain experts to AI developers in Project 1 (versus their *illegibility* in Project 2) and shaped the *concentrated* nature of expertise in the hands of domain experts in Project 1 (versus the *dispersed* nature of expertise in Project 2), which became another important anchor for our theorizing.

Findings

We find that the interplay between the task and organizational structures within which domain experts are embedded influenced the effectiveness with which AI developers elicited expertise from domain experts over whom they had no formal authority. In the first project, when domain experts were performing tasks that were central to their occupation, had clear jurisdiction over these tasks, and enacted these tasks homogeneously relative to other members of their occupation, AI developers were able to identify and gain access to relevant domain experts, and effectively elicit their domain expertise. In the second project, when domain experts were performing tasks that were peripheral to their occupation, had ambiguous jurisdiction over these tasks, and enacted these tasks heterogeneously relative to other members of their occupation, the same set of AI developers found it difficult to identify and gain access to domain experts, and were unable to piece together the disparate and dispersed information they had gathered to build a robust understanding of the domain expertise needed to build AI tools.

We organize our findings as follows. The first subsection unpacks how the interplay between task and organizational structures that rendered the key domain experts legible (or illegible) to the AI developers, which in turn helped (or hindered) the developers from identifying and gaining access to the domain experts. The second subsection explores the developers' attempts to elicit expertise from the domain experts they had gained access to. Here, we discuss how task and organizational structures shaped how domain expertise was concentrated (versus dispersed) in the hands of key domain experts, which made it easier (versus more difficult) for developers to elicit this expertise. Appendix A provides additional evidence for these findings, while Table 2 provides an illustrative summary of key analytical concepts.

=== Insert Table 2 here ===

AI Developers' Attempts to Identify and Gain Access to Domain Experts

At the onset of both projects, the AI developers did not know which domain experts might be best suited to providing expertise about supply chain and retail processes followed at Weave. As such, they initially explored their own network within Weave to obtain a high-level understanding of these two processes, to narrow down specific domain experts to talk to and follow. Once they had identified their targets, they moved to gain access to these domain experts. Beyond simply scheduling one-off meetings, they sought to ensure that these experts would be willing to cooperate with their requests throughout the duration of their project: i.e., to have their work practices observed, be interviewed whenever necessary, and expend time and effort to support emerging project needs to build the AI tools. As we show, in Project 1, developers were able to clearly identify and gain access to domain experts, whereas in Project 2, they struggled to even identify who the relevant domain experts were, let alone how to gain access to them.

Effective Identification and Access to Domain Experts in Project 1. At the onset of Project 1, the developers were uncertain about how Weave's supply chain processes were organized, as well as which domain experts might be best placed to help them understand these processes. To identify their target domain experts, the AI developers leveraged their contacts to schedule a meeting with the company's Head of Supply Chain, who provided a broad overview of the different tasks comprising the supply chain process. As the developers learned:

> "Within [Weave]'s supply chain, a catalog of roughly 5,000 products… are shipped from different manufacturing facilities to a regional fulfillment center (FC) … and then divided through a capillary network of roughly 180 stores within the region, where customers can purchase them. This *last-mile-delivery* from FC to retail stores is technically called the "allocation process", and it is supervised by a team of allocators *who decide which store to ship available products, which determines the level of inventory of each store.*" (Field notes from a meeting between developers and the Head of Supply Chain (#RET_001); emphasis added)

The developers returned from this meeting with a relatively clear understanding of the different activities and tasks comprising Weave's supply chain process: including the design of products, macro- and micro-budgeting, production, and the distribution of produced inventory across stores. As they learned, each of these tasks was controlled by a specific occupational group: designers were responsible for designing products, merchandisers planned long-term budgets across regions, buyers made decisions about product assortment, and allocators distributed inventory across retail chains. Moreover, each task—and its associated occupational group—was coordinated by a manager who directly reported to the Head of Supply Chain. At an internal discussion shortly following their meeting with the Head of Supply Chain, the developers were able to produce a clear map of this process. Refer to Figure 1 for a process map constructed by developers following their meeting with the Head of Supply Chain.

==== Insert Figure 1 here ====

As the Head of Supply Chain mentioned, allocators were responsible for the "last-mile delivery" of inventory across retail stores, making decisions about how much inventory should be shipped to each store. This point seemed to resonate with developers—during subsequent internal meetings, they came to a consensus that any inventory problems faced by retail outlets could reasonably be attributed to the decisions made by allocators. Circling 'allocation' in the process map they had created (Figure 1), they agreed that improving the task performance of allocators would lead to marked improvements in the overall performance of the supply chain. As such, they decided to focus their efforts on optimizing the allocation process and began making moves to gain access to Weave's *allocators*.

Jurisdictional Clarity and Clear Lines of Authority. To gain access to these allocators, the developers leveraged their relationship with the Head of Supply Chain to secure a meeting with the Head of Allocators—who was his direct report. As they learned during this meeting, Weave employed a total of 10 allocators to manage the allocation of different product categories to retail stores, all of whom directly reported to the Head of Allocators. The Head of Allocators reaffirmed that these 10 subordinates were indeed best placed to help the developers obtain an understanding of the domain of "allocation" and its underlying process. Once the nature and scope of the developers' project were fully explained to him, he offered to broker connections with each allocator. As he asserted:

> "I can email the group and I'll set you up with different allocators. I think it's really important to see, you know, how different products get allocated. Women's handbags can be a little bit different than men's bags, for example. Each allocator can provide a little clue on the process—even though they are similar overall. So, I think talking with them would be cool and quick… Let me take that back to my team, and I'll definitely let you know." (Field notes from a meeting between developers and the Head of Allocators, #ALL_001)

The Head of Allocators proved to be an ideal intermediary for the AI developers. Since his primary responsibility was to coordinate the activities of the ten allocators, he interacted with them on a daily basis, and had detailed knowledge of how different allocators went about performing their work. Crucially, his role as supervisor guaranteed him direct formal authority over all ten allocators. His sponsorship was therefore particularly instrumental in motivating the allocators to expend time and effort in interacting with developers to share their expertise about allocation processes. In individual emails to each allocator, he introduced the developers, explained the scope of their project, and instructed the allocators to "connect [with] and send some timeslots to [the developers] for interviews". Allocators, in turn, followed this direct instruction, and the AI developers were able to set up initial meetings with the allocators shortly thereafter.

==== Insert Figure 2 here ====

As Figure 2 indicates, Weave's supply chain process was characterized by a *single, clear line of authority*: starting from the Head of Supply Chain who directly supervised the Head of Allocators, who in turn directly supervised all of Weave's allocators. As such, developers were able to identify and gain access to the domain experts (i.e., allocators) who worked on-the-ground by starting at the top (i.e., with the Head of Supply Chain) and navigating this single line of authority. People in supervisory roles (i.e., the Head of Supply Chain and the Head of Allocators) were both able to speak comprehensively about what their subordinates were doing in their daily work, as well as broker connections to their direct reports. In this way, developers were able to both identify that allocators were best placed to provide them with the domain expertise they were seeking to develop the AI tool, as well as effectively establish connections with these target domain experts.

Task Centrality. As the developers started to interact with allocators, they got a window into how the allocators viewed their own work:

Developer: "I'm trying to understand how you decide whether to stick with the recommendation that I can see in this column [sales forecasts] or deviate from it."
Allocator: "Well, it's not just about the numbers the system spits out. We look at the forecast, sure, but we also consider things like upcoming events or recent marketing campaigns. There's a lot of nuance to it."
Developer: "But how do you decide when to override the recommendation?"
Allocator: "It's hard to explain. It's a feeling you get after doing this for a while. Sometimes, the system might suggest sending more stock to a store, but I know that store won't move that product as quickly. It's not just about data!"
(Field note from an interaction between an AI developer (#DEV_002) and an allocator (#ALL_002))

As this exchange suggests, allocators viewed themselves as playing an important role in parsing imperfect sales forecasts to produce optimal allocation decisions. Making allocation decisions was a core task for them, and they felt that they had unique capabilities for performing this task well. Throughout their subsequent interactions, allocators would often and actively check in with developers: to ask about the status of the project, and to give pointed, specific suggestions to the developers about what kind of tool they should develop. Allocators viewed the AI developers' efforts as having the potential to significantly reshape the core tasks of their jobs, and they were particularly invested in ensuring that the developers had as much information as they needed to build a useful tool for them. The developers, in turn, found it relatively easy to convince the allocators to expend time and effort in cooperating with them and supporting their project.

Ineffective Identification and Access to Domain Experts in Project 2. The brief for Project 2 had called on AI developers to create a tool that could optimize "retail productivity": defined as the ratio of payroll expenditures to revenues for Weave's ~180 retail stores. Similar to Project 1, during internal meetings at the start of Project 2, the developers were unsure about where to begin their exploration: it was not clear to them what exactly they were being asked to optimize. However, they knew that Weave's retail finance managers, who happened to be co-located with them in the same administrative office, were in charge of monitoring payroll costs and revenues across Weave's stores—and as such, might have some insight into how retail productivity was being evaluated. This seemed like an appropriate entry point, and so they decided to conduct exploratory interviews with this team. During these interviews, the retail finance managers explained how the "retail productivity" metric was measured and evaluated, and how they interpreted differences in productivity between Weave's different retail stores. As the developers learned:

"Retail finance measures "retail productivity" by the ratio of payroll to sales. Sales are the expenditures of customers who enter the store and make a

> purchase. Payrolls are the hourly cost of the people who are in the store on a given day. [Retail finance managers] track this metric primarily because they want to see if the stores have a sufficient level of sales compared to the cost of payroll. [...] *They know that for the same level of sales, there are different payroll costs. They think this is due to a lack of skills on the part of the store managers*." (Field notes from a meeting between developers and a retail finance manager (#BUS_15); emphasis added)

Over multiple interviews with the retail finance managers, the developers learned that store managers had discretion in assigning labor hours to each employee: comprising a fixed number of hours allocated to all employees, plus optional additional flexible hours depending on store needs. They also learned that payroll costs varied significantly across different stores, even those with similar sales levels. The finance managers did not have any explanations for this difference but seemed to suspect that certain store managers simply lacked the skills to make efficient workforce scheduling decisions, and for some inexplicable reason, were regularly scheduling much more labor hours than they actually needed. Absent better information, the AI developers took the reasons provided by the retail finance managers seriously, and decided to focus their efforts on optimizing workforce scheduling. In turn, they set out to gain access to the domain experts putatively in charge of workforce scheduling: i.e., *store managers*.

Jurisdictional Ambiguity and Overlapping Lines of Authority. It quickly became clear to the AI developers that connecting with all relevant store managers involved in workforce scheduling would be a difficult undertaking. Weave employed approximately 180 store managers, and given limited resources, the AI developers judged that it would be impossible to interact with all of them. Therefore, the developers deliberated about how they might sample a smaller number of store managers to follow. Using data about retail productivity they received from the retail finance managers, they constructed nine ‘clusters’ of stores with similar productivity metrics, and randomly sampled stores—and store managers—from each group. Several heated internal discussions about the suitability of this approach followed. As one developer put it:

> “We want to be able to ensure we're interviewing store managers from each of these groups… I don’t know if 9 interviews will be enough? Right now there are nine [clusters], and … with a minimum of one store [per cluster], that's nine interviews. But in some of these buckets, there are 43 [stores]. 9 interviews would be the *absolute* minimum [his emphasis]! We want to be able to capture all these different aspects that we're interested in looking at...” (Field notes from an internal meeting between AI developers; Speaker: #DEV_005)

In private conversations with our participant observer, several AI developers admitted that they were skeptical about whether their sampling strategy would be sufficient for obtaining a clear understanding of store managers' different scheduling practices. However, after multiple inconclusive internal meetings, in the interest of moving the project forward, the AI developers decided to start by interviewing one store manager from each of the nine ‘clusters.’

Their difficulties, however, did not cease with this decision: developers also faced considerable practical difficulties even in connecting with the target store managers. Unlike the allocators, the ~180 store managers were not grouped together under the supervision of a single leader. Instead, they were grouped into 36 clusters, each led by a district manager responsible for overseeing 2 to 6 stores. These 36 district managers were themselves embedded in complex and overlapping lines of authority, comprising different area managers and regional managers, leading up to the VP of North America for Operations Management. As illustrated in Figure 3, given this complex organizational structure, no one was able to broker connections to the target domain experts—the nine store managers—directly. Their high-level contacts (i.e., retail finance managers) were unable to broker connections to store managers, or to directly order store managers to spend time and effort in interacting with developers given they were part of a different authority-order. Instead, the developers had to go sequentially to regional managers, who then connected them to district managers, who then connected them to the respective store manager under their supervision. The AI developers then had to repeat this process for *each* of their identified subset of store managers.

==== Insert Figure 3 here ====

In summary, as Figure 3 illustrates, Project 2 was characterized by *multiple, overlapping lines of authority*. Consequently, the group in charge of monitoring and evaluating the “retail productivity” metric that developers had been tasked with optimizing (i.e., retail finance managers) were neither able to speak authoritatively about how workforce scheduling decisions were made on the ground, nor could they broker connections with store managers who made these decisions on a daily basis. Moreover, Weave's complex and varied reporting structures, including different district and regional managers, as well as the sheer larger number of employees involved in this process, introduced further complications to the developers' efforts to establish connections with store managers. As such, the developers' approach of starting as close as possible to the top of Weave's retail operations processes—an approach *similar* to the one that had worked well for them in Project 1—proved ineffective both for *identifying* exactly which actors had the domain expertise they were seeking, as well as *establishing connections* with these domain experts.

Task Peripherality. Nonetheless, through painstaking and drawn-out efforts, AI developers eventually managed to connect with five out of the nine store managers they had sampled and identified (as well as their supervisors, the district managers). Yet, their difficulties continued: developers found that these domain experts were not particularly forthcoming with details about the workforce scheduling practices as well as their everyday work practices. During their site visits and interviews, AI developers noticed that the store managers seemed distracted and generally uninterested in cooperating with them. When our participant observer followed the developers to a store visit in Los Angeles, the group was made to wait unattended in an administrative office for close to half an hour, and even then, only managed to speak with an assistant store manager. The store manager, as they were told, was "*busy elsewhere, because they were changing the products displayed in the store windows*." Similar experiences were observed during other site visits: store managers often seemed "too busy to spend time" meeting with and attending to the developers' requests, and only begrudgingly consented to one-off interviews and brief shadowing sessions.

These difficulties in securing the time and attention of store managers may be traced to the task structure of these domain experts' jobs. During site visits, AI developers saw store managers juggling a wide variety of tasks, including coordinating customer-facing activities, tracking delivery orders, organizing visual merchandising assortments, managing external relationships, and reporting to district managers, to name a few. Workforce scheduling seemed to be *peripheral* to the store managers' jobs relative to these other so-called "office tasks". As such, developers found it difficult to convince store managers to expend time toward supporting efforts to augment a task—workforce scheduling—that the store managers believed was a peripheral task in their everyday work.

Analytical Summary. As the preceding two subsections have shown, the ease or difficulty by which developers identified and gained access to domain experts was shaped by the task and organizational structures within which both groups were embedded.

In Project 1, the developers found it easy to gain access to allocators, for two interrelated reasons. First, as illustrated in Table 2 (Rows #2 – 10), *jurisdictional clarity* surrounding the task of allocation—i.e., (a) the fact that allocators had undisputed control over the turf of allocation tasks, and (b) allocators were embedded in clear lines of authority—meant that developers could identify which domain experts they should focus on with relative ease, and in turn, indirectly leverage the formal authority of key brokers (i.e., Head of Supply Chain, Head of Allocators) to convince these target domain experts (i.e., allocators) to interact with them. Second, as illustrated in Table 2 (#11 – 12), *task centrality*—i.e., the fact that the task of allocation was a central task performed by allocators—meant that allocators had clear incentives to cooperate with developers, since any tool that developers built to augment the task of allocation could significantly reshape the core of their daily work. As such, allocators were motivationally invested in ensuring that developers would obtain a rich understanding of their domain expertise, such that they could build a useful tool to improve Weave's allocation processes.

Conversely, in Project 2, the developers' efforts to identify and gain access to store managers remained mired in difficulties. Mirroring Project 1, these difficulties can be traced to two interrelated reasons. First, as illustrated in Table 2 (#3 – 10), the task of workforce scheduling was characterized by *jurisdictional ambiguity*. Developers found it difficult to pin down who exactly was involved in performing and evaluating workforce scheduling decisions, and their decisions to interact with the retail finance team and store managers were largely governed by choices of convenience, rather than the kind of deliberative certainty that had emerged in their choice of allocators in Project 1. Moreover, store managers were embedded in complex and overlapping lines of authority, comprising retail finance managers, district managers, and area managers, to name a few. No one was in a position to broker connections with all store managers, and as such, the developers instead had to painstakingly navigate complex organizational hierarchies to gain access to each of their target domain experts. Second, as illustrated in Table 2 (#11 – 12), the developers' difficulties in persuading domain experts to support their efforts may also be traced to the *task peripherality* of workforce scheduling to the job of store managers. Workforce scheduling occupied a relatively minor aspect of the daily work of store managers, and as such, they were not particularly invested in helping developers obtain a deep understanding of the domain expertise surrounding this task.

From Domain Experts to Domain Expertise: Enablers and Barriers in AI Developers' Attempts to Elicit Domain Expertise

After AI developers had identified and gained access to domain *experts*, they moved to elicit their domain *expertise*. At this stage, developers worked to observe and interview domain experts as they performed their job, to understand the expertise involved in making allocation (Project 1) and workforce scheduling decisions (Project 2). In so doing, they hoped to be able to define specific problems in these two domains that could be addressed with AI tools. As we discuss below, in Project 1, domain experts were able to clearly communicate their expertise to the developers, who, in turn, were able to narrow down to a well-defined problem and develop AI solutions to address that problem. However, in Project 2, different domain experts provided partial and conflicting accounts of the expertise involved in workforce scheduling, which ultimately proved insufficient for the developers to fully specify what kind of tool they should build.

Effective Elicitation of Domain Expertise in Project 1. In Project 1, it seemed to take only a few interactions with allocators for the developers to glean a clear picture of how allocators made their decisions, and what tools they used in their work.

Jurisdictional Clarity and Exclusive Control over the Turf of Allocation Tasks. During their interactions, allocators seemed well-equipped to provide the AI developers with a clear overview of how day-to-day allocation decisions were made at Weave. As the developers learned during one such interaction:

> "[Allocator_2] showed us the interface that he was using for making his decision. [...] The decision of how many units to send to each store is just based on the forecast of sales per store. Products were divided based on the proportion of total sales that a store was supposed to make. For example, they showed us a store that was forecasted to sell about 20% of the entire demand of a certain SKU, so the system recommended sending 20% of the total inventory of that SKU to that store." (Field notes from a meeting between AI developers and an allocator (#ALL_002))

The developers learned that allocators used a single integrated software platform to evaluate and submit allocation decisions. This platform provided sales forecasts for all Weave stores, based on which the allocators made decisions about the inventory (i.e., what type of product and how many) to be shipped to each store. Therefore, for allocators, 'optimal' allocation decisions entailed accepting the platform's recommendations when they felt that the sales forecast for a given store was reasonable, and deviating when they felt they had a better estimate of how much inventory the store might need. Such deviations were a frequent occurrence: the allocators seemed to all agree that, given their extensive experience, their own estimates were often better than the system's forecasts. "It's not just about the data", as one allocator put it. "If you had perfect [sales forecasts], you wouldn't need us anymore… But you don't!"

Throughout their observations and interviews with allocators, the developers were struck by how authoritatively the allocators were able to outline the specific processes they followed when making allocation decisions. They clearly had exclusive control over the turf of allocation decision-making tasks: i.e., they did not need to check in with any other supply chain personnel as they made their day-to-day decisions, and their decisions were accepted without further scrutiny by Weave's distributors who went on to move inventory to different stores as they had decided. The developers did not feel the need to speak to any other occupational groups to improve their understanding of Weave's allocation processes.

Task Enactment Homogeneity. Over time, the developers also found that reports from different allocators were largely consistent with each other. Even though different allocators were working on significantly different product categories and store contexts, their day-to-day tasks seemed mostly homogenous. All allocators brought up similar complaints about inaccuracies in sales forecasts, and spoke similarly about how they would override the system as often as they needed to. Therefore, the developers felt that they likely did not need to repeatedly interview all ten allocators, and could instead focus on the few who seemed particularly articulate and interested in supporting their efforts. Reflecting on his experiences interviewing two different allocators, one developer suggested:

> "[#Allocator_2] is the one that we understand better. He gives us valuable insights. *After all, they're all essentially doing the same thing.* Talking with him is like talking with [Allocator_6]" (Field notes from an internal meeting between AI developers; Speaker: #DEV_002; emphasis added)

In this case, domain expertise surrounding the task of allocation was *concentrated* in the hands of Weave's allocators, and all ten allocators had broadly consistent views of what this expertise entailed. As a result, by closely following a small number of allocators who were willing and able to provide detailed insights about how allocation decisions were made, developers were able to elicit a rich understanding of the domain expertise they were seeking.

Ineffective Elicitation of Domain Expertise in Project 2. For Project 2, eliciting domain expertise proved much more cumbersome. The developers went into the field with a “working theory” shaped by their prior conversations with the retail finance managers—i.e. that *some* individual store managers were scheduling their workforces less efficiently than others—with higher payroll costs per unit of sales. The challenge was to understand *why* this was the case and conceptualize and develop an AI tool to overcome that issue. However, after a few interviews with store and district managers, the developers realized that modeling the task of workforce scheduling was much more complex than they had initially imagined.

Jurisdictional Ambiguity and Contested Control over the Turf of Workforce Scheduling. During an initial meeting with a district manager, the developers learned that there was an ‘ideal’, standardized way in which workforce scheduling decisions were supposed to be made. As the district manager explained:

> “Each week, employees are mandated to submit their availability for the upcoming two weeks via the workforce management platform. Subsequently, store managers retrieve sales targets from the corporate platform and use the Census model to determine the maximum number of work hours to allocate for that week. This model provides the total amount of hours that store managers then distribute among employees, adhering strictly to the availability each employee has provided. This distribution is done iteratively until they are able to find an optimized configuration… Once finalized, this schedule is delivered to employees." (Field notes from a meeting between developers and a district manager, #RET_005)

This ‘ideal’ model, in large part, reflected the interests of the retail finance managers, whose mandate was to ensure that stores were meeting corporate targets in terms of revenues and costs. As the retail finance team confirmed, optimal workforce scheduling entailed the optimization of a key metric: Payroll Percentage of Sales (PPoS), calculated as the ratio of sales and the average cost of labor per the total number of hours allocated for a given week. To claim turf over the task of workforce scheduling—i.e., to ensure that store managers were sensitive to this PPoS metric when making scheduling decisions—retail finance managers had commissioned the development of the “Census model”: a tool that provided store managers with an upper-bound estimate of how many total work hours they could assign each week.

To get these estimates, store managers had to download weekly sales forecasts from a centralized corporate platform and input them into the Census model spreadsheet. The model would then help them estimate what would happen to their PPoS targets at different levels of total payroll hours. As a store manager illustrated:

> "For example, this week's sales are supposed to be [...]$ [pointing her finger at a cell in the spreadsheet], so to not get over 4.4% [PPoS target] we cannot schedule more than 1900 hours summing the workload of all the store employees" (Field notes from a site visit to a retail store; Speaker: #RET_009 (store manager))

In practice, however, the developers found that the Census model did not serve a meaningful role in guiding workforce scheduling decisions for many store managers. Several store managers reported that the Census model consistently underestimated their labor needs, and as a result, many had negotiated arrangements with their district managers to bypass the model whenever necessary. One store manager even openly admitted to 'gaming' the model:

> "Eventually, you start to play with [the Census model] to understand how to reach the [target] 5% payroll sales—you put in a random number and then adjust it until it reaches 5%" (Interview with store manager, #RET_010)

Rather than using the tool as a workforce scheduling guide, certain store managers were often manipulating their input to meet their desired payroll targets. Moreover, even when the model specified the maximum number of hours a given store could schedule, this recommendation was not always followed. During a site visit, a store manager (#RET_009) showed the developers a spreadsheet, pointing her finger to a column for a week in February 2023 where she had scheduled 2116 hours—far above the Census model's recommended maximum of 1814 hours. To the developers, it seemed like store managers were simply ritualistically complying with the retail finance team's push to constrain payroll costs—making sure to input their forecasts into the sheets so that their use of the tool would be formally logged, but not actually incorporating its output into their decision-making.

To further complicate matters, the developers also realized that a third group of experts—district managers—had their own view of what workforce scheduling should look like. District managers seemed to view workforce scheduling primarily as a traffic-matching problem. The metrics they cared the most about were conversion rates and idle times for store workers. They seemed particularly concerned about over-assigning too many workers to certain stores and time slots where customers were unlikely to show up. When asked about what workforce scheduling decisions should ideally be based on, one district manager responded:

> "As a store manager, you should be active. You know, even outside of the holiday season, you should be aware of what is happening in your mall—what sorts of events they are doing…these are all traffic drivers! … There was a big campaign earlier this year for the mayoral elections of [US city]: one of the candidates running was the owner of this mall. So, in November, he hosted these special events that impacted our store traffic. *This* [his emphasis] should determine our scheduling." (Interview with district manager, #RET_005).

However, while considerations about store traffic predictions did occasionally arise during developers' interactions with store managers, it seemed that their primary consideration was altogether different. When making workforce scheduling decisions on the ground, store managers seemed to care most about how these decisions might affect the morale and motivation of their key employees. Each store employed a small number of workers on 'P1' contracts—Weave's highest tier of full-time contracts—who were typically experienced salespeople who often also took on additional supervisory and administrative responsibilities to support the store managers. As the developers learned, most store managers made sure that P1 workers were always assigned the maximum allowable hours on their contract: 40 hours per week. Occasionally reducing their working hours would likely save payroll costs—something retail finance managers would strongly endorse—but store managers worried that if P1 workers faced the prospect of fewer hours and lower wages, they would feel less motivated in their work. As the store managers saw it, making sure that P1 workers were happy with their schedules was critical for keeping their stores running smoothly. As our participant observer noted during a site visit to one of Weave's flagship stores:

> "While observing the scheduling, I noticed that P1 workers always got their full 40 hours. When I asked her why this was the case—knowing that P1 workers could be given anywhere between 30 and 40 hours by contract—she doubled down: "P1 workers always get 40 hours!"…She emphasized that one of her concerns was to ensure that the "leadership team" [P1 workers] received a stable salary… since cutting hours would have "created tension and reduced satisfaction". I perceived a strong sense of camaraderie between the store manager and her leadership team (P1). Instead, she would have cut hours for P2 and P3 without hesitation—and she even showed me that she had recently done so..." (Field note from a site visit to a retail store, including store manager #RET_011)

As such, in Project 2, domain expertise about workforce scheduling seemed to be *dispersed* across the three occupational groups involved—retail finance managers, district managers, and store managers—

who each had their own views about what *'optimal' scheduling* should look like. Following each of these experts left developers with partial fragments of the expertise they were seeking, which they would have to piece together in ways that would help them narrow down on a specific tool that they should develop.

Task Enactment Heterogeneity. Throughout the data collection process, the developers were also struck by the amount of variation in how workforce scheduling decisions were made across stores, and sometimes even in the same store over time. They saw this heterogeneity as resulting from three main contextual factors. First, Weave's stores varied considerably in size: the smallest stores had as few as two workers on site at any given time, and an equally small roster of available workers to rotate shifts between. In such stores, store managers had significantly less flexibility in making scheduling decisions. By contrast, the largest stores had over seventy workers on their roster: here, workforce scheduling required much more concerted effort, and store managers had more flexibility with scheduling. Second, some stores were often visited by "porters": i.e., commercial traders who purchased discounted items in bulk to resell in other states. Porters typically did not require assistance, and did not need to be actively convinced to make purchases. For stores that were reliably and frequently transacted with porters, store managers did not seem to require as many payroll hours to meet their revenue targets. Finally, some store managers were significantly more experienced than others. Experienced store managers were allowed more leeway in negotiating the number of hours they could schedule, even if this required pushing back some of their PPoS targets. Some even started to play an active role in influencing the hiring and training strategies for their district, which further amplified their autonomy in making scheduling decisions.

This heterogeneity in task enactment mattered because the developers were only able to interview and shadow a small subset of Weave's ~180 store managers. While they were, to some extent, able to tease out a few factors that explained differences in scheduling approaches among the stores they observed, they remained unsure if there were any other important considerations that had not been reflected in their sample. As such, if they were to conceptualize a tool based on the insights they gleaned from their limited interviews and observations, they were unsure if such a tool would be widely adopted by all store managers.

By the end of their observations and interviews, the developers had failed to develop a rich understanding of the domain expertise surrounding the task of workforce scheduling. Since domain expertise surrounding the task of workforce scheduling seemed to be dispersed across multiple occupational groups, as well as the fact that members of the same occupational group (here, store managers) enacted this task differently on the ground, developers remained unsure in their understanding of Weave's workforce scheduling processes despite repeated interactions with different groups of domain experts.

Analytical Summary. As the preceding two subsections have shown, the developers' effectiveness in eliciting domain expertise from domain experts was, once again, shaped by the task and organizational structures within which both groups were embedded.

The effective elicitation of domain expertise in Project 1 may be traced to two interrelated reasons. First, and as illustrated in Table 2 (Rows #8—10), *jurisdictional clarity*—i.e., the fact that allocators had an exclusive say in how the task was performed—entailed that there were convergent views of how allocation decisions should be made. Therefore, allocators were able to speak authoritatively and definitively on this matter in their interactions with developers. Second, as illustrated in Table 2 (Rows #13—14), *homogenous task enactment*—i.e., the fact that allocation decisions were made broadly similarly by allocators across different store contexts and product categories—meant that AI developers were able to quickly achieve saturation over repeated interviews and observations.

In Project 2, however, developers were unable to synthesize the fragments of dispersed expertise they collected from the various domain experts they followed. Mirroring Project 1, this low effectiveness can be similarly traced to two reasons. First, as illustrated in Table 2 (Rows #8—10) the task was characterized by *jurisdictional ambiguity*—i.e., multiple groups of experts, including retail finance managers, district managers, and store managers, each had a say in how workforce scheduling decisions ought to be made. As the developers found, each of these groups had radically divergent interests and views on how optimal scheduling should be done, rooted in their own domain expertise (Table 2, Row #8). As such, by largely focusing their interviews and observations on a particular group of experts (i.e., store managers), AI developers were only able to obtain a partial picture of how scheduling decisions should be made. Second, as illustrated in Table 2 (Rows #13—14), *heterogeneous task enactment*: i.e., the fact that scheduling decisions were highly sensitive to contextual factors such as store size, customer types, and manager tenure, meant that this partial picture was itself even more partial. Developers were unsure about the validity and applicability of their findings even after multiple rounds of interviews and observations.

==== Insert Figure 4 here ====

As Figure 4 illustrates, a crucial difference between the two projects was that in the former, domain expertise was *concentrated* in the hands of one group of domain experts (i.e., allocators), whereas in the latter, domain expertise was *dispersed* across multiple groups of domain experts (i.e., store managers, retail finance, district managers, etc.). Through repeated interactions with allocators, developers were able to obtain a rich and consistent understanding of how allocation decisions (Project 1) were made on the ground and what it might mean to optimize these decisions. On the other hand, repeated interactions with store managers, the retail finance team, and district managers (Project 2) left developers with fragmented, conflicting accounts of how workforce scheduling decisions should be made. Developers had to synthesize these various fragments of expertise from different stakeholders to piece together an understanding of what sort of AI tool they should build—which proved to be an extremely challenging endeavor.

Project Outcomes

In line with the variation in effectiveness of the developers' attempts to elicit domain expertise, the outcomes of the two projects also varied considerably. In Project 1, developers were able to parse the domain expertise they had elicited to narrow down on a specific conceptualization of the AI tool they needed to develop. In turn, they proceeded to build a tool that saw rapid and widespread uptake by the allocators. However, Project 2's tool underwent a much longer development period, during which the developers were unable to reach a consensus. Pressured by their supervisors to quickly wrap up the development process, they made an arbitrary choice and proceeded to develop a tool that had no discernible impact on the performance of store managers. Consequently, Weave's executives halted the internal development process, and instead decided to acquire a third-party solution for workforce scheduling.

Variation in the Developers' Ability to Synthesize Domain Expertise and Narrow Down on a Problem Definition. In each project, after months of shadowing domain experts in order to elicit their expertise, the developers hoped to synthesize what they had learned and, ideally, narrow down on a specific problem that they could address using their technical AI expertise. However, their ability to do so varied considerably between the two projects.

In Project 1, domain expertise was largely concentrated in the hands of one occupational group (i.e., allocators), who were able to clearly and authoritatively communicate this expertise to the developers. Since the allocators' accounts were largely consistent with one another, and also accorded with their work practices that developers had been observing, developers felt confident that they had acquired a rich understanding of how day-to-day allocation decisions were made at Weave. One problem that had come up consistently throughout these interactions was the poor quality of sales forecasts that allocators relied upon to make their decisions. Allocators seemed to routinely disregard existing sales forecasts and instead trust their own intuitions about how much inventory different stores in their purview would need. In turn, this problem became a central point of discussion in internal meetings among developers once they had wrapped up their interactions with allocators. Judging by how often this problem had come up in their interactions, the developers felt confident that building a new sales forecasting tool would significantly improve the ways in which allocators performed their daily work. Moreover, improving the quality of sales forecasts was precisely the kind of prediction problem that their AI expertise was particularly well-suited to solving. They quickly put together a brief proposal outlining the nature of the tool they planned to build, and sought the approval of allocators and their supervisor—who all enthusiastically endorsed the idea.

By contrast, Project 2 remained mired in confusion, as the developers struggled to narrow down on a conceptualization of what tool they should develop. After nearly a year of following various domain experts and deliberating about what they had learned, the developers still had multiple potential options on the table. One option, following the advice of district managers, was to build a tool to improve traffic forecasting for retail stores—i.e., to predict how many people are likely to visit each store each month. District managers were particularly enthusiastic about this idea, to the point where they regularly emailed the developers with specific advice on what sorts of predictions such a tool should be able to make, and how this might help improve workforce scheduling. However, developers remained skeptical: as they saw it, most store managers were actually using a standard template for scheduling work, and so adding an extra piece of information (i.e., forecasted traffic) was unlikely to influence their decision-making process. The second option, following the advice of the retail finance team, was to build a tool to optimize the Payroll Percentage of Sales (PPoS) metric. The retail finance team hoped that developers would help them identify stores with staffing imbalances, such that they could reduce the budget for stores that were overstaffed and increase the budget for stores with low sales and high traffic—which could provide an incentive for store managers to improve their workforce scheduling. Again, the developers were skeptical. Optimizing for the PPoS metric seemed to have things backwards, since sales were dependent on—rather than a predictor of—the number of hours scheduled in any given store. At one internal meeting, the developers also briefly considered incorporating the store managers' idea of maximizing the hours assigned to critical ('P1') employees, to ensure that these key performers were kept motivated and happy. However, this idea was quickly discarded: they felt that no one besides store managers would consider such factors as constituting 'optimal' workforce scheduling, and it would be very difficult to convince the retail finance team and district managers to buy into such a proposal. As such, the developers were left at an impasse, with multiple conflicting options on the table. As one developer colorfully expressed his frustration:

> "*This project is like surfing…You see a wave that's coming out from one [group of domain experts], you're like ah, I can ride this wave to do something*. But we're still waiting for the right wave. So, there was a wave with [the retail finance team] who was like 'Let's experiment with payroll percentage of sales', and we were like, 'ok, let's surf on that wave now.' But then we met [district manager], and he has his own wave… So, we're like 'maybe we can ride this wave.... But, you know, he has a completely different idea with respect to [retail finance manager] ... *and we don't know which wave to ride*!" (Field notes from an internal meeting between developers; Speaker: #DEV_005 emphasis added)

The developers understood that, contrary to Project 1, they were unlikely to pick a solution that would be endorsed by the various groups of domain experts involved. However, as time went on and the pressure to deploy a tool continued to mount, they decided—rather arbitrarily—to follow the proposal of the finance team and focus on optimizing the PPoS metric. In large part, this was because the retail finance team was consistently the most involved in providing guidance and advice to the developers, and as such, appeasing them seemed like the easiest way to get this project across the finish line.

Variation in the Rollout and Uptake of AI tools. In Project 1, buoyed by the endorsement of allocators and their supervisor, the developers proceeded to construct a new AI tool to improve sales forecasts. Carefully perusing the documentation of the sales forecast tool that allocators were already using, they judged that the quality of sales forecasts could be significantly improved if the tool's underlying prediction model (based on a 'weighted moving average' algorithm) was replaced with a more complex Recurrent Neural Network (RNN)-based AI model. Choosing this approach also meant that there would be no changes to the front-end of the tool that allocators had already become used to—they would simply get better sales forecasts on their existing familiar interface. The developers hoped that this would minimize the disruption to allocators' existing work practices and increase the likelihood of allocators incorporating the improved tool into their decision-making.

For the next few months, the developers worked to develop and test their new prediction model against Weave's historical sales data, to ensure that their forecasts were as accurate as possible. By early 2021, their first prototype was ready. This prototype was initially deployed in the summer of 2021 to a small subgroup of allocators to test the impact on their performance. This test was judged to be a success—based on the experimental results the developers collected, it seemed that the quality of allocation decisions had measurably improved. Based on this, Weave's executives decided to roll out the tool to all allocators—which was completed by early 2022. This completed rollout was what the developers were hoping for—marking their first major deployment since their team was founded. Shortly after the rollout, the developers were invited to give a presentation of their success to the CEO and shareholders.[1]

In Project 2, the developers made an arbitrary decision to build an AI tool that could help better enforce the retail finance team's Payroll Percentage of Sales (PPoS) metric as a target for store managers. As before, they started by perusing the documentation of the various tools that store managers were already using to plan and input their workforce scheduling decisions. However, since there was no existing tool for calculating optimal PPoS targets that they could simply tweak, they decided to develop an AI tool from scratch that would classify stores based on their historical PPoS performance. This tool would help retail finance managers adjust the PPoS targets for different stores, thus optimally limiting the maximum weekly expenditure in payroll for each store. By the developers' own admission, this was a problematic solution,

[1] In fact, the AI tool developed through Project 1 was widely adopted and seen as very successful by Weave's senior executives that by the time we concluded our observations (Dec 2023), the number of allocators employed at Weave had approximately halved. Generating allocation decisions based on the now highly-accurate sales predictions seemed to require much less human intervention than the allocators had perhaps imagined, and Weave's executives seemed to have decided that their new tool made it no longer necessary to hire as many allocators.

as it appeared from the site visits that store managers were not evaluating payroll recommendations in the Census model to make their scheduling decisions. The retail finance team, however, seemed very pleased with this new tool, and encouraged its rollout. Similar to Project 1, they rolled out this tool to a small sample of twenty stores. However, this experiment showed no impact on store revenue and costs for the treated stores compared to a control group. In fact, there seemed to be no discernible changes at all in how workforce scheduling decisions were being made—it seemed that store managers were largely ignoring the new tool. Most developers had already anticipated such an outcome, and yet they seemed quite disappointed by this failure. The tone of an internal meeting following this failed experiment was noticeably sullen:

> **Developer_#005:** "Are we really able to signal [to the store managers] that there's a better way of scheduling labor? Do they even trust us?..."
> **Developer_002:** "Don't think about it, man… After this meeting, I'll be going to the liquor store on my block!" (Field note from an internal meeting between developers)

Weave's executives were similarly frustrated with the ineffectiveness of this new tool, despite the immense time and resources that had been expended on its development. In November 2023, a decision was made to scrap the tool and discontinue its internal development. Instead, around the time we concluded our fieldwork, the company's executives had started to explore options to purchase a readymade workforce scheduling tool from a third-party vendor.

Discussion and Implications

Initiatives where professionals (such as sustainability officers, technology developers, DEI workers, and legal aid lawyers) are tasked with eliciting domain expertise from other actors (such as procurement managers, supply chain analysts, hiring managers, nurses and physicians) within their organization to execute strategic projects are becoming increasingly common. However, such professionals typically do not have formal authority over these domain experts, which significantly impacts their effectiveness in eliciting domain expertise, and can even contribute to the failure of such strategic projects. Despite the prevalence of this problem, it is unclear when and how professionals are able (or unable) to elicit expertise from domain experts over whom they have no formal authority. Drawing on four years of qualitative fieldwork at Weave and focusing on the case of Weave's AI developers' interactions with domain experts (e.g., allocators, retail finance managers, store managers, etc.), we find that the interplay between task and organizational structures shape the effectiveness of professionals' attempts to elicit domain expertise in the absence of formal authority.

A Model of Eliciting Domain Expertise in the Absence of Formal Authority

Below, we synthesize our findings to develop a model (refer to Figure 5) on eliciting domain expertise in the absence of formal authority, and its consequences.

==== Insert Figure 5 here ====

As illustrated in Figure 5, the interplay between task and organizational structure—as instantiated in lines of authority, professional jurisdictions, and the task composition of domain experts' work— enables (or constrains) the effectiveness of professionals in eliciting domain expertise, impacting the ways in which they are able to identify and gain access to domain experts, as well as elicit their expertise. In particular, the interplay between task and structure shapes (a) the extent to which domain experts are *legible* (versus *illegible*) to their professional collaborators, and (b) the extent to which domain expertise is *concentrated* in the hands of domain experts (versus *dispersed* across multiple experts).

When attempting to elicit domain expertise in the absence of formal authority, professionals first need to identify and then gain access to domain experts who have this expertise. In cases where task and organizational structures render these experts legible (versus illegible), professionals find it easier (versus more difficult) to evaluate which specific experts they need to reach out to and connect with. Once they have gained access to their target domain experts, professionals can move to eliciting domain expertise. When expertise is concentrated in the hands of domain experts (e.g., in our empirical setting, allocators in Project 1), professionals are likely to be more effective in building a robust understanding of domain expertise by interacting with and observing these experts in action. However, when expertise is dispersed across multiple groups of domain experts with varied interests and goals (e.g., in our empirical setting, retail finance managers, store managers, and district managers in Project 2), interactions with each group of domain experts only provide a partial—and at times, even conflicting—account of domain expertise, which professionals then need to reconcile and synthesize together into a complete whole. However, such reconciliation and synthesis of dispersed expertise is difficult to accomplish in practice.

Our findings reveal three underlying factors that shape the effectiveness of professionals' attempts to elicit domain expertise in the absence of formal authority. As depicted in the top half of Figure 5, in situations characterized by (a) *jurisdictional clarity*—i.e., when domain experts have exclusive control over the turf of a certain task (cf. Chown, 2020) about which expertise is being solicited, and are embedded in clear lines of authority, (b) *task centrality*—i.e., when the focal task is central to the work of domain experts, and (c) *task enactment homogeneity*—i.e., when the focal task is enacted similarly among different domain experts, professionals would be more effective in accessing domain experts and eliciting their expertise. This leads to both epistemic and practical consequences—professionals are able to develop a rich understanding of the domain expertise they seek, and, in turn, incorporate this expertise in their pursuit of practical objectives. As we discussed in our empirical case of AI developers and domain experts, these three factors enabled AI developers to acquire a deep understanding of supply chain allocation processes in Project 1 despite not having any formal authority over the relevant domain experts (i.e., allocators), and in turn, to build a tool that usefully improved the work performance of these domain experts.

Conversely, as depicted in the bottom half of Figure 5, in situations characterized by (a) *jurisdictional ambiguity*—i.e., when different groups of domain experts seek to control the turf of the focal task, and are embedded in complex, overlapping lines of authority, (b) *task peripherality*—i.e., when the focal task is less central to the everyday work of domain experts, and (c) *task enactment heterogeneity*—i.e., when different groups of domain experts enact the focal task differently on the ground, then professionals would be less effective in accessing domain experts and eliciting their expertise. This leads to adverse epistemic and practical consequences—professionals are unable to develop a rich understanding of domain expertise, and, in turn, are hindered in their pursuit of objectives for which this expertise is necessary. In our empirical case, these three factors hindered AI developers in that they were unable to acquire a deep understanding of workforce scheduling processes, and thus, were also unable to build a useful tool to improve the work of relevant domain experts.

Together, our model explains how and when professionals, in the absence of formal authority, are able to identify and gain access to domain experts, and effectively elicit their expertise. Below, we discuss our contributions to the literature on cross-occupational collaboration for eliciting domain expertise.

Contributions to the Literature on Cross-Occupational Collaboration for Eliciting Domain Expertise

Beyond our main contribution—i.e., a model of eliciting domain expertise in the absence of formal authority (Figure 5)—our study makes other contributions that advance our understanding of cross-occupational collaboration for eliciting domain expertise. Although prior research has not directly examined attempts by professionals to elicit domain expertise in the absence of formal authority, there has been related prior work to improve our understanding of both cross-occupational collaboration across social hierarchies, as well as the elicitation of expertise in the specific context of our empirical case (i.e., technology development). Our theorizing bridges across, and makes contributions to, both these literatures.

First, our findings highlight the importance of *structural* factors—especially the interplay between task and organizational structure—in shaping the effectiveness of cross-occupational collaboration (Monteiro, 2024). Prior literature has identified and examined *demographic* and *relational* factors that impact the effectiveness of cross-occupational collaboration. For instance, studies have identified the importance of demographic factors, such as cross-cutting demographics between occupational groups (DiBenigno & Kellogg, 2014), in shaping the effectiveness of cross-occupational collaboration. Other studies have examined relational tactics such as "upward and lateral influence" (Kipnis & Schmidt, 1988; Yukl & Falbe, 1990; Dutton & Ashford, 1993; Kellogg, 2019), "scut work" (Huising, 2015) and "peer publicizing" (Karunakaran, 2022) in impacting the success of cross-occupational collaboration. While underscoring the importance of demographic and relational factors, our study highlights the overlooked role of structural factors in shaping the effectiveness of cross-occupational collaboration for eliciting domain expertise, especially when the professionals tasked with eliciting domain expertise do not have formal authority over the very domain experts they need to interact with. For instance, both of the projects we tracked involved the same set of professionals (i.e., the same AI developers) interacting with demographically homogeneous groups of domain experts. Across the two projects, the AI developers, given their lack of formal authority over the domain experts, also enacted a consistent set of relational tactics to upwardly influence the managers of domain experts as well as laterally influence the domain experts. They also enacted a consistent set of "best practices" that they learned through the practitioner literature on "user research" and "requirements elicitation" that involved "empathizing with," and "observing, following, and learning from" domain experts. However, the effectiveness of cross-occupational collaboration for eliciting domain expertise as well as the respective outcomes of these two projects varied significantly. Our study thus highlights the role of structural factors in the form of the interplay of task and organizational structures in rendering the illegibility (versus legibility) of domain experts, and the dispersion (versus concentration) of domain expertise across multiple groups of experts with varied interests and goals (especially given their respective embeddedness in different organizational units) made it difficult (versus easy) for developers to identify, gain access to, and learn from domain experts, and elicit their expertise. In this regard, our study answers recent calls that urge scholars to theorize "the role of *organizational structures* in shaping expertise, rather than *simply housing it*" (Monteiro, 2024, p. 1, emphasis added).

Second, prior work on the importance of seeking domain expertise in technology development processes has largely assumed that domain expertise is concentrated in the "minds and hands" of relevant domain experts. As such, existing literature has largely focused on the relational dynamics between developers and domain experts—assuming that as long as tensions between the two groups are resolved (e.g., when the two groups develop shared representations and common ground), developers will be able to effectively capture domain expertise by 'following' domain experts. Our findings, however, complicate this notion. Merely following domain experts to develop a grounded understanding of organizational workflows and practices can, under some conditions, lead to ineffective technology development processes, in part because structural factors could hinder the evolution of mutual learning processes (van den Broek et al., 2021) between AI developers and domain experts that are needed to build grounded and useful AI tools. In cases where the interplay between task and organizational structures render domain experts illegible and domain expertise dispersed—especially when multiple groups of domain experts with conflicting interests and goals are all trying to simultaneously shape the AI development process—cycles of mutual learning where AI developers and domain experts engage "deeply with the technology" (van den Broek et al., 2021, p. 1557) and with each other to come up with a human-AI "hybrid practice" (p.1573) are constrained. In such cases, "following the domain experts" is unlikely to be sufficient for eliciting domain expertise. In this vein, our findings prompt an epistemic shift from a *substantive* view of expertise as something "possessed" by certain domain experts, towards a view of expertise as *emergent* from a network of domain experts embedded in a web of task and organizational structures (cf. Anteby & Holm, 2021; Pakarinen & Huising, 2023). By building on prior works (e.g., Eyal, 2013; Collins & Evans, 2002) and conceptually distinguishing experts from expertise, we can start to appreciate the ways in which expertise might be, under some conditions, concentrated in the "minds and hands" of a group of experts, and at other conditions, dispersed across a broad range of actors.

Beyond just a conceptual distinction, this distinction between experts and expertise has important implications for technology development. When expertise is concentrated in the hands of a single occupational group of domain experts, developers can, as expected, effectively get access to those experts and elicit their expertise. However, when domain expertise is dispersed across a wide range of occupational groups, developers would only collect fragments of expertise from their interactions with the different occupational groups involved, and would then need to reconcile conflicting accounts and synthesize "bits and pieces" of information together in order to develop a robust understanding of relevant domain expertise. This poses a significant hindrance to eliciting domain expertise, especially when the developers do not have formal authority over domain experts. Therefore, an important implication of our findings is that organizations either need to ascribe some formal authority to the technology developers and institute processes (for e.g., maximum turnaround time for domain experts when they receive a meeting request from developers) that create enabling conditions for domain experts to start considering requests from developers more seriously, or create boundary spanning roles (Levina & Vaast, 2005) such as project/product managers and business analysts who have some formal authority over the domain experts.

Third, while prior research suggests that domain experts will be reluctant to cooperate with technology developers and share their expertise when they view the proposed technology as a "threat" to the core tasks they perform, for fear of undermining their occupational identity and loss of power and status (Forsythe, 1993; Anthony, 2018; Nelson & Irwin, 2014), our findings suggest a contrasting picture. As our findings indicate, the allocators in Project 1 were willing to cooperate with the developers even though the proposed AI tool was supposed to augment their "core" task of product allocation, while the store managers in Project 2 were reluctant to cooperate with the developers despite the fact that the proposed AI tool was supposed to augment what they considered as a "peripheral" task. One possible explanation for these seemingly disparate findings is that domain experts are willing to cooperate with technology developers *in the short run* insofar as they view technology development efforts as having the potential to significantly augment the core tasks of their jobs *in the long run*. In turn, the potential to shape long-run improvements in their core tasks makes them more (and not less) invested in the process, and more motivated to ensure that the technology developers have as much information as possible to build a useful tool. In contrast, when domain experts view the proposed technology as augmenting tasks they consider "peripheral" to their everyday work, then they are less invested in the process, and less motivated to cooperate with the developers as they may find it a "waste" of their time in the short run, even though the proposed technology might produce long-term benefits precisely because they augment peripheral tasks. More generally, our findings suggest an alternative explanation of core-task augmentation (versus peripheral-task augmentation) in shaping the proclivity of domain experts to cooperate with technology developers.

Finally, prior research on the elicitation of domain expertise for technology development has largely treated the organizational context as a background condition—or a "container"—within which collaborative efforts for technology development take place (see Monteiro, 2024 for related arguments). However, as our findings show, organizations serve as more than mere containers, and processes of technology development are deeply shaped by a complex web of task and organizational structures within which technology developers and domain experts are embedded. By paying closer attention to such structural factors, we might move toward a deeper understanding of how and when technology development processes, as they unfold in organizational contexts, are more or less likely to be effective.

Our findings also have important practical implications regarding the development and use of AI systems in large organizations (Wiesenfeld et al., 2022). While organizations have been increasingly investing in the development of AI tools for improving various organizational processes, returns on these investments have been unexpectedly slow (Benanav, 2020; Brynjolfsson et al., 2019; McElheran et al., 2023). Explanations for this discrepancy have tended to focus on the lack of complementary organizational capabilities (e.g., internal workflows, processes, and roles) that could leverage the benefits of these tools, and/or suboptimal conditions for employees on the ground to incorporate these tools into their daily work practices (cf. Lebovitz et al., 2021; Anthony et al., 2023). For instance, prior research has suggested that attempts to implement new technology in organizations may stall because employees might perceive an existential threat to their jobs or that they might lose power and status when the tool is deployed within the organization (Granulo et al., 2019; Longoni et al., 2019; Markus, 1983). Prior research has also shown that employees may be disproportionately averse to accepting AI recommendations, or may not trust the outputs of opaque 'black-box' AI systems (Dietvorst et al., 2015; Glikson & Woolley, 2020). Our findings complicate this picture in two key ways. First, building on recent related calls in organizational scholarship (Bailey & Barley, 2020; Sergeeva, 2023; van den Broek et al., 2021), we pay closer attention to the early stages of designing intelligent technologies—focusing especially on the role of the designers and developers of these technologies—and show how issues emerging *upstream* in AI development may have *downstream* consequences for their subsequent deployment and use. Second, our findings also underscore the need to move beyond a focus on cognitive factors—i.e., the extent to which different stakeholders might or might not adopt the right "frames" of reference to effectively develop and use AI systems—toward a consideration of the structural (and relational) factors of authority and expertise that shape work practices surrounding the development and use of AI.

At a time when widespread hype about the promised benefits of AI systems seems to be disconnected from their adoption and use in practice, our findings advance our understanding of the role of task and organizational structure—and their interplay—in shaping the contingent impacts of AI on work, occupations, and organizations.

References

Aghion, P., & Tirole, J. (1997). Formal and Real Authority in Organizations. *The Journal of Political Economy*, *105*(1), 1–29.

Almandoz, J., & Tilcsik, A. (2016). When Experts Become Liabilities: Domain Experts on Boards and Organizational Failure. *The Academy of Management Journal*, *59*(4), 1124–1149.

Anteby, M., & Holm, A. L. (2021). Translating Expertise across Work Contexts: U.S. Puppeteers Move from Stage to Screen. *American Sociological Review*, *86*(2), 310–340. https://doi.org/10.1177/0003122420987199

Anthony, C. (2018). To question or accept? How status differences influence responses to new epistemic technologies in knowledge work. *Academy of Management Review*, *43*(4), 661–679.

Anthony, C., Bechky, B. A., & Fayard, A.-L. (2023). "Collaborating" with AI: Taking a System View to Explore the Future of Work. *Organization Science*. https://doi.org/10.1287/orsc.2022.1651

Augustine, G. (2021). We're not like those crazy hippies: The dynamics of jurisdictional drift in externally mandated occupational groups. *Organization Science*, *32*(4), 1056–1078.

Autor, D. (2014). *Polanyi's paradox and the shape of employment growth*. National Bureau of Economic Research.

Bailey, D. E., & Barley, S. R. (2020). Beyond design and use: How scholars should study intelligent technologies. *Information and Organization*, *30*(2), 100286. https://doi.org/10.1016/j.infoandorg.2019.100286

Barley, S. R. (1986). Technology as an occasion for structuring: Evidence from observations of CT scanners and the social order of radiology departments. *Administrative Science Quarterly*, 78–108.

Basbug, G., Cavicchi, A., & Silbey, S. S. (2023). Rank has its privileges: Explaining why laboratory safety is a persistent challenge. *Journal of Business Ethics*, *184*(3), 571–587.

Beane, M., & Anthony, C. (2023). Inverted Apprenticeship: How Senior Occupational Members Develop Practical Expertise and Preserve Their Position When New Technologies Arrive. *Organization Science*.

Benanav, A. (2020). *Automation and the Future of Work*. Verso Books.

Blau, P. M. (1968). The hierarchy of authority in organizations. *American Journal of Sociology*, *73*(4), 453–467.

Boland, R. J., & Tenkasi, R. V. (1995). Perspective making and perspective taking in communities of knowing. *Organization Science*, *6*(4), 350–372.

Bourgoin, A., Bencherki, N., & Faraj, S. (2020). "And Who Are You?": A Performative Perspective on Authority in Organizations. *Academy of Management Journal*, *63*(4), 1134–1165.

Brooks, F. P., & Bullet, N. S. (1987). Essence and accidents of software engineering. *IEEE Computer*, *20*(4), 10–19.

Browne, G. J., & Rogich, M. B. (2001). An empirical investigation of user requirements elicitation: Comparing the effectiveness of prompting techniques. *Journal of Management Information Systems*, *17*(4), 223–249.

Bruns, H. C. (2013). Working Alone Together: Coordination in Collaboration Across Domains of Expertise. *The Academy of Management Journal*, *56*(1), 62–83.

Brynjolfsson, E., Rock, D., & Syverson, C. (2019). Artificial intelligence and the modern productivity paradox. *The Economics of Artificial Intelligence: An Agenda*, *23*, 23–57.

Byrd, T. A., Cossick, K. L., & Zmud, R. W. (1992). A Synthesis of Research on Requirements Analysis and Knowledge Acquisition Techniques. *MIS Quarterly*, *16*(1), 117. https://doi.org/10.2307/249704

Chakraborty, S., Sarker, S., & Sarker, S. (2010). An Exploration into the Process of Requirements Elicitation: A Grounded Approach. *Journal of the Association for Information Systems*, *11*(4).

Chan, C. K., & Hedden, L. N. (2023). The Role of Discernment and Modulation in Enacting Occupational Values: How Career Advising Professionals Navigate Tensions with Clients. *Academy of Management Journal*, *66*(1), 276–305. https://doi.org/10.5465/amj.2020.1014

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. sage.

Chown, J. (2020). Financial Incentives and Professionals' Work Tasks: The Moderating Effects of Jurisdictional Dominance and Prominence. *Organization Science*, *31*(4), 887–908.

Cohen, L. E. (2013). Assembling Jobs: A Model of How Tasks Are Bundled Into and Across Jobs. *Organization Science*, *24*(2), 432–454. https://doi.org/10.1287/orsc.1110.0737

Collins, H., & Evans, R. (2002). The third wave of science studies: Studies of expertise and experience. *Social Studies of Science*, *32*(2), 235–296.

DiBenigno, J. (2018). Anchored Personalization in Managing Goal Conflict between Professional Groups: The Case of U.S. Army Mental Health Care. *Administrative Science Quarterly*, *63*(3), 526–569.

DiBenigno, J., & Kellogg, K. C. (2014). Beyond Occupational Differences: The Importance of Cross-cutting Demographics and Dyadic Toolkits for Collaboration in a U.S. Hospital. *Administrative Science Quarterly*, *59*(3), 375–408. https://doi.org/10.1177/0001839214538262

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114.

Dutton, J. E., & Ashford, S. J. (1993). Selling issues to top management. *Academy of Management Review*, *18*(3).

Evans, J. (2021). How Professionals Construct Moral Authority: Expanding Boundaries of Expert Authority in Stem Cell Science. *Administrative Science Quarterly*, *66*(4), 989–1036.

Eyal, G. (2013). For a Sociology of Expertise: The Social Origins of the Autism Epidemic. *American Journal of Sociology*, *118*(4), 863–907. https://doi.org/10.1086/668448

Feldberg, A. C. (2022). The task bind: Explaining gender differences in managerial tasks and performance. *Administrative Science Quarterly*, 67(4), 1049-1092.

Forsythe, D. E. (1993). Engineering Knowledge: The Construction of Knowledge in Artificial Intelligence. *Social Studies of Science*, *23*(3), 445–477.

Glaser, B., & Strauss, A. (2017). *Discovery of grounded theory: Strategies for qualitative research*. Routledge.

Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Goguen, J. A., & Linde, C. (1993). *Techniques for requirements elicitation*. 152–164.

Granulo, A., Fuchs, C., & Puntoni, S. (2019). Psychological reactions to human versus robotic job replacement. *Nature Human Behaviour*, *3*(10), 1062–1069.

Hassenzahl, M., & Tractinsky, N. (2006). User experience—A research agenda. *Behaviour & Information Technology*. https://doi.org/10.1080/01449290500330331

Heimstädt, M., Koljonen, T., & Elmholdt, K. T. (2024). Expertise in Management Research: A Review and Agenda for Future Research. *Academy of Management Annals*, *18*(1), 121–156.

Hickey, A., & Davis, alan. (2004). A Unified Model of Requirements Elicitation. *Journal of Management Information Systems*, *20*(4), 65–84. https://doi.org/10.1080/07421222.2004.11045786

Huising, R. (2015). To Hive or to Hold? Producing Professional Authority through Scut Work. *Administrative Science Quarterly*, *60*(2), 263–299. https://doi.org/10.1177/0001839214560743

Karunakaran, A. (2022). Status–Authority Asymmetry between Professions: The Case of 911 Dispatchers and Police Officers. *Administrative Science Quarterly*, *67*(2), 423–468. https://doi.org/10.1177/00018392211059505

Karunakaran, A., & Etzion, D. (2024). *Mandate without Authority? Sustainability Professionals in Large Organizations.* Stanford University Working Paper.

Kellogg, K. C. (2014). Brokerage Professions and Implementing Reform in an Age of Experts. *American Sociological Review*, *79*(5), 912–941. https://doi.org/10.1177/0003122414544734

Kellogg, K. C. (2019). Subordinate activation tactics: Semi-professionals and micro-level institutional change in professional organizations. *Administrative Science Quarterly*, *64*(4), 928–975.

Kipnis, D., & Schmidt, S. M. (1988). Upward-influence styles: Relationship with performance evaluations, salary, and stress. *Administrative Science Quarterly*, 528–542.

Koljonen, T., & Chan, C. K. (2024). Balancing Professional Autonomy and Managerial Goals amid Broad Technology Adoption Pressures: Intraprofessional Segmentation at a Finnish School. *Academy of Management Journal*, *67*(3), 798–828. https://doi.org/10.5465/amj.2022.1093

Koppman, S., Bechky, B. A., & Cohen, A. C. (2022). Overcoming conflict between symmetric occupations: How “creatives” and “suits” use gender ordering in advertising. *Academy of Management Journal*, *65*(5).

Lebovitz, S., Levina, N., New York University, Lifshitz-Assa, H., & New York University. (2021). Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts’ Know-What. *MIS Quarterly*, *45*(3), 1501–1526. https://doi.org/10.25300/MISQ/2021/16564

Levina, N., & Vaast, E. (2005). The emergence of boundary spanning competence in practice: Implications for implementation and use of information systems. *MIS Quarterly*, 335–363.

Levine, T. (2023). *Council Post: Unlocking The Power Of Industry Domain Expertise In A Technology Team*. Forbes. https://www.forbes.com/councils/forbesbusinesscouncil/2023/10/25/unlocking-the-power-of-industry-domain-expertise-in-a-technology-team/

Lifshitz-Assaf, H. (2018). Dismantling knowledge boundaries at NASA: The critical role of professional identity in open innovation. *Administrative Science Quarterly*, *63*(4), 746–782.

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, *46*(4), 629–650.

Mao, Y., Wang, D., Muller, M., Varshney, K. R., Baldini, I., Dugan, C., & Mojsilović, A. (2019). How Data Scientists Work Together With Domain Experts in Scientific Collaborations: To Find The Right Answer Or To Ask The Right Question? *Proceedings of the ACM on Human-Computer Interaction*, *3*, 237: 1-237:23.

Margolin, V. (1997). Getting to know the user. *Design Studies*, *18*(3), 227–236.
Markus, M. L. (1983). Power, politics, and MIS implementation. *Communications of the ACM*, *26*(6), 430–444.
McElheran, K., Li, J. F., Brynjolfsson, E., Kroff, Z., Dinlersoz, E., Foster, L. S., & Zolas, N. (2023). *AI Adoption in America: Who, What, and Where* (Working Paper No. 31788). National Bureau of Economic Research.
Merton, R. K. (1987). Three fragments from a sociologist's notebooks: Establishing the phenomenon, specified ignorance, and strategic research materials. *Annual Review of Sociology*, *13*(1), 1–29.
Monteiro, P. (2024). Generating, Grading, and Ghosting: How Organizing Experts Shapes Expertise. *Journal of Management Studies*. https://doi.org/10.1111/joms.13056
Nelson, A. J., & Irwin, J. (2014). "Defining What We Do—All Over Again": Occupational Identity, Technological Change, and the Librarian/Internet-Search Relationship. *Academy of Management Journal*, *57*(3), 892–928.
Nielsen, J. A., Elmholdt, K. T., & Noesgaard, M. S. (2024). Leading digital transformation: A narrative perspective. *Public Administration Review*, 84(4), 589-603.
Pakarinen, P., & Huising, R. (2023). Relational Expertise: What Machines Can't Know. *Journal of Management Studies*, *n/a*(n/a). https://doi.org/10.1111/joms.12915
Palumbo, S., & Edelman, D. (2023, July 1). What Smart Companies Know About Integrating AI. *Harvard Business Review*. https://hbr.org/2023/07/what-smart-companies-know-about-integrating-ai
Piorkowski, D., Park, S., Wang, A. Y., Wang, D., Muller, M., & Portnoy, F. (2021). How AI Developers Overcome Communication Challenges in a Multidisciplinary Team: A Case Study. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 131:1-131:25. https://doi.org/10.1145/3449205
Polanyi, M. (1966). The tacit dimension. In *Knowledge in organisations* (pp. 135–146). Routledge.
Pugh, D. S., Hickson, D. J., Hinings, C. R., & Turner, C. (1968). Dimensions of organization structure. *Administrative Science Quarterly*, 65–105.
Sandhu, S., & Kulik, C. T. (2019). Shaping and being shaped: How organizational structure and managerial discretion co-evolve in new managerial roles. *Administrative Science Quarterly*, *64*(3), 619–658.
Scott, W. R. (1975). Organizational Structure. *Annual Review of Sociology*, *1*, 1–20.
Sergeeva, A. V. (2023). Why developers matter: The case of patient portals. *Health Informatics Journal*, *29*(1).
Soderstrom, S. B., & Weber, K. (2020). Organizational structure from interaction: Evidence from corporate sustainability efforts. *Administrative Science Quarterly*, *65*(1), 226–271.
Somers, M. (2024, September 11). *The secret to successful AI implementations? Worker voice | MIT Sloan*. https://mitsloan.mit.edu/ideas-made-to-matter/secret-to-successful-ai-implementations-worker-voice
Sosa-Hidalgo, M. A., Hafermalz, E., Günther, W., & Huysman, M. (2024). The Ongoing Quest For Complicatedness: How Data Science Practitioners Manage Their Emerging Role In Organizations. *ECIS 2024 Proceedings*. https://aisel.aisnet.org/ecis2024/track06_humanaicollab/track06_humanaicollab/12
Stice-Lusvardi, R., Hinds, P. J., & Valentine, M. (2023). Legitimating Illegitimate Practices: How Data Analysts Compromised Their Standards to Promote Quantification. *Organization Science*.
Tiwari, S., Rathore, S. S., & Gupta, A. (2012). Selecting requirement elicitation techniques for software projects. *Proceedings of the CSI Sixth International Conference on Software Engineering (CONSEG)*.
Truelove, E., & Kellogg, K. C. (2016). The Radical Flank Effect and Cross-occupational Collaboration for Technology Development during a Power Shift. *Administrative Science Quarterly*, *61*(4), 662–701.
van den Broek, E., Sergeeva, A., & Huysman, M. (2021). When the Machine Meets the Expert: An Ethnography of Developing Ai for Hiring. *MIS Quarterly*, *45*(3), 1557–1580. https://doi.org/10.25300/MISQ/2021/16559
Weeks, K. P., Taylor, N., Hall, A. V., Bell, M. P., Nottingham, A., & Evans, L. (2024). "They Say They Support Diversity Initiatives, But They Don't Demonstrate It": The Impact of DEI Paradigms on the Emotional Labor of HR&DEI Professionals. Journal of Business and Psychology, 39(2), 411-433.
Wiesenfeld, B. M., Aphinyanaphongs, Y., & Nov, O. (2022). AI model transferability in healthcare: a sociotechnical perspective. *Nature Machine Intelligence*, 4(10), 807-809.
Wilmers, N. (2020). Job Turf or Variety: Task Structure as a Source of Organizational Inequality. *Administrative Science Quarterly*, *65*(4), 1018–1057. https://doi.org/10.1177/0001839220909101
Wilson, T. D. (1981). On user studies and information needs. *Journal of Documentation*, *37*(1), 3–15.
Wrong, D. (1979). *Power: Its forms, bases and uses*. Routledge.
Yukl, G., & Falbe, C. M. (1990). Influence tactics and objectives in upward, downward, and lateral influence attempts. *Journal of Applied Psychology*, *75*(2), 132.
Zowghi, D., & Coulin, C. (2005). Requirements elicitation: A survey of techniques, approaches, and tools. *Engineering and Managing Software Requirements*, 19–46.

TABLES AND FIGURES

Table 1: Overview of Data Collection

<table>
<tr><th></th><th>Data</th><th>Usage in Analysis</th></tr>
<tr><td rowspan="2">Observations</td><td>Project 1
Approximately 300 hours between May 2020—May 2022:
Weekly meetings between developers,
Developers’ meetings with key stakeholders in this project (allocators, Head of Supply Chain, etc.),
Monthly meetings with developers and the Chief of Data Science,
Informal company events and social gatherings
3 presentations made by the developers to Weave’s C-suite</td><td rowspan="2">To track how developers’ understanding of the domain expertise they were seeking shifted over time.

To document interactions between developers and different domain experts and stakeholders involved in the two projects.

To understand how developers evaluated the quality of their interactions with domain experts, and what they felt about the different domain experts they were interacting with.

To document how developers were evaluating the overall status of the two projects over time, and their decisions to move each project forward.

To understand what developers felt about the two projects, and how they were evaluating their prospects of successfully completing each project.</td></tr>
<tr><td>Project 2
Approximately 200 hours between July 2022—Dec 2023:
Weekly meetings between developers,
Developers’ meetings with key stakeholders in this project (retail finance managers, district managers, store managers, etc.)
Monthly meetings with developers and the Chief of Data Science,
Informal company events and social gatherings
5 store visits made by developers.</td></tr>
<tr><td rowspan="2">Interviews</td><td>Project 1
16 semi-structured interviews:
15 Interviews with allocators and their supervisor,
1 Interview with the Head of Supply Chain</td><td rowspan="2">To independently track the information that domain experts were communicating to developers.

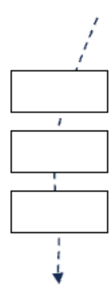
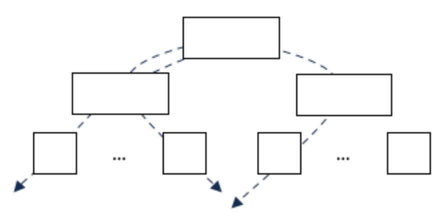
To understand how domain experts evaluated their interactions with developers.

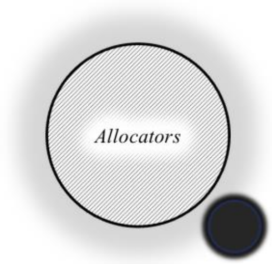
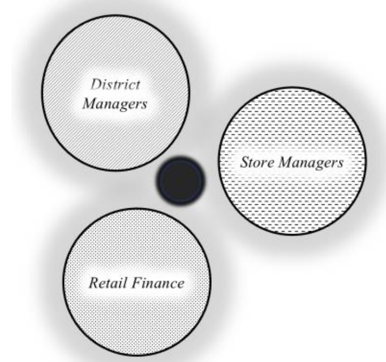
To understand how domain experts perceived the importance and/or stakes of the developers’ projects, and what outcomes they were hoping to see from these</td></tr>
<tr><td>Project 2
15 semi-structured interviews:
5 Interviews with Store Managers
2 Interviews with District Manager
1 Interview with Regional Manager
7 Interviews with Retail Finance Managers</td></tr>
<tr><td rowspan="2">Archival Documents</td><td>Project 1
Memos of task and process maps created with AI developers during their exploration, describing:
Supply chain processes, tasks, and tools used.
Stakeholders involved and their relationships.
Documentation on allocators’ decision-making tools.
Mail exchanges between AI developers and stakeholders.</td><td rowspan="2">To evaluate how the developers’ solutions differed (or not) from the tools that domain experts were already using.

To track how developers were documenting (or not) their evolving understanding of the two processes they were trying to learn about.

To track written interactions between developers and different domain experts and stakeholders involved in the two projects.</td></tr>
<tr><td>Project 2
Documentation of the tools employed for communication between the store managers, the district managers and the company.
Mail exchanges between AI developers and stakeholders.</td></tr>
</table>

Table 2: The Interplay Between Task and Organizational Structures and their Impact on the Elicitation of Domain Expertise

#	Dimensions	Project 1	Project 2
1	Focal task for the AI developers	**Allocation** of products across Weave's stores	Weekly **workforce scheduling**
		JURISDICTIONAL CLARITY/AMBIGUITY	
2	Occupations that had a say in how the focal task should be performed	*Single occupation:* • Allocators	*Multiple occupations:* • Store Managers • Retail Finance Managers • District Managers
3	Target domain experts	10 Allocators	9 Store Managers
4	Presence of a broker with direct authority over all target domain experts	*Yes* The Head of Allocators could broker connections to all 10 allocators, who were all his direct reports.	*No* Different regional and district managers could each broker connections only to a few store managers who directly reported to them.
5	Steps taken by developers to access brokers	*'Start from the top'* Leveraged their connections with Weave's senior executives (in this case, the Head of Supply Chain) to ask to be introduced to brokers.	*'Start from the top'* Leveraged their connections with Weave's senior executives (in this case, the Head of Retail Finance) to ask to be introduced to brokers.
6	Effectiveness of the developers' attempts to access brokers and domain experts	*Effective* Their senior-level contact (the Head of Supply Chain) made introductions to the Head of Allocators, who in turn, made introductions to all 10 allocators.	*Ineffective* Their senior-level contact (the Head of Retail Finance) did not have direct authority over regional and district managers, and were unable to make introductions to them. Similarly, each regional/district manager could only make introductions to the few store managers they had authority over.
7	**Lines of authority**	**Single, clear line of authority**  The developers needed to traverse a single line of authority starting from the Head of Supply Chain, down to the Head of Allocators, to their target domain experts (allocators).	**Multiple, overlapping lines of authority**  The developers needed to traverse multiple, overlapping lines of authority—spanning different groups of retail finance, district, and regional managers—to access their target domain experts (store managers).
8	Views about how the focal task should be performed	*Convergent:*	*Divergent:*

		• Products should be allocated based on the proportion of total sales that a store was forecasted to make.	• According to store managers, workforce scheduling should be based on keeping key (‘P1’) *employees happy and motivated*. • According to retail finance managers, workforce scheduling should *optimize store revenues and costs*. • According to district managers, workforce scheduling should be based on *store traffic forecasts to maximize sales*.
9	**Control over the turf**	**Exclusive control over the turf of focal task**  Allocators had exclusive say over how the focal task (allocation) should be performed.	**Contested control over the turf of focal tasks**  Multiple occupational groups (store managers, retail finance managers, district managers) had diverging views of how the focal task (workforce scheduling) should be performed.
10	**Jurisdiction**	**Jurisdictional Clarity** Weave’s organizational structure surrounding the task of allocation was such that (a) a single occupation had control over the turf of this task, and (b) members of this occupation were embedded in a single, clear line of authority	**Jurisdictional Ambiguity** Weave’s organizational structure surrounding the task of workforce scheduling was such that (a) multiple occupations contested for control over the turf of this task, and (b) members of these occupations were embedded in complex, overlapping lines of authority.
		TASK CENTRALITY/PERIPHERALITY	
11	Task-Composition of the Focal Occupation **Focal task for the AI developers*	Allocators’ tasks, in decreasing order of priority: • **Allocating products from the Fulfillment Center to Stores* • Checking Inventory status across stores • Solving misallocation problems • Preparing and sending weekly updates to the Head of Allocators	Store managers’ tasks, in decreasing order of priority: • Handling customer complaints • Managing daily store operations (e.g., morning stand-up meetings, opening/closing registers) • Preparing and analyzing financial reports for district managers • Analyzing sales data to identify trends and areas for improvement • Visual merchandising (e.g., maintaining store visual standards and displays, etc.) • Implementing sales strategies and promotions.

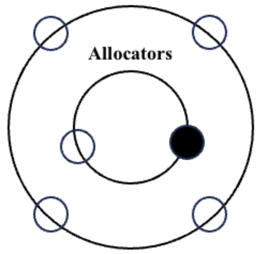
			• Hiring, onboarding and training employees • *<u>*Weekly workforce scheduling</u>* • Conducting regular inventory audits • Building relationships with local businesses (e.g., shopping centers) and participating in community events • Managing contracts with vendors
12	**Task Centrality/Peripherality**	**Task Centrality** The task that the AI developers were working to augment (**allocation**) was central to the job of allocators.  	**Task Peripherality** The task that the AI developers were working to augment (**workforce scheduling**) was peripheral to the job of store managers. 
		TASK ENACTMENT HOMOGENEITY/HETEROGENEITY	
13	Contextual factors that impacted task enactment	*None/minimal:* • The process of making allocation decisions was largely similar across all 10 allocators.	*Significant:* • Large stores had more flexibility to optimize staffing compared to small stores where staffing was fixed. • Store managers with control over hiring/firing decisions had more leeway to control costs. • Stores with a higher percentage of "porters" had less need for sales staff to convert traffic into sales.
14	**Task Enactment Homogeneity/ Heterogeneity**	**Homogeneous task enactment**  Different allocators enacted the focal task (allocation) in a largely similar manner, regardless of the types of products they were allocating.	**Heterogeneous task enactment**  Different store managers enacted the focal task (workforce scheduling) in different ways depending on contextual factors such as store size, manager tenure, and customer profiles.

Figure 1: Reproduction of a process map constructed by developers following their meeting with the Head of Supply Chain for Project 1 (emphasis theirs)

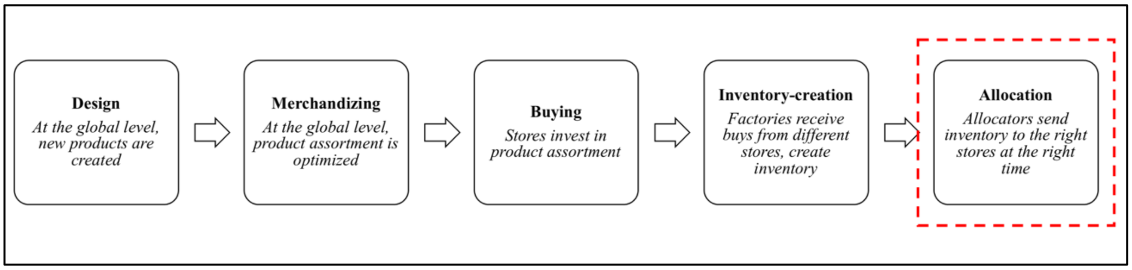


Figure 2: A Single, Clear Line of Authority in Project 1 (Supply Chain Allocation Project)

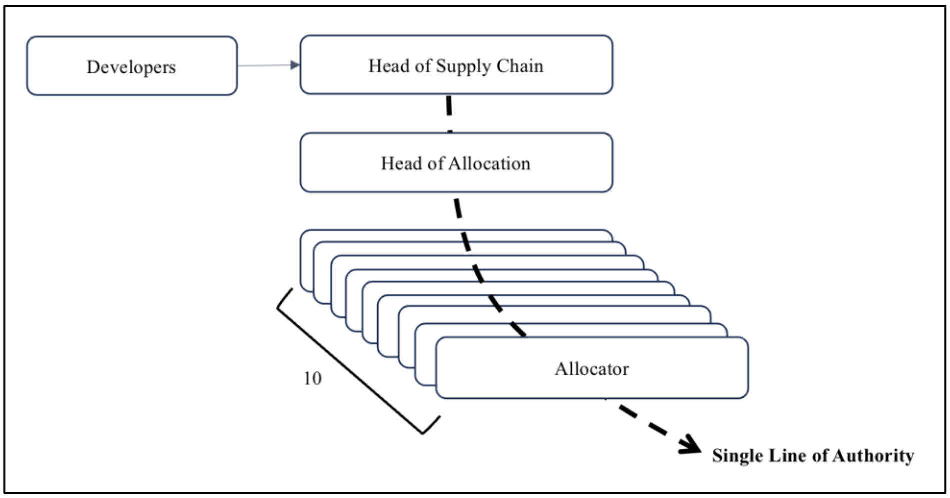


Figure 3: Multiple, Overlapping Lines of Authority in Project 2 (Workforce Scheduling Project)

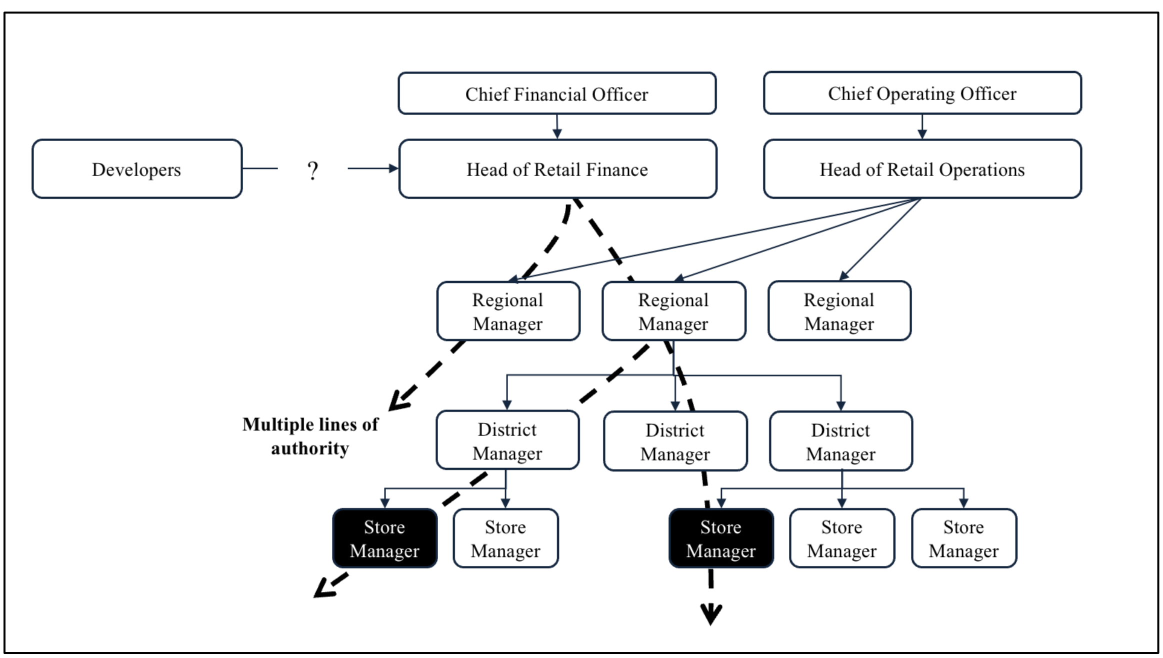


Figure 4: Concentration versus Diffusion of Domain Expertise Across the Two Projects

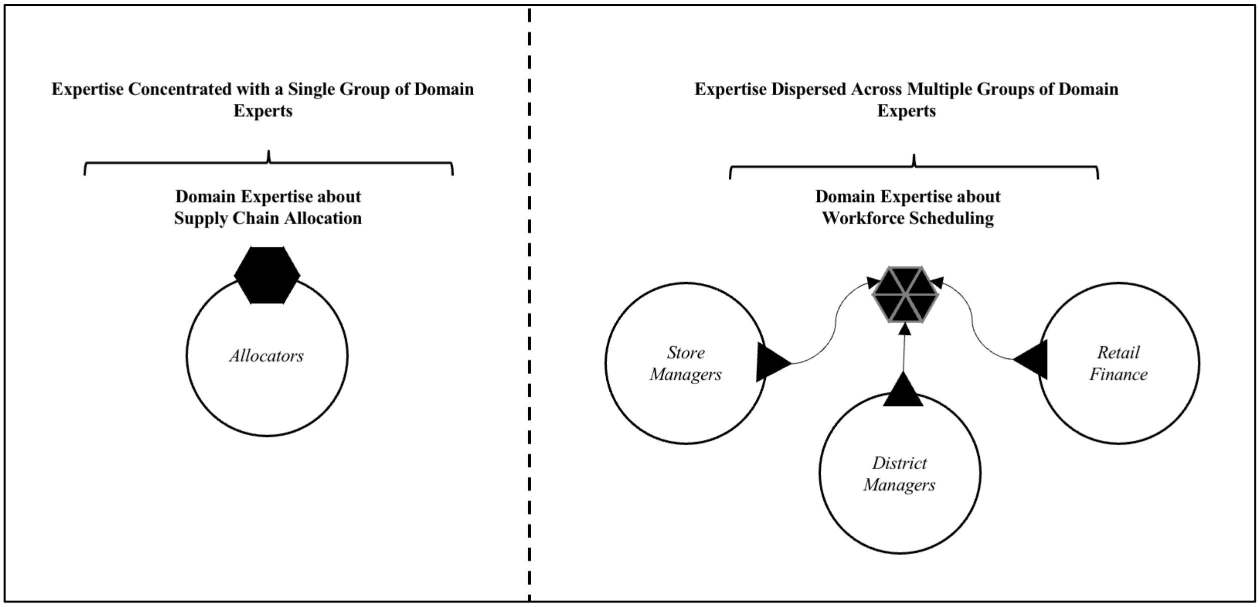
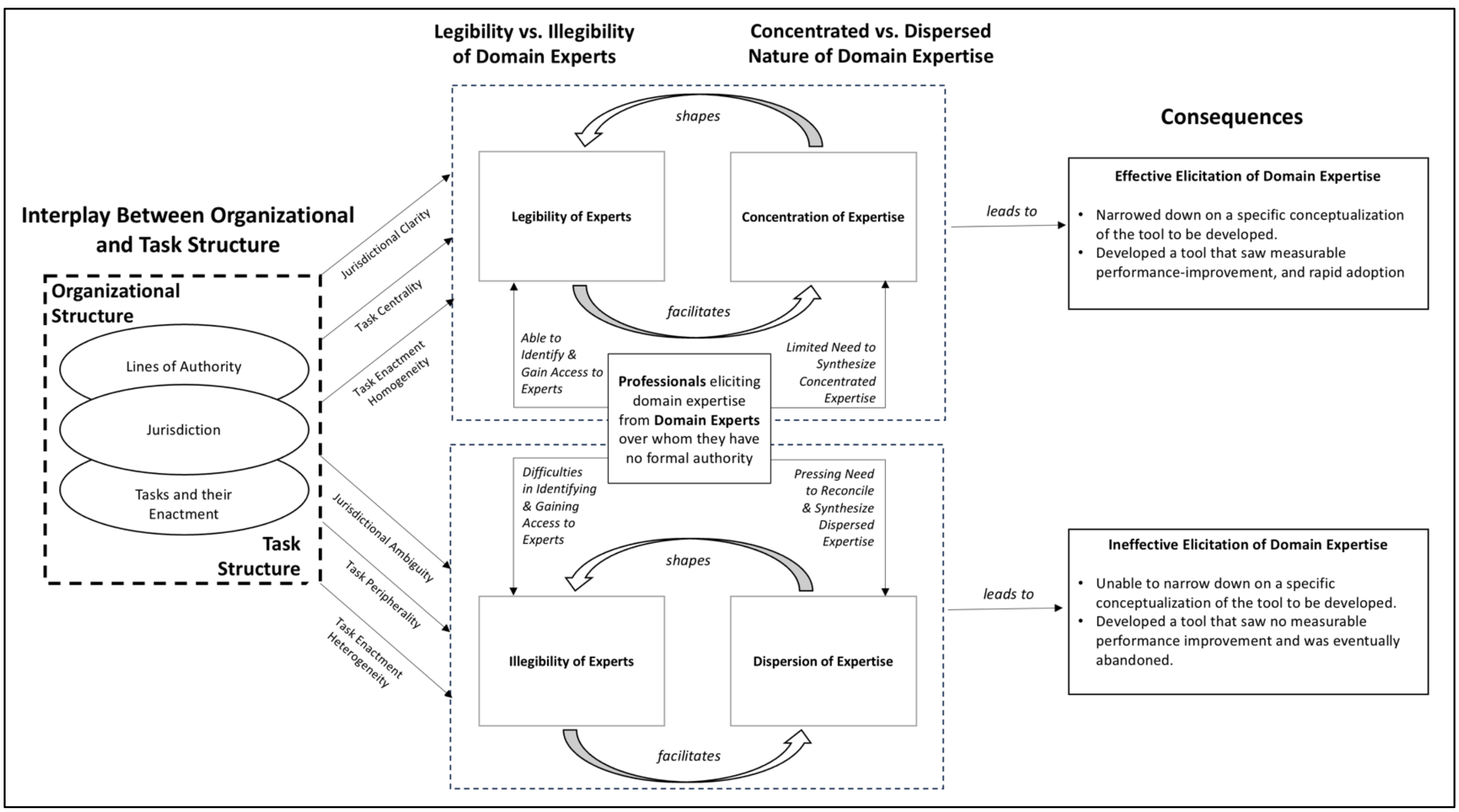


Figure 5: A Model of Eliciting Domain Expertise in the Absence of Formal Authority



Appendix A: Additional Evidence

Themes and Sub-themes	Evidence	Interpretation
Jurisdictional Clarity (Project 1)		
Allocators had exclusive control over the turf of allocation decisions.	*Interviewer:* Why do you over-allocate? *RET_001 (Head of Supply Chain):* Because it is difficult to predict where the products should be shipped. On a detailed level as to *why*, we are not super good at this, I cannot help you with that. I don't even know what software we use. I mean, you should talk to ALL_001. He will know for sure. I can make an introduction". ——— *Interviewer:* How frequently do they [allocators] allocate products? *ALL_001 (Head of Allocators):* "Oh, it's their decision. They know when it's needed. Look, I think you should really talk with them if you need this level of detail. I just coordinate them, and I do a lot of stakeholder interface type work, you know, telling people that we can't do what they're asking us for, and so on…"	When discussing processes of allocation, both the Head of Supply Chain and the Head of Allocators seemed hesitant to get into the technicalities of how this task should be done. They did not seem to have a say in how allocation decisions were made on the ground, and were happy to defer to the expertise of allocators in this matter.
Allocators were embedded in a single line of authority.	From: DEV_002 (Developer) To: ALL_001 (Head of Allocators) Subject: Interviews Dear ALL_001, Thanks again for taking the time to schedule the first set of interviews. This is a really great step in the right direction. In addition to these, it would also be really helpful if we could, over the next two weeks speak with ALL_002 ALL_003 and ALL_004 … Kind regards, DEV_002 ——— From: ALL_001 (Head of Allocators) To: DEV_002 (Developer) CC: ALL_005; ALL_006; ALL_007 (Allocators) Subject: RE: Interviews Hi [DEV_002],	As this email thread illustrates, in the first project, developers were able to access allocators by traversing a single line of authority. The Head of Allocators introduced all allocators to the developers, and directed them to make time for interviews – which they duly complied with.

	… We can work to schedule you to have 15-minute interviews with all relevant allocators tomorrow. Note that we have two brand new allocators who started less than 2 weeks ago that I think would probably be exempt from the interviews but let me know your thoughts. They are still onboarding! @ALL_005 @ALL_006 @ALL_007, you all are next up on the interview list! Could you three connect and send some times over to DEV_002 for the next round of interviews? Should only take 15-20 minutes per person. Thanks, ALL_001 --- From: ALL_006 (Allocator) To: DEV_002 (Developer) CC: ALL_001 (Head of Allocator) Subject: RE: Interviews … Sure! I can’t do tomorrow, but I can do 7/21 or 7/22. Best, ALL_006	
Task Centrality (Project 1)		
Allocation is a central task in the allocators job.	"*That's what we do*: we do manual adjustments because the system gives us crazy numbers. We're always doing manual adjustments --annoying, but *I guess that’s the reason why we have jobs right now*.” (ALL_001; emphasis added) --- "[…explaining how he makes allocation decisions] So, across all categories, I have about 2.5k products available in DC, if I want to see what I have available by size, I have to click here [points]… It's all very slow today. *But this is what I do every day*." (ALL_006; emphasis added)	Making allocation decisions was the main task that allocators carried out on a day-to-day basis, and they believed that their ability to do this task well was the main reason they still had a job.
Allocators perform few other tasks besides making allocation decisions.	"It takes a little bit of time, usually on Monday, to go through and evaluate the stores, to understand if there are any major problems. But then, the rest of the week, we just allocate products unless there are major problems.” (ALL_004)	Usually on one day a week, before starting to allocate the products, allocators would

		review the status of the inventory per store resulting from their previous allocation decisions. The rest of their time was spent predominantly on making allocation decisions.
Homogeneous Task Enactment (Project 1)		
Allocators followed broadly consistent procedures across different product categories and stores.	*Interviewer:* "If you have 10 units in stock [in the central warehouse], is your strategy to allocate all 10 units or to allocate the minimum number of units necessary for that store? ALL_006: "We *always* allocate all available units (his emphasis). ——— *Researcher:* Is there anyone who allocates from other Distribution Centers instead of [DC Location]? *ALL_006:* "No, we all allocate all products from [DC location]. We just have different categories of products, but they are distributed in such a way that our workload is balanced among us.	Allocators have a lot of standard procedures. As the first quote illustrates, the allocator seemed confident that they should "always" allocate all available products, as if this was a corporate policy. In the second quote, the allocator specifies that even when dealing with different product categories, all allocators focus on a single distribution center.
Jurisdictional Ambiguity (Project 2)		
Store managers, district managers, and the retail finance managers each contested for control over the turf of workforce scheduling decisions.	*[During a store visit]* *DEV_002 (developer):* So, your assumption is that store workers get overwhelmed when store traffic increases to a certain rate, which basically explains the declining trend in conversion. *RET_005 (district manager):* Yeah, I mean… [not clear if he understood what the developer said], the longer the interactions between store workers and customers, the higher the probability we can gather customers to buy something. *DEV_002:* Are you suggesting then - in this case - if you increase the sales force within the store, your conversion rate should go up? *RET_005:* Yes, correct. ——— "You know, when there is a peak of traffic in the weekend, [store managers] are more likely	As the first two quotes illustrate, district managers strongly believed that optimal workforce scheduling was a matter of adjusting scheduling based on anticipated customer traffic. They also seemed to believe that store managers were not paying enough attention to anticipated store traffic in their scheduling decisions. As the third quote illustrates, store managers viewed the retail finance team's 'Census model' to unduly constrain their decision-making. As such, they had negotiated with their managers to be able

	to react. But when it happens during the week – say Monday – [store managers] are usually unprepared" (RET_004 (district manager)) "So there's a document [from retail finance] that comes in on Mondays - and it basically tells me what my payroll spend was last week and if I was over or under. But I know that [district manager] gave me a certain amount of hours that I could assign regardless of what [retail finance] said at the beginning of the month. So I'm kind of filling [the Census model] out because I need to, but I also know that I'll be okay to have certain expenditure regardless… (RET_008)	to bypass this tool to make scheduling decisions as they saw fit.
Store managers were embedded in multiple, complex lines of authority	From: RET_002 (regional manager) To: DEV_002 (developer); RET_005 (district manager) Subject: West Coast Data Labs visit Hello All, I wanted to get this group together as an introduction email to DEV_002. He will have a team representative on the west coast, and would like to have his team meet some [store managers] with the purpose of discussing payroll planning and how stores are executing with current tools today. … DEV_002, take it away! Best, RET_002 From: RET_005 (district manager) Hi DEV_002, Looking forward to meeting you and contributing on this project! Wanted to reach out to coordinate on visits, and to clarify scope. The email from RET_002 indicates that this is primarily a project looking at store managers of outlet stores; I cc-ed RET_10, who is a Retail store manager – wanted to verify that you will want to meet with both channels? … Any additional info about what you'd like to accomplish in stores and 'shadowing' would also be helpful (reports, data needed, etc.) Happy to connect live to discuss if that's easier! Best, RET_005	In the second project, access to store managers required traversing multiple lines of authority. This email exchange illustrates one instance of connecting to a regional manager, who connected developers to a district manager, who finally connected them to a store manager. However, each such interaction would at most help them access 2-3 store managers within a specific region (here, the 'West Coast').

Task Peripherality (Project 2)		
Maximizing sales conversion [rather than workforce scheduling] is central to the job of store managers	*RET_011:* So, we have the "conversion rate" metric, which is the main number for me. *You'd always want your conversion rate to be high – that's our job.* I'm not happy with my conversion from last week… So my average conversion throughout the year is 8.2%. [...] but last week my conversion was down 25%... *Interviewer:* Why is that? Did you made a mistake in scheduling your workforce? *RET_011:* "Yeah, no, customers didn't want to buy anything. When I think about last week, there was a lot of rain, the weather was terrible..."	Store managers seemed to care most about maximizing their conversion rate, and viewed this as the core of their job. They did not see workforce scheduling as inextricably connected to maximizing sales conversion – instead, they would often attribute bad or good sales to exogeneous factors such as the weather.
Store managers perform a wide variety of different tasks [besides workforce scheduling].	*Interviewer:* When do you usually do the scheduling? *RET_011:* "I usually do Mondays – just one day a week – when I do all my *office work*" (emphasis added). "There are a lot of activities that I must control. Scheduling is one of them, and I'm sorry if it's not always perfect… but I'm overwhelmed!" (RET_007)	Store managers often talked about scheduling as a marginal activity that needed to be done weekly, but that they would try to get out of the way as soon as possible. They would make scheduling decisions in the afternoon when there were few customers in their store, or when they otherwise had free time.
Heterogeneous Task Enactment (Project 2)		
Store managers used various heuristics to make their scheduling decisions	*RET_009 (store manager):* "… I schedule 425 hours a week: that gives me seven to eight people a day, with nine on the weekends. On Friday and Saturday I definitely need nine, because [points to store traffic data] look at the traffic… And Sunday is my highest traffic day, so I'll always layer in a bit more. This schedule basically stays the same". *DEV_005 (developer):* "So, let me understand: you keep the scheduling constant for every week of the year?" RET_009: "Yes. I mean outside of the holidays and seasons. In November, December, and	These quotes illustrate how store managers use various heuristics to plan their schedules. All three kept their schedules largely unchanged on a week-to-week basis, for different reasons. The first basically copies and pasts the same schedule every week, only making changes for busy seasons or holidays. The second would allocate more hours on the weekend and fewer on weekdays to balance things

	January, you'll need more hours – I'll hire more than 10 people. Then in late January everything gets back to normal…" "… I already planned that Friday, Saturday, and Sunday are naturally going to be a bit [more crowded], so we'll need more people. So I'll do that, and then schedule a little less on Monday and Tuesday" (RET_012 (store manager)). "So, my base schedule is not really going to change much every day. For me to make this business work, I need to have at least seven people, then I will add an eighth or ninth person on the weekend. So that's kind of the puzzle. You start by putting all the hours for [P1 workers: i.e., experienced full-time workers with administrative responsibilities]. And then you fill in the holes, early shift, mid shift, late shift… So it's basically a copy and paste every week, unless something very big or small happens…" (RET_006)	out. And the third would start by maxing out the hours for his leadership team [P1 workers], and then dividing up the remaining hours among his contract staff.
Store managers used a variety of different information management practices.	"We keep our own ledger, which is where we type out our daily activities – e.g., what we did yesterday, how much we earned, how much traffic we had, etc… we'll write notes about all of it. So we can use this ledger to plan, for instance, what promotion we run each year at this time" (RET_011). Researcher: "Oh, there are a lot of emails printed here!" [points to the right of the store manager's desk. There was a calendar hanging on the wall, surrounded by several printed cut-outs of emails describing various store events]. RET_007 (store manager): "So this is (our) Outlook calendar. If I get an Outlook invitation for something, it pops up on this calendar… If the email has a note for an event, I print that out to keep track of it. When the date comes up, I just take it off the wall."	Store managers seemed to use a variety of different tools to gather the information they needed to make scheduling decisions. The first store manager reported having a private ledger where she keeps relevant information to use for workforce scheduling; the second store manager instead printed out emails she received to keep track of upcoming events.
	"I think I've been with a brand for 14 years… I do a lot of the recruiting and hiring and training for the district. So even though it's my store and I'm the store manager,… I also do a lot of the recruiting. So, for example, say I need to hire someone. I would go out and find someone that I'm like, "Okay, I think this candidate has great potential. I want to hire them." I would then pass them on to my district manager for a second interview. If we agree about the fit, we'll hire this person. I kind of took on a bit of extra work – I'm not supposed to be	The two quotes contrast the experiences of tenured versus non-tenured store managers. In the first case (first quote), the store manager seems to have much more influence over hiring decisions, being able to select the people she wants in her store, and negotiating for more labor as needed.

	involved in this, but I'm going to help [district manager] find the right people…" (RET_010) "As a store manager, you don't fire, you don't hire… Sometimes we get people that we don't like, and we have to find a way to make them sell. That's it." (RET_012)	In the second case (second quote), new hires seemed to be decided arbitrarily by the district managers, and had to be accepted by the store manager.
Store size affects the type and number of workers available in the roster	"I have 30 people on staff, which looks really great, but I have a majority of P3 [temporary contract workers]… For P3, I can assign anywhere from four to 19 hours. But a lot of them have school, or they have second jobs. They work one day a week. So, it's not great to have them." (RET_009; store manager of a small store) "I do not have a P3. I don't like P3s – they usually only work one or two days, up to 18 hours. In my store, everybody is P2: which is 22 to 25 hours, or three days a week. And all the management is full time. If I had somebody who could only come in one day a week… they're just playing catch-up for half of their shift. So it doesn't really make sense… You can't really have that person out there selling." (RET_010; store manager of a large store) "So my store did a little over $2.7 million last year… I should be ranked either number two or number three in the district. When volume is higher you can have very different complexities in managing a store. We are a flagship location… When you're flagship, there's a lot of different elements that kind of come into place – from product assortment to staff size… We have a lot of people!" (RET_010; store manager of a large store) "Okay, so I have eight people. At least four people have to own keys: that is supposed to include the store manager and assistant store manager… But I don't have an assistant store manager underneath me. So, we have four people who hold keys, handle registers, handle all the opening duties, and so on… We don't really need to manage people, but everyone is assisting with opening and closing the store" (RET_012; store managers of a smaller store)	Store size affected the type and number of workers available for store managers to schedule. As the first two quotes illustrate, managers of smaller stores were worried about having too many [P3] temporary, part-time workers that could not be trained properly. Larger stores, however, were able to make sure their roster comprised workers with longer [P2] contracts. We later found that district managers seemed to force small stores to hire students or other part-time seasonal workers, to keep labor costs low. Store size also affected the number of workers available. As the latter two quotes illustrate, larger stores had a larger roster and could play with this flexibility to allocate more or less hours as needed. In small stores, managers only seemed to have the bare minimum number of people to keep the store open.