Rebalancing Worker Initiative and AI Initiative in Future Work: Four Task Dimensions

Jessica He
IBM Research
Yorktown Heights, NY, USA
jessicahe@ibm.com

David Piorkowski
IBM Research
Yorktown Heights, NY, USA
djp@ibm.com

Michael Muller
IBM Research
Cambridge, MA, USA
michael_muller@us.ibm.com

Kristina Brimijoin
IBM Research
Yorktown Heights, NY, USA
kbrimij@us.ibm.com

Stephanie Houde
IBM Research
Cambridge, MA, USA
stephanie.houde@ibm.com

Justin D. Weisz
IBM Research
Yorktown Heights, NY, USA
jweisz@us.ibm.com

ABSTRACT

Organizations have recently begun to deploy conversational task assistants that collaborate with knowledge workers to partially automate their work tasks. These assistants evolved out of business robotic process automation (RPA) tools and are becoming more intelligent: users can initiate task sequences through natural language, and the system can orchestrate those tasks if they have not previously been defined. As these tools become more automated, system designers tend to optimize overall process efficiency, but at the cost of shifting agency away from workers. Particularly in high stakes work environments, this shift raises questions of how to re-delegate agency such that workers feel sufficiently in control of automated tasks. We explored this through two studies comprised of interviews and co-design activities with knowledge workers and identified four task dimensions along which their automation and interaction preferences vary: process consequence, social consequence, task familiarity, and task complexity. These dimensions are useful for understanding when, why, and how to delegate agency between workers and conversational task assistants.

CCS CONCEPTS

• **Human-centered computing** → **Natural language interfaces**; *Collaborative interaction*; **HCI theory, concepts and models**; *Participatory design*; • **Applied computing** → Business process management.

KEYWORDS

Task assistant, Intelligent agent, Orchestration of work, Agent skills, Participatory design.

ACM Reference Format:

Jessica He, David Piorkowski, Michael Muller, Kristina Brimijoin, Stephanie Houde, and Justin D. Weisz. 2023. Rebalancing Worker Initiative and AI Initiative in Future Work: Four Task Dimensions. In *Annual Symposium on Human-Computer Interaction for Work 2023 (CHIWORK 2023), June 13–16, 2023, Oldenburg, Germany*. ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3596671.3598572

This work is licensed under a Creative Commons Attribution International 4.0 License.

CHIWORK 2023, June 13–16, 2023, Oldenburg, Germany
© 2023 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0807-7/23/06.
https://doi.org/10.1145/3596671.3598572

1 INTRODUCTION

Intelligent assistants are becoming important to support business users in workplace settings, and are likely become even more important in the near future of work. A common way to interact with these assistants is through a conversational interface, in which human domain-experts and other end-users seek AI-based task assistance for the more mundane parts of their work - freeing them to put their energies into other aspects of their work that require human expertise. Today, we see conversational agents used in *assistant* roles of answering frequently asked questions [36], handling customer service tasks [40, 87, 118], and performing personal tasks such as scheduling meetings [24, 49, 70], ordering food [64], or guiding decision-making processes [4, 28, 53, 85, 89, 108]. The technical capabilities underpinning the agents have also grown, moving toward multi-agent systems with complex orchestration, memory, and planning components [21, 103]. Agents that specialize in one task are now part of a collective that do more complex tasks, calling into question the balance between human and AI initiative [39, 77, 90, 99].

Addressing these questions requires reconsidering the design of such technology to account for user agency, yet as recently as 2017, Lee et al. noted the absence of guidance for the design of task assistants [63]. In this paper, we focus on design issues for task assistants that support more complex activities of business users through conversational user interfaces. These assistants are capable of executing business processes in back-end systems on behalf of the user [20, 21, 91]. Such systems are enabled via multi-agent *orchestration* technologies, in which a front-end dialogue-manager transforms a user's natural-language request into a series of actions, and then into an executable sequence of operations (i.e., the orchestration), powered by an underlying set of AI *skills* that connect to back-end services [20, 21, 91]. The goal is often to relieve the user from some of the tedious aspects of typical business workflows. However, these systems are derived from existing business automation (BA) systems [92], such as robotic process automation (RPA) [3], and as with other automation applications, there are design tensions about building more process rigidity into automated tasks to prioritize workflow efficiency, but at the cost of loss of human control and hence worsened user productivity and satisfaction [8, 17, 99].

To address this tension, it may be helpful to cede some control over task management and execution away from the AI agent and back to the business user. However, this supposition raises questions about when to cede control and what kinds of task-related controls business users want. Hence, we sought to address the following research questions:

RQ1: What factors affect business users' task control needs and automation preferences?
RQ2: Given the findings from RQ1, how should the user experience of a conversational task assistant be designed to give workers sufficient control over their tasks?
RQ3: What design considerations are needed to support future design of conversational task assistants?

We ran two studies to answer these questions. To address RQ1, we conducted an interview and participatory activity with knowledge workers and observed that certain dimensions of tasks affect participants' automation preferences.[1] These learnings informed a follow-up co-design study that sought to address RQ2 by exploring which task dimensions are applicable in the context of a conversational task assistant and what implications they have on system design. Synthesizing across both studies, we describe the task dimensions and associated design considerations to address RQ3.

Our paper makes the following contributions to the human-centered AI (HCAI) community [74, 98, 99]:

(1) We identify four dimensions of tasks that influence people's preferences for task automation: the consequences of task errors, social consequences, users' familiarity with a task, and task complexity.
(2) We elaborate on types of agency and control and ways to delegate them based on these dimensions.
(3) We provide recommendations for how to factor these dimensions into the design of a conversational task assistant to drive user-centered human-AI collaborations in future workplaces.

2 RELATED WORK

We begin by briefly introducing existing business process automation systems. We then detail prior work on task management challenges and task dimensions, and finally touch on applications of AI to support complex tasks.

2.1 Business process automation & RPA

Business processes are typically described as a collection of tasks that are connected together to accomplish an organizational goal [21]. A simple example is a seller asking an AI sales-support agent to parse a request for proposals (RFP). The AI finds components, accesses current pricing information, and proposes a discount rate for this particular customer. The seller can adjust the discount rate and other aspects of the proposal, and then the AI can produce a well-formatted proposal and (with seller approval) submit the proposal electronically to the customer.

The design goal of business process automation is to apply AI to automate portions of a business process to allow humans to use human creative and social skills [68] while computers perform more tedious support operations [39, 90, 97]. One key enabling technology for automating repetitive software tasks is Robotic Process Automation (RPA) [117]. In RPA systems, software "robots" (i.e., AI agents) perform repetitive and highly-structured activities, such as email responses, cancellation and refund of an airline ticket, and information retrieval from audit documents [16, 67]. Recent advances in multi-agent orchestration systems enable more intelligent behaviors to be *composed* (i.e., orchestrated) out of a set of simpler behaviors [103]. For example, a calendar scheduling agent may combine the individual *skills* of natural language understanding, sending emails, satisfying constraints, and creating a calendar invitation (e.g., [59]). However, little is known about how people actually need to interact with multi-agent orchestration systems and how they will integrate them into their existing work practices.

2.2 Challenges of task management

To understand the design space of multi-AI-agent task-support systems we turn to the HCI, IUI, and CSCW literatures on human task management. Much of the early work dealt with complex and intermixed tasks that tended to be imposed on office workers [9, 12, 19, 115, 119]. While some treatments of task management involved interpersonal or team collaboration [22, 42, 66, 71, 75, 82, 114], much of the work focused on the individual task-performer and how they coped with the work-to-be-done [18, 25, 43, 45, 52, 116]. Geyer et al. analyzed the "flood" of an overwhelming number of tasks, which were "scattered" across multiple channels ([42]; see also [43]).

There is also evidence that people's task management practices are highly varied [54, 55, 60, 75] especially in multi-disciplinary teams [43, 71]. This variance has made it difficult to create technologies that support diverse tasks and diverse ways of working [14, 106]. Designs that provided flexibility in structuring tasks, collaborations, and representations seem to have been readily adopted [6, 76, 96], with one such approach being re-implemented as a major collaborative task management product that supported individual users and teams to create their own task structures and collaborative dependencies [42, 73, 81]. But which aspects of tasks and task-management practices have supported this adoption, and how can new AI technologies contribute to task-management practices?

2.3 Task dimensions

Task management and support present complex design and implementation challenges. Researchers have characterized this space in terms multiple *task dimensions* based on classifications and trade-off dimensions. Tasks may be analyzed along dimensions of retrospective vs. prospective [50], informative vs. actionable [101], reminding vs. being-reminded [75], visible vs. invisible [23, 94, 104], content-oriented vs. relationship-oriented [22, 75, 114], and holistic [6] vs. itemized [10, 25, 34, 51]. In a series of studies, Bellotti and colleagues considered task dimensions organized into the diverse functions of

[1] Participatory design is practiced in different ways, in different types of institutions, and for different purposes [15, 80, 83]. In this project, we adapted our participatory approach to industry constraints (e.g., [13, 56]), in which institutional pressures often constrain employees' time to a single participatory session [79], with little opportunity to continue engagement over time [84].

task representations, such as reminders, work estimators, prioritization aids, and status trackers ([9–12, 34]; see also [22, 75]). Brdiczka and colleagues added temporal dimensions and time-patterning ([18, 19]; see also [7, 50]), factors which become crucially important in emergency response [66, 88] and urgent healthcare [62]. Most knowledge workers must manage a multitude of diverse tasks [25, 42, 45, 75]. In this paper, we consider how this analytic space may be both complicated and also re-configured by AI support.

2.4 AI assistance for complex tasks

In recent years, intelligent assistants have become more advanced, taking on more complex tasks and even involving multiple agents in conversation [29, 38]. However, users primarily use these assistants for simple tasks, such as checking the weather and setting reminders. They are much less likely to trust them for complex, socially-sensitive tasks, such as starting a call and sending long emails, in part due to uncertainty over the their capabilities and contextual state [65].

Task-based work has always been socially entangled [13, 75], because one person's work *output* often serves as *input* to their colleague. The use of AI agents leads to new concerns about those entanglements. Employees are concerned with their workplace reputations if an AI agent takes action on their behalf, because AI agents may not understand social dynamics or relational consequences [48] - a set of workplace concerns that Goffman described in terms of impression management [44].

Lack of trust may also be caused by issues of agency, initiative, and control. Humans generally prefer to develop their own strategies and tactics, and to choose their own actions [31, 99, 110]. However, organizations often prefer to direct, channel, and control human actions through operationalizations of tasks in the form of workflows [2, 8, 17, 27, 32, 46]. Multiple *fixed allocation* models have been proposed, in which humans should perform certain types of tasks while computers should perform other types of tasks (e.g., [39, 90, 97]). By contrast, other researchers have centered their projects around the concept of *mixed initiative interfaces*, in which some actions may be flexibly assigned to either human or AI agent, depending on work demands and human and machine resources [30, 47, 57, 77, 78, 102, 111]. Conversational task assistants provide a pragmatic workplace setting to explore these mixed-initiative issues in a conversational structure, where initiative may shift from one party to the other [72].

3 METHODS

We conducted two studies. Previous work had created taxonomies of to-do actions and atomic task components [43, 61]. Large-scale empirical work provided catalogs of user actions through surveys [109] and through automated mining of activity logs [95]. In this paper, we wanted to complement those methods with users' own statements of their preferences and needs, and therefore we turned to more participatory approaches.

To answer RQ1, we ran an interview-based participatory analysis with 13 business users from IBM, a large international technology and services company based in North America, with offices in nearly 100 countries. We asked participants about their current task management and execution practices, then conducted a participatory analysis based on [79] to understand and elicit considerations relevant to their task control and automation preferences. To address RQ2 and contextualize these considerations, we moved from the analysis of RQ1 into direct participatory design based on [37] of how control and automation features should be incorporated into a conversational task assistant, with an additional 15 business users. We synthesized design considerations across the two studies to address RQ3.

3.1 Participants

We recruited participants from several company-wide Slack[2] channels for communities of interest. Participants self-selected to contribute to the research and were paid the equivalent of $25 USD for participating. Thus, while we used a convenience sample, our colleagues decided whether or not to be part of that sample. Participants were screened through a brief survey that asked them to describe their day-to-day work. For the first study, we selected 13 participants (61% male, 39% female; 31% non-US) who described a tedious or repetitive workflow as part of their knowledge work — i.e., work in which the employee's domain knowledge, skills, and relationships were major determinants of successful performance. We believed that AI-assistance with routine tasks within their work would allow employees to focus on the more intellectually and socially fulfilling aspects of their skilled work. We summarize our participants in Table 1 (top). We report participant numbers for Study 1 as P1xx.

Using the same pool of participants collected for Study 1, we recruited 15 new participants (33% male, 67% female; 20% non-US) for the second study who had not participated in Study 1. We invited participants from three roles: business, design, and research — i.e., roles that had experience with the types of workflows that were candidates for automation. We present an overview of these participants in Table 1 (bottom). We report participant numbers for Study 2 as P2xx.

3.2 Study 1: Participatory Analysis of High-Level Considerations for Automating Work Tasks

3.2.1 Study 1 Design. The first study included two parts that took place in an individual, one-hour, online study session led by one researcher. The first phase consisted of a semi-structured interview on how the participant currently organizes, prioritizes, switches between, and executes their work tasks. The second phase aimed to identify aspects of these practices that had implications for users' task control needs and automation preferences (RQ1). We asked participants to identify a work task that they wanted to automate — one with multiple steps, tools, or collaborators, that a task assistant could facilitate — and conducted a co-analysis to understand when they would want to retain vs. delegate control in an automation scenario and *why*. We then repeated the same analytic activity for one of two "standard" business tasks (either travel approval or procurement, selected by the participant) to understand variations in preferences for same task across different people.

[2] https://slack.com

P#	Role	Work location	Task
P101	Developer + customer records	Costa Rica	Submit expense claim
P102	Research Scientist (Quantum)	US	Manage & organize courses
P103	Sys Admin + sourcing materials	Colombia	Deliver machines to customers
P104	Sr. Design Researcher (review work + coordinate)	US	Schedule customer calls
P105	Supplier Manager (records + reporting)	US	Filter report
P106	Research Scientist (procurement + conferences)	US	Request extra cloud storage
P107	Market Analyst (team coord. + mentoring)	US	Submit expense report
P108	Developer (Agile team coord.)	US	Resolve HR payroll issue
P109	Designer + admin tasks	US	Set up employee benefits
P110	Iteration Manager (monthly ops review)	Mexico	Fill out organization chart
P111	Solutions Manager (set price + contracts)	US	Develop contract for seller
P112	Global Manager (Enterprise Ops)	US	Change a password
P113	Visual Designer (planning + reports + IT issues)	UK	Renew license key
P201	Public Relations + Communications	US	Find decision-maker
P202	Infrastructure Delivery	US	Pull utilization data
P203	Strategy Consultant	Mexico	Approve vacation requests, claim hours
P205	Business	US	Find opportunities with customers
P206	Design Researcher UX Strategy	US	Get survey reports
P209	Executive Design Director	US	Create interview guide
P210	Finance Service Design	US	N/A
P211	Sr. Research Scientist (Carbon Capture)	Brazil	Add new team member to project tools
P212	Researcher (Linguistics)	US	File expense reports, track projects
P213	Sr. Research Manager (Exploratory Science)	US	N/A
P214	Research Scientist (CSCW)	US	Filter emails
P215	Research Scientist (AI-HCI)	US	File expense reports
P216	UX Designer	Mexico	Create report
P217	Brand Designer	US	Get projects for checkpoint
P218	Digital Sales Specialist	US	Send automated emails

Table 1: The participants' roles and tasks used in the participatory analysis activity for Study 1 (top, P101—P113) and Study 2 (bottom, P201—218).

3.2.2 Study 1: Co-Analysis. For the co-analysis, we used a more structured version of the CARD method [79], adapted for remote participation via Mural[3] (see Figure 1). Participants described their task steps, which were recorded onto virtual sticky-notes, and then placed into the appropriate boxes in the template in Mural. These became reference points for discussions that informed RQ1. We discussed the participant's goals, how existing computing resources could help or hinder that work, and how the participant sometimes had to take creative steps to overcome hindrances.

Next, we explored participants' automation preferences for each step in their task. For each step or task, we asked them to indicate the level of automation they would be comfortable with, using the slider under each step (bottom of Figure 1) and *why*. We also probed about their information needs and the kinds of AI collaboration and controls they envisioned for each of their selections. These questions helped us understand the needs and barriers that participants had regarding automating work tasks.

3.3 Study 2: Participatory Design of a Task Automation Experience

3.3.1 Study 2 design. In the second study, we moved from analysis of workers' considerations for task automation, to *how* they wanted to specify that automation. We ran a participatory design activity using an online version of paper-prototyping for co-design [37, 105] in the low-fidelity Balsamiq[4] prototyping tool to contextualize findings from Study 1 in a conversational task assistant. This activity supported a discussion of how to design an automation experience that provides users with an appropriate amount of control over their tasks (RQ2). Each study session involved one participant and one researcher, was held remotely, and lasted approximately one hour. We began each session by introducing the concept of the conversational task assistant we were designing. The researcher described the capabilities and limitations of the assistant at a high level, then showed a brief video demo of a user working with the assistant. Figure 2 shows the core components of the co-design activity: an empty conversational user interface (CUI) on the left (Figure 2A) and an empty workspace view (WS) on the right (Figure 2B), with a set of sketched UI components that participants could place into the CUI or the WS (Figure 2C). To provide participants with flexibility in specifying their preferred interaction modalities, we provided a split-view interface and both conversational and GUI controls. During each session, both the researcher and participant could move components on the screen, type text and contribute to the interface through the shared screen in Balsamiq.

3.3.2 Study 2: Co-design activities. We co-designed for three scenarios:

[3]https://www.mural.co

[4]https://balsamiq.com

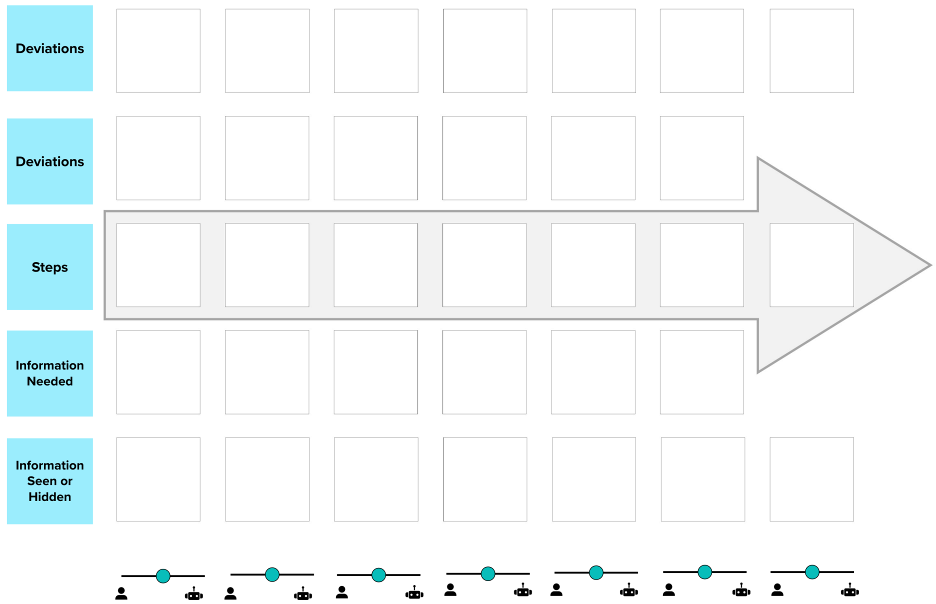

Figure 1: The template used for users to break down tasks into steps, deviations, and information needs. The expected path is captured in the row outlined by the arrow labeled "steps". For each step added, participants indicated how much automation they would be comfortable with using a slider widget along the bottom. Alt-text: The figure depicts a blank template with five rows of either six or seven empty squares. Each row has a label on the left. From top to bottom, they are: deviations, deviations, steps, information needed, information seen or hidden. The middle row, labeled steps, is outlined by a big arrow pointing to the right. Underneath all the rows of squares are seven sliders with the toggle set in the middle, one under each square. The left end of each slider has an icon of a person; the right end has an icon of a robot.

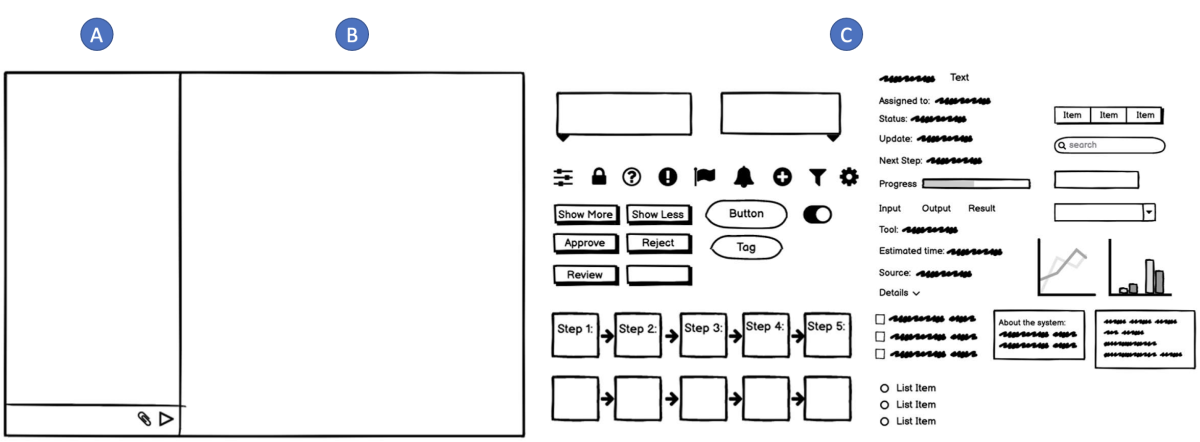

Figure 2: Low-fidelity prototypes and controls. We provided this low-fidelity version of the conversational task assistant and a kit of controls for the participatory design activity. Within the assistant's window there are two main views: (A) conversational user interface (CUI) on the left and (B) workspace view (WS) on the right. Participants were encouraged to use whatever controls (C) were appropriate for their task, including designing their own controls. Alt-text: On the left (A) is a blank low-fidelity wireframe of a chat interface (CUI) and a (B) blank workspace on the right. To the right of the wireframe is (C) a kit of common GUI controls, including chat bubbles, icons, buttons, squiggly lines to represent text, tags, checkboxes, radio buttons, dropdown menus, text fields, a search bar, tabs, and charts.

(1) *Managing multiple tasks*: Intelligent assistants often support work on multiple tasks at the same time, and we wanted to understand workers' needs for agency in managing concurrent tasks. We gave participants a scenario in which they are working on several ongoing tasks with the assistant and asked how they would like to organize and manage these tasks.
(2) *Scheduling a meeting*: We asked participants how they would like to work with an assistant to schedule a meeting, one that included attendees with unknown availability. Asking all participants to co-design for the same task allowed us to compare desired agency and controls across individuals.
(3) *Participant's choice*: Participants also selected a task of their own that they would like automated. Since the first study revealed that needs vary by different types of tasks, this allowed us to get a greater breadth of tasks.

For each activity, we asked what UI controls the participant needed and how they envisioned interacting with the assistant. These activities served as bases for discussions on participants' needs and expectations for the assistant.

Finally, given observations from Study 1 that characteristics of tasks can affect automation needs, we probed participants on scenarios where aspects of their tasks were altered to identify salient characteristics and how they impacted their needs. For example, if the participant described a scheduling task as low-consequence, we asked them to reconsider the scheduling scenario with higher-stakes and modify the design to accommodate this change as needed.

3.4 Data and Analysis

In total, we collected 28 hours of video interview data and approximately 1200 pages of transcripts from Webex[5]. We made annotations into the transcript as needed to clarify participants' words by their actions as recorded in the video. Using a combination of participants' responses and the videos' transcriptions, we conducted an inductive thematic analysis [107] to analyze the data for factors that affected participants' automation preferences and task control needs. We constructed four major themes, which we describe as characteristics of tasks. These formed the basis of what we termed task dimensions, described in the first four subsections of Section 4. We then analyzed how users preferred to work with the assistant in the context of these dimensions to understand design implications for task assistants. We also list two follow-on consequences of those four themes, in the remaining two subsections. The following section presents results across both studies.

4 RESULTS

In RQ1, we asked what factors are relevant to business users' task control needs and automation preferences. Through two studies comprised of interviews and co-design activities with knowledge workers, we identified four dimensions of tasks along which these needs and preferences varied: process consequence, social consequence, familiarity, and complexity (Table 2). These dimensions are useful for understanding **when**, **why**, and **how to delegate agency** to between humans and an intelligent task automation assistant (RQ2, RQ3). We detail each of these below.

4.1 Process consequence

When asked for tasks to automate, participants described a mix of tasks with both minimal and significant consequences of error. For example, P110 considered filling out an organizational chart to be low-consequence and thus acceptable to be automated. If there was a mistake, the outcome would be relatively inconsequential and easy to fix. Conversely, P109 described an onboarding task that was high-consequence due to the sensitive information requested of them. They were concerned about how the system would use the information: *"if I was entering personal information, like tax information or health insurance, I would want to know if that information is secure, if it's going anywhere after that, or if they just delete everything and it doesn't stay in any kind of database."* There could be grave consequences if a task assistant were to disclose personal information to the wrong person, like identity theft and unauthorized purchasing. In an industry setting where disclosure may be on a "need to know" basis, sharing information should be a *human* decision, made with contextual and organizational understandings, regarding both the nature of the information and the needs of the recipient. Participants required additional oversight and control to facilitate these kinds of decisions. Hence, we find that the **process consequences** associated with a task is a dimension that affects users' information and agency needs.

4.1.1 Implications of process consequence. Our studies revealed several instances where significant process consequences would deter users from automating a task. These examples imply that only low-consequence tasks would be considered for automation, but further probing into participants' concerns revealed ways in which they could enjoy expedited work with human-centric safeguards to mitigate risks. Among them were the ability to preview to-be-automated actions in greater detail (exemplified by P203's designs in Appendix A), verify outputs, and identify consequential steps of the upcoming task. This granular insight could make the user feel comfortable with and in control of the automation. In P109's onboarding task, insight into how the system would handle personal information would provide them with the agency to decide whether to provide the system with sensitive input. Such features can draw attention to the risks within a task to help the user understand the system's behavior and give them the control to mediate potential risks.

4.2 Social consequence

Some participants chose tasks involving other people, including teammates or clients, as candidates for automation. For example, P212 co-designed a scenario in which the assistant would *"email everyone on [a]... project,"* and P104 wanted to schedule customer calls. When asked about automation preferences, P104 emphasized that initial contact with a new customer should be handled by a human *"to establish the... relationship with the customer."* Similarly, P111 felt that following up with customers should be human-driven, and P106 wanted *"to be able to control how widely communication [goes]."* P215 co-designed a notification initiated by the assistant that informs them if a task involves important people, which alerts them to

[5] https://www.webex.com/

Dimension	Description
Process consequence	The user's perceived cost of failure when the assistant makes a mistake.
Social consequence	The user's perceived risks in allowing an assistant to represent them.
Familiarity	The user's knowledge about how to complete a given task.
Complexity	The overall difficulty or time required to complete a task.

Table 2: The four task dimensions, identified from both Study 1 and Study 2.

pay closer attention (see Appendix B). Across both studies, we saw that the involvement of other people, and sometimes specifically who is involved, reduced the desire for automation. Where interpersonal skills and emotional sensitivity are required, participants felt that the task assistant lacked the emotional intelligence to handle the task on its own. This insight indicates that **social consequence**, including reputation-management and impression-management, is another dimension of tasks along which information needs and desired control vary.

4.2.1 Implications of social consequence. Automated assistance was generally thought to be best left out of tasks involving other people due to concerns about social errors. However, there were exceptions. Similar to the process consequence dimension, participants felt more comfortable letting the assistant deal with human communication when stakes were low. They described using the assistant to facilitate collaborative steps such as gathering information and documents or managing access permissions in collaborative tools. For example, P108 described how they wanted the back-and-forth with HR tickets automated away once they had described their issue. P110 was likewise comfortable with the assistant sending out requests for colleagues to fill in gaps in their organizational chart.

4.3 Familiarity

Another way that participants' tasks varied was by **familiarity** to the user, which became an important determinant of how much system assistance and initiative participants wanted. Participants explored tasks with low familiarity (i.e. they had little to no knowledge of the task) and high familiarity (i.e. they had sufficient knowledge to execute the task). Note that familiarity is dynamic – a familiar task may become unfamiliar if left untouched over time, and experienced workers can encounter new tasks that are unfamiliar to them. In low-familiarity scenarios, participants asked for more from the system in two ways: (1) more transparency into system actions and (2) guidance on how to complete the task. As unfamiliar tasks become more familiar over time, participants expected the system to reduce its support accordingly. P212 co-designed low and high familiarity versions of the same expense approval task, showing how they would no longer require step-by-step guidance when the process became familiar and would instead initiate with *"Submit an expense report for <event>. Here are the dates and locations. Here's a folder of the receipts"* (see Appendix C).

4.3.1 Implications of familiarity. The co-design activities revealed that the amount of desired automation was inversely proportional to the level of familiarity with the task, suggesting the need for a system that uses knowledge of its user to calibrate its automation. Such a feature would support participants' overarching rationale for using automation in the first place: efficiency. P105 said, *"I'm trying to save time if I'm trusting an automated procedure."* Providing more assistance in low-familiarity tasks reduces the cognitive load of understanding how to execute the task, while minimal guidance for high-familiarity tasks reduces tedium and saves users time.

There are several ways in which an assistant can support efficiency within the familiarity dimension. When more guidance is desired for unfamiliar tasks, the assistant might, as P203 and P216 described, present the tasks as a sequence of steps with explanations, much like a checkout workflow for online purchases. Demos or tutorials may accompany unfamiliar tasks, as suggested by P210's and P211's designs. P209 described their ideal scenario: *"[the assistant]... calibrates the experience, you know, so I have more support or I have more dialogue to... step me through what I need to do."* Similarly, P212 said, *"If it's something I haven't done before, having kind of explicit guidance, like questions or prompts, could be helpful... maybe it would give me some instructions."*

If the system has awareness of when unfamiliar tasks shift to familiar, it can even request autonomy from the user, as P215 co-designed in an interaction: *"We can see that you've submitted expense reports with this title to this cost center and submitted to your manager for approval before. Do you want us to do the same thing for the newest costs?"*.

4.4 Complexity

In the co-design activities of Study 2, participants adapted their designs to reflect the **complexity** of tasks and steps. Participants showed that they preferred chat for simple workflows and more traditional graphical user interfaces (GUIs) for more complex, information-rich workflows. One reason for this difference was due to the perceived richness of these interactions. Namely, natural language-based interactions were more efficient when participants had small amounts of information to convey to the assistant, such as an everyday meeting scheduling task that they considered straightforward. In contrast, for tasks that they considered more complex, the richer interactions of GUIs afforded participants with more control, such as direct manipulation of a calendar (exemplified by P217's designs in Appendix D).

4.4.1 Implications of complexity. Similar to the social consequence dimension, participants did not wish to delegate agency to the assistant when the task was complex, despite the reduction of cognitive load that automation would afford. However, participants suggested that the additional details and user control provided by a richer interface could make automation of complex tasks more desirable. For a meeting scheduling task that was more complex than usual, P203 said, *"If the meeting is with more people... probably I would prefer another kind of interaction, maybe a traditional one where I* *can see the schedule of the people."* Here, the assistant could highlight available times in a visual schedule or request that an invitee free up slots their calendar. Similarly, when designing a vacation approval automation, P203 preferred to converse with the assistant for a single, previously-cleared approval (in other words, a simple approval) but required a GUI to handle a more complex scenario with multiple unknown requests. In both cases, a GUI grants more information and control than a conversational interface, making it more conducive to automating complexities.

4.5 Evolving agency with trust

Across the task dimensions, trust was a key determinant of participants' willingness to delegate agency to the conversational task assistant. In tasks with potential for damaging consequences such as unintended release of confidential information, distrust of the system played into people's desire to maintain their agency. For instance, P107 spoke of submitting an expense report – a task that could have financial ramifications if done incorrectly: *"if I hadn't gotten that trust yet, then I'd probably ask the system, 'prepare the expense report for my review' rather than letting the system submit on its own."* P105 also commented that the assistant would have to prove its trustworthiness over time, adding that they wanted to *"watch it first"* to build that trust (similar to [33, 113]). This sentiment aligns with prior work on trust-building between people and AI systems [41, 69, 100].

4.6 Managing automated tasks

We hypothesized that automating multiple varied tasks at the same time would also require the task assistant to surface meta-information about tasks, such as progress status, to facilitate human-AI collaboration and foster trust. To understand how to achieve this, we asked participants how they currently manage multiple concurrent work tasks to identify aspects that could be applied to a conversational task assistant. Participants' tools and approaches varied greatly, echoing prior work that reported high-variability in task practices [54, 55, 60, 75]. However, across varied approaches, several characteristics of existing tools emerged as valuable to participants and relevant to the design of task management capabilities in conversational task assistants. These characteristics are detailed in Table 3.

These characteristics have several implications on managing automated tasks, which participants detailed through co-design. The need for visual tracking implies that a purely conversational interface may be insufficient for efficient task management. When asked how they would track multiple tasks in a task assistant, 12 out of the 15 participants in Study 2 co-designed a graphical, rather than conversational, interface. Their designs sometimes included project tags, icons denoting priority or urgency, and progress bars – visual indicators to glean information at a glance.

Participants also co-designed ways to filter and sort (e.g. by priority, project, deadline) as another layer of control. Finally, participants hoped that the assistant could help them prioritize and push reminders. P203 envisioned that as a deadline approached, the assistant could send a message that said, *"We haven't completed this high priority task. Do you want me to help complete that?"* Overall, we see that in a system that's doing work on a user's behalf, providing insight and functionality for task management might improve users' experiences working with the assistant.

5 DISCUSSION

Our results may inform the design of mixed-initiative human-AI collaborations for future systems and future workplaces (e.g., [30, 57, 78]). We begin with a review of the four task dimensions identified in our studies. We contextualize them in issues reviewed in Related Work (Section 2) and elaborate on the novel ways in which they should be considered in future task automation systems. Finally, we consider how future workers may want to personalize their AI agents and AI assistants. We hope that these user-stated concerns can be combined with earlier taxonomies of to-dos, tasks, and actions [12, 43, 59, 63], to guide the orchestration of more complex and contingent activity sequences.

5.1 Task dimensions

Through two studies with business users, we identified four dimensions of tasks which we believe will be important in designing future conversational AI tools for workers.These dimensions mediated business users' task control needs and automation preferences (RQ1), and participants' comments and designs also informed ways in which these needs could be realized in the design of a conversational task assistant to support a more human-centered automation experience (RQ2). We consider each of these dimensions in terms of the prior research summarized in Section 2.3, then reflect on broader considerations for designing future task assistants (RQ3).

5.1.1 Process Consequences. Several papers have discussed users' needs in relation to potential risks and costs of process-oriented errors to the self and to the team, including losing track of tasks (especially delegated tasks) [75]; integrating information from multiple, scattered channels [114]; and costs of task-interruptions [25, 45]. Our findings in Section 4.1 extends these issues to include *degrees* of consequences and participants' concerns that contributed to them. In particular, we identified low-stakes vs. high-stakes risks, privacy issues, and participants' needs for insight into and control over these kinds of risks. Some of these attributes may be calculated from a worker's history with similar tasks, or from privacy-protective socially-translucent [35] comparisons with other workers who are doing similar jobs.

As AI assistants move into these complex and previously human-responsible domains, end users will need ways to know if the assistants are doing what humans would want them to do. To support this outcome, we propose that future task orchestration systems should offer transparency into their task sequences, along with review and confirmation steps that provide users with decision-making authority over steps with potentially costly errors. For some tasks, workers may want to examine and choose alternative orchestrations.

5.1.2 Social Consequences. Social consequences in task completion and coordination have been discussed by Bardram et al. in a healthcare context [5] and by Chasanidou et al. in a task management context [22]. Bardram et al. focused on the division of well-structured labor. Chasanidou et al. described an interaction of

Participants (N)	Feature	Description
P103, P107, P109, P110, P112, P201, P203, P206, P209, P211 (10)	Visual indicators	Using visual indicators, such as color coding and tags, to glean important information at a glance supports efficient task management.
P103, P112, P203, P206, P211, P212 (6)	Prioritization and reminders	With so much work to manage and constantly changing priorities, some participants appreciated systems that helped them prioritize tasks and pushed reminders for upcoming deadlines.
P107, P108, P112, P209 (4)	Filtering	Filtering was a useful feature for participants to quickly find what's relevant to them, especially in management systems with many unrelated tasks.

Table 3: Features of task management systems valued by participants.

task and team structure with tool features and affordances. Similar work has emerged in the context of police and fire services [58], emergency medical teamwork [62], and crisis informatics [88].

We expanded the scope of these concerns to consider more *relational* social consequences that participants reported as being important in less structured tasks and technologies. Participants stated concerns about the emotional intelligence of conversational task assistants acting as their communication proxies (e.g., sending email messages in their name). Participants were keenly aware of power and hierarchy issues in their organizations, and wanted to be able to inspect automatically-generated messages if important people were among the addressees. It is possible that conversational task assistants may become capable of managing or mediating social relationships, but participants require more transparency and experience before they would trust assistants to do so.

5.1.3 Familiarity. Task familiarity has been an important consideration in urgent or life-critical activities, such as emergency response [58, 88] and acute medical care [62]. Familiarity has been shown to have complex interactions with explanations and outcomes in simple binary games [93].

We expanded the temporal basis for familiarity effects. Participants wanted a conversational task assistant that understood the state of their task knowledge, perhaps similar to a mental model of the user [86]. Through sensing user understanding, participants hoped for an adaptive user interface experience (e.g., [1, 112]) in which the assistant would gradually reduce the amount of tutorial or remedial support as the user became more familiar with a task.

5.1.4 Complexity. As summarized in Section 2.4, the research literature primarily addresses complexity in terms of a passive human who has to deal with imposed complexity. Challenges include the sheer number and scatter of tasks [12, 42, 114]; the diversity of task types [25, 75]; the distribution of labor across interdependent tasks [5, 22]; the expenditure of time [18, 19] and effort [9, 12]; the demands of multi-tasking [75, 114] and consequent needs for task switches and interruptions [25, 45].

We extended this analysis from passive coping to questions of participants' active *choices* for interaction modalities. Participants preferred chat interfaces for simple tasks but wanted graphical user interfaces for complex, information-rich activities. This sentiment is in line with Media Richness Theory, which states that a lean interface such as text chat may be ineffective for ambiguous tasks [26]. To address this need, task assistants could provide flexibility in interaction methods, or invite users to choose their preferred interaction for a task. For the latter, the assistant might apply principles from task familiarity (Section 4.3) to help users adjust their interface settings and experiences. A more intelligent conversational agent could learn the user's preferences for particular interface modalities associated with particular tasks or degrees of task complexity, or particular collaboration partners.

5.2 Agency, initiative, control, and decision-making considerations in future work automation

The four dimensions are useful when considering the design of a mixed-initiative conversational task assistant because they provide a user-centered perspective into *when*, *why*, and *how* agency should be delegated between the system and the user. Although some models propose that humans and AI assistants may be equally capable of doing certain tasks [47, 78, 102], conversational task assistant architectures are based on more asymmetric principles in which the assistant retains execution capability for many operations, as described in the allocation models of Section 2.4 [39, 90, 97]. Participants' comments and designs encouraged us to re-examine *whether and when* the assistant should retain control. Understanding what *not* to automate is equally important [77], as there is work that users do not trust assistants with at all. Furthermore, in considering *how* control should be delegated, we observed three types of agency: initiative (who gets a task or step going?), control (who decides how the task gets done?), and decision-making (who has a say in what gets done?).

In Sections 4.1 and 4.2, we saw that concerns about process consequence and social consequence – i.e., how the system might mishandle tasks with high consequences of error – made participants less willing to allow the assistant to control and make decisions on such tasks. The assistant has ultimate decision-making authority, because under organizational requirements, the work must be done. However, participants wanted to exercise *control* to inspect, adjust, and potentially prevent the assistant's operation. In these scenarios, additional information and oversight on the task can provide users with the agency to decide whether or not to entrust the assistant with the task. Tasks with extreme risk of consequence, such as renewing a visa, may not be suitable for any kind of automation, even with additional user controls. Nine participants cited consequence concerns as a reason for not wanting automation in such situations.

In discussing familiarity, (Section 4.3), we observed that for unfamiliar tasks, participants wanted to relinquish *initiative* to the assistant but maintain *control* and *decision-making* rights. In these scenarios, they wanted the assistant to initiate steps but they wanted to decide for themselves whether to follow them, and also retain control over the presentation of data and task information. As participants gained knowledge of the task, they wanted control over task options to shift back to themselves to determine which party (human or AI) should perform each task component.

In matters of trust (Section 4.5), we saw that the prior task dimensions could help the user to determine their degree of trust in the conversational task assistant. As with the discussion of familiarity (Section 4.3), participants wanted to inspect and monitor the assistant's operations, gradually releasing agency (in the form of control and even decision-making capabilities) back to the assistant once the human was assured that it would operate as desired.

Participants' discussions of managing automated tasks (Section 4.6) described two principal ways in which they wanted to calibrate trust. First, they required greater *visual* oversight over the assistant's execution of complex and consequential tasks (either procedurally or socially consequential). Second, they proposed a complex *interplay* of *initiative* and *control*, in which the human would choose *what* to do and *when* to do it, and would then trust the assistant to take the initiative to remind the human. We note that, even though the assistant could initiate reminders, participants wanted the human to retain *decision-making* capabilities in regard to whether or not to act on that reminder.

Earlier, we proposed that users might want to exercise control over diverse attributes of their individual and team tasks (Sections 1 and 2.2). Participants showed us how they wanted to architect complex patterns of exchanges in initiative, control, and decision-making, both across and within tasks. For tasks that were required to be done, they nonetheless wanted to control the quality of task execution for consequential outcomes, and they sought to exercise that control by modifying the user interface and system feedback to allow them the right degree of visibility into the assistant's actions. Thus, far from the rigid and predetermined control structures of earlier allocation models [39, 90, 97], we propose that participants saw themselves as co-creators of their task automation experience. This view of end-users as experience-architects may provide a useful extension of the primarily responsive/reactive role for users in Shneiderman's recent framework [99].

5.3 Towards personalized assistants

Participants expected interactions with a conversational task assistant to be adaptive to the them (Section 4.3), proactive to their needs (Section 4.6), and personalized to their way of working. Becoming more efficient in the parts of their jobs that they personally cared about was the primary motivator that participants cited for working with a conversational task assistant. Participants described process goals such as efficiency and time management, as well as priorities related to collaborating with others, as motivating examples. Identifying such priorities through user studies with specific work-groups can help prioritize development of task assistance that focuses on functionality that matters most.

Personalization plays a larger role as assistants become more adaptive and collaborative. Our study revealed that users want to customize messages at higher fidelity, with different messaging for specific persons or groups. Similarly, we saw that users expected to control how they interact with the assistant on a per-task basis both in terms of the interaction and how the task was represented. We expect that as such assistants become more collaborative, the need for personalization will similarly increase and the design of these assistants should take this into account.

5.4 Positionality and Limitations

We conducted this work as members of the Research department of IBM. Our project was intended to inform the design of an enterprise AI-based tool for knowledge workers. Our task was to discover participants' needs and preferences and communicate them to product developers.

All participants spoke English. Because we were holding interactive sessions, we had to find meeting times that participants and researchers could join across timezones. Participants self-selected according to their own convenience. As a result, nearly all participants were based in the North America. Due to this limited sample, we were unable to analyze for meaningful differences between locales. Other institutions, cultures, and settings might find additional task dimensions, or might find different weightings or importances among task dimensions. Additionally, although participants were recruited from different job roles, we were unable to identify significant differences in automation preferences between roles due to insufficient sample size in each role.

6 CONCLUSION

We conducted two studies to understand business users' automation preferences and needs for working with a conversational task assistant. This work identified important user considerations in the context of task dimensions for the design of conversational task assistants:

- We identified four dimensions of tasks along which participants' automation preferences varied: **process consequence, social consequence, familiarity, and complexity**. These task dimensions provide a human-centered perspective into **when, why, and how to delegate agency and control** between system and user.
- Along these dimensions, we elicited several types of agency that could be delegated or exercised by the user: initiating tasks, controlling the task experience (including interaction modalities and amount of system feedback), and making decisions about if or how a task gets done (e.g., via review and confirmation steps).
- The dynamic nature of these dimensions means that users' expectations of how to interact with a conversational task assistant will change over time. Therefore, conversational task assistants should allow users to situationally adapt their task controls for a comfortable and efficient automation experience.

REFERENCES

[1] Pierre A Akiki, Arosha K Bandara, and Yijun Yu. 2014. Integrating adaptive user interface capabilities in enterprise applications. In *Proceedings of the 36th International Conference on Software Engineering*. 712–723.
[2] Aggeliki Androutsopoulou, Nikos Karacapilidis, Euripidis Loukis, and Yannis Charalabidis. 2019. Transforming the communication between citizens and

government through AI-guided chatbots. *Government information quarterly* 36, 2 (2019), 358–367.

[3] Wasique Ali Ansari, Paritosh Diya, Sahishnu Patil, and Sunita Patil. 2019. A review on robotic process automation-the future of business organizations. In *2nd International Conference on Advances in Science & Technology (ICAST)*.

[4] Maryam Ashoori and Justin D Weisz. 2019. In AI we trust? Factors that influence trustworthiness of AI-infused decision-making processes. *arXiv preprint arXiv:1912.02675* (2019).

[5] Jakob Bardram, Jonathan Bunde-Pedersen, and Mads Soegaard. 2006. Support for activity-based computing in a personal computing operating system. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 211–220.

[6] Jakob E Bardram. 2009. Activity-based computing for medical work in hospitals. *ACM Transactions on Computer-Human Interaction (TOCHI)* 16, 2 (2009), 1–36.

[7] Deborah Barreau and Bonnie A Nardi. 1995. Finding and reminding: file organization from the desktop. *ACM SigChi Bulletin* 27, 3 (1995), 39–43.

[8] LS Baumann and RD Coop. 1980. Automated workflow control: A key to office productivity. In *Proceedings of the May 19-22, 1980, national computer conference*. 549–554.

[9] Victoria Bellotti, Brinda Dalal, Nathaniel Good, Peter Flynn, Daniel G Bobrow, and Nicolas Ducheneaut. 2004. What a to-do: studies of task management towards the design of a personal task list manager. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 735–742.

[10] Victoria Bellotti, Nicolas Ducheneaut, Mark Howard, and Ian Smith. 2002. Taskmaster: recasting email as task management. *PARC, CSCW* 2 (2002).

[11] Victoria Bellotti, Nicolas Ducheneaut, Mark Howard, and Ian Smith. 2003. Taking email to task: the design and evaluation of a task management centered email tool. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 345–352.

[12] Victoria Bellotti, Nicolas Ducheneaut, Mark Howard, Ian Smith, and Rebecca E Grinter. 2005. Quality versus quantity: E-mail-centric task management and its relation with overload. *Human–Computer Interaction* 20, 1-2 (2005), 89–138.

[13] Susanne Bødker and Morten Kyng. 2018. Participatory design that matters—Facing the big issues. *ACM Transactions on Computer-Human Interaction (TOCHI)* 25, 1 (2018), 1–31.

[14] David Bogen. 1994. Do Winograd and Flores have politics? *Computer Supported Cooperative Work (CSCW)* 3, 1 (1994), 79–83.

[15] Claus Bossen, Christian Dindler, and Ole Sejer Iversen. 2016. Evaluation in participatory design: a literature survey. In *Proceedings of the 14th Participatory Design Conference: Full papers-Volume 1*. 151–160.

[16] Clint Boulton. 2018. What is RPA? A revolution in business process automation. https://www.cio.com/article/3236451/what-is-rpa-robotic-process-automation-explained.html. [Online; accessed 11-Jun-2020].

[17] Steen Brahe and Kjeld Schmidt. 2007. The story of a working workflow management system. In *Proceedings of the 2007 international ACM conference on Supporting group work*. 249–258.

[18] Oliver Brdiczka. 2010. From documents to tasks: deriving user tasks from document usage patterns. In *Proceedings of the 15th international conference on Intelligent user interfaces*. 285–288.

[19] Oliver Brdiczka, Norman Makoto Su, and James Bo Begole. 2010. Temporal task footprinting: identifying routine tasks by their temporal patterns. In *Proceedings of the 15th international conference on Intelligent user interfaces*. 281–284.

[20] Tathagata Chakraborti, Shubham Agarwal, Yasaman Khazaeni, Yara Rizk, and Vatche Isahagian. 2020. D3BA: a tool for optimizing business processes using non-deterministic planning. In *International Conference on Business Process Management*. Springer, 181–193.

[21] Tathagata Chakraborti, Vatche Isahagian, Rania Khalaf, Yasaman Khazaeni, Vinod Muthusamy, Yara Rizk, and Merve Unuvar. 2020. From Robotic Process Automation to Intelligent Process Automation. In *International Conference on Business Process Management*. Springer, 215–228.

[22] Dimitra Chasanidou, Brian Elvesæter, and Arne-Jørgen Berre. 2016. Enabling team collaboration with task management tools. In *Proceedings of the 12th International Symposium on Open Collaboration*. 1–9.

[23] Juliet M Corbin and Anselm L Strauss. 1993. The articulation of work through interaction. *The sociological quarterly* 34, 1 (1993), 71–83.

[24] Justin Cranshaw, Emad Elwany, Todd Newman, Rafal Kocielnik, Bowen Yu, Sandeep Soni, Jaime Teevan, and Andrés Monroy-Hernández. 2017. Calendar. help: Designing a workflow-based scheduling agent with humans in the loop. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2382–2393.

[25] Mary Czerwinski, Eric Horvitz, and Susan Wilhite. 2004. A diary study of task switching and interruptions. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 175–182.

[26] Richard L Daft and Robert H Lengel. 1986. Organizational information requirements, media richness and structural design. *Management science* 32, 5 (1986), 554–571.

[27] Peng Dai, Daniel Sabey Weld, et al. 2010. Decision-theoretic control of crowd-sourced workflows. In *Twenty-Fourth AAAI Conference on Artificial Intelligence*.

[28] Ben Dattner, Tomas Chamorro-Premuzic, Richard Buchband, and Lucinda Schettler. 2019. The Legal and Ethical Implications of Using AI in Hiring. *Harvard Business Review* (25 April 2019). Retrieved 07-Sep-2021 from https://hbr.org/2019/04/the-legal-and-ethical-implications-of-using-ai-in-hiring

[29] M Gatti de Bayser, Claudio Pinhanez, Heloisa Candello, MA Vasconcelos, M Pichiliani, M Alberio Guerra, Paulo Cavalin, and Renan Souza. 2018. Ravel: a mas orchestration platform for human-chatbots conversations. In *The 6th International Workshop on Engineering Multi-Agent Systems (EMAS@ AAMAS 2018), Stockholm, Sweden*.

[30] Sebastian Deterding, Jonathan Hook, Rebecca Fiebrink, Marco Gillies, Jeremy Gow, Memo Akten, Gillian Smith, Antonios Liapis, and Kate Compton. 2017. Mixed-initiative creative interfaces. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 628–635.

[31] Virginia Dignum. 2017. Responsible autonomy. *arXiv preprint arXiv:1706.02513* (2017).

[32] Paul Dourish. 2001. Process descriptions as organisational accounting devices: the dual use of workflow technologies. In *Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work*. 52–60.

[33] Jaimie Drozdal, Justin Weisz, Dakuo Wang, Gaurav Dass, Bingsheng Yao, Changruo Zhao, Michael Muller, Lin Ju, and Hui Su. 2020. Trust in AutoML: exploring information needs for establishing trust in automated machine learning systems. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*. 297–307.

[34] Nicolas Ducheneaut and Victoria Bellotti. 2001. E-mail as habitat: an exploration of embedded personal information management. *interactions* 8, 5 (2001), 30–38.

[35] Thomas Erickson, Christine Halverson, Wendy A Kellogg, Mark Laff, and Tracee Wolf. 2002. Social translucence: designing social infrastructures that make collective activity visible. *Commun. ACM* 45, 4 (2002), 40–44.

[36] Kshitij Fadnis, Pankaj Dhoolia, Li Zhu, Q Vera Liao, Steven Ross, Nathaniel Mills, Sachindra Joshi, and Luis Lastras. 2021. Doc2Bot: Document grounded Bot Framework. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 16026–16028.

[37] Vivienne Farrell, Graham Farrell, Kon Mouzakis, Chris Pilgrim, and Pauline Byrt. 2006. PICTIOL: a case study in participatory design. In *Proceedings of the 18th Australia conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*. 191–198.

[38] Ethan Fast, Binbin Chen, Julia Mendelsohn, Jonathan Bassen, and Michael S Bernstein. 2018. Iris: A conversational agent for complex tasks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.

[39] Paul M Fitts, MS Viteles, NL Barr, DR Brimhall, Glen Finch, Eric Gardner, WF Grether, WE Kellum, and SS Stevens. 1951. *Human engineering for an effective air-navigation and traffic-control system, and appendixes 1 thru 3*. Technical Report. Ohio State Univ Research Foundation Columbus.

[40] Asbjørn Følstad, Cecilie Bertinussen Nordheim, and Cato Alexander Bjørkli. 2018. What makes users trust a chatbot for customer service? An exploratory interview study. In *International conference on internet science*. Springer, 194–208.

[41] David Gefen. 2000. E-commerce: the role of familiarity and trust. *Omega* 28, 6 (2000), 725–737.

[42] Werner Geyer, Michael J Muller, Martin Thomas Moore, Eric Wilcox, L-T Cheng, Beth Brownholtz, Charles Hill, and David R Millen. 2006. Activity Explorer: Activity-centric collaboration from research to product. *IBM Systems Journal* 45, 4 (2006), 713–738.

[43] Yolanda Gil, Varun Ratnakar, Timothy Chklovski, Paul Groth, and Denny Vrandecic. 2012. Capturing common knowledge about tasks: Intelligent assistance for to-do lists. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2, 3 (2012), 1–35.

[44] Erving Goffman et al. 1959/2002. The presentation of self in everyday life. 1959. *Garden City, NY* 259 (1959/2002).

[45] Victor M González and Gloria Mark. 2004. " Constant, constant, multi-tasking craziness" managing multiple working spheres. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 113–120.

[46] Shinsuke Goto, Toru Ishida, and Donghui Lin. 2016. Understanding crowdsourcing workflow: modeling and optimizing iterative and parallel processes. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 4.

[47] Imke Grabe, Miguel González-Duque, Sebastian Risi, and Jichen Zhu. 2022. Towards a Framework for Human-AI Interaction Patterns in Co-Creative GAN Applications. *HAIGEN 2022: 3rd Workshop on Human-AI Co-Creation with Generative Models* (2022).

[48] Jones Granatyr, Nardine Osman, João Dias, Maria Augusta Silveira Netto Nunes, Judith Masthoff, Fabrício Enembreck, Otto Robert Lessing, Carles Sierra, Ana Maria Paiva, and Edson Emílio Scalabrin. 2017. The need for affective trust applied to trust and reputation models. *ACM Computing Surveys (CSUR)* 50, 4 (2017), 1–36.

[49] Jonathan Grudin and Richard Jacques. 2019. Chatbots, humbots, and the quest for artificial general intelligence. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–11.

[50] Jacek Gwizdka. 2000. Timely reminders: a case study of temporal guidance in PIM and email tools usage. In *Chi'00 extended abstracts on human factors in computing systems*. 163–164.

[51] Jacek Gwizdka. 2002. Email Land–An Exploration of Email User Interfaces Supporting Pending Tasks. In *Conference Companion of UIST 2002*. Citeseer, 9–10.

[52] Jacek Gwizdka and Mark Chignell. 2007. 12. Individual Differences. *Personal information management* (2007), 206.

[53] Karen Hao. 2019. AI is sending people to jail – and getting it wrong. *MIT Technology Review* (21 January 2019). Retrieved 07-Sep-2021 from https://www.technologyreview.com/s/612775/algorithms-criminal-justice-ai/

[54] Mona Haraty, Joanna McGrenere, and Charlotte Tang. 2016. How personal task management differs across individuals. *International Journal of Human-Computer Studies* 88 (2016), 13–37.

[55] Mona Haraty, Diane Tam, Shathel Haddad, Joanna McGrenere, and Charlotte Tang. 2012. Individual differences in personal task management: a field study in an academic setting. In *Proceedings of Graphics Interface 2012*. 35–44.

[56] Nanna Holdgaard and Signe Louise Yndigegn. 2018. Participatory design negotiations in project funding regimes. In *Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial-Volume 2*. 1–5.

[57] Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 159–166.

[58] Amanda L Hughes, Lise AA St. Denis, Leysia Palen, and Kenneth M Anderson. 2014. Online public communications by police & fire services during the 2012 Hurricane Sandy. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1505–1514.

[59] Richard Jacques, Asbjørn Følstad, Elizabeth Gerber, Jonathan Grudin, Ewa Luger, Andrés Monroy-Hernández, and Dakuo Wang. 2019. Conversational agents: Acting on the wave of research and development. In *Extended abstracts of the 2019 CHI conference on human factors in computing systems*. 1–8.

[60] Amirrudin Kamsin, Ann Blandford, and Anna L Cox. 2012. Personal task management: my tools fall apart when I'm very busy! In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*. 1369–1374.

[61] Jihie Kim, Marc Spraragen, and Yolanda Gil. 2004. An intelligent assistant for interactive workflow composition. In *Proceedings of the 9th international conference on Intelligent user interfaces*. 125–131.

[62] Diana Kusunoki, Aleksandra Sarcevic, Zhan Zhang, and Maria Yala. 2015. Sketching awareness: A participatory study to elicit designs for supporting ad hoc emergency medical teamwork. *Computer Supported Cooperative Work (CSCW)* 24, 1 (2015), 1–38.

[63] Sang-su Lee, Jaemyung Lee, and Kun-pyo Lee. 2017. Designing intelligent assistant through user participations. In *Proceedings of the 2017 conference on designing interactive systems*. 173–177.

[64] Xi Yu Leung and Han Wen. 2020. Chatbot usage in restaurant takeout orders: A comparison study of three ordering methods. *Journal of Hospitality and Tourism Management* 45 (2020), 377–386.

[65] Ewa Luger and Abigail Sellen. 2016. " Like Having a Really Bad PA" The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 5286–5297.

[66] Faisal Luqman and Martin Griss. 2010. Overseer: a mobile context-aware collaboration and task management system for disaster response. In *2010 Eighth International Conference on Creating, Connecting and Collaborating through Computing*. IEEE, 76–82.

[67] Somayya Madakam, Rajesh M Holmukhe, and Durgesh Kumar Jaiswal. 2019. The future digital work force: robotic process automation (RPA). *JISTEM-Journal of Information Systems and Technology Management* 16 (2019).

[68] James Manyika, Susan Lund, Michael Chui, Jacques Bughin, Jonathan Woetzel, Parul Batra, Ryan Ko, and Saurabh Sanghvi. 2017. Jobs lost, jobs gained: Workforce transitions in a time of automation. *McKinsey Global Institute* 150 (2017).

[69] D Harrison McKnight, Larry L Cummings, and Norman L Chervany. 1998. Initial trust formation in new organizational relationships. *Academy of Management review* 23, 3 (1998), 473–490.

[70] Namita Mhatre, Karan Motani, Maitri Shah, and Swati Mali. 2016. Donna interactive chat-bot acting as a personal assistant. *International Journal of Computer Applications* 140, 10 (2016).

[71] Felix Michel, Yolanda Gil, Varun Ratnakar, and Matheus Hauder. 2015. A task-centered interface for on-line collaboration in science. In *Proceedings of the 20th International Conference on Intelligent User Interfaces Companion*. 45–48.

[72] Robert J. Moore and Raphael Arar. 2019. *Conversational UX Design: A Practitioner's Guide to the Natural Conversation Framework*. Association for Computing Machinery, New York, NY, USA.

[73] Thomas P Moran, Alex Cozzi, and Stephen P Farrell. 2005. Unified activity management: supporting people in e-business. *Commun. ACM* 48, 12 (2005), 67–70.

[74] Michael Muller, Plamen Agelov, Hal Daume, Q Vera Liao, Nuria Oliver, David Piorkowski, et al. 2022. HCAI@ NeurIPS 2022, Human Centered AI. In *Annual Conference on Neural Information Processing Systems*.

[75] Michael Muller, Casey Dugan, Michael Brenndoerfer, Megan Monroe, and Werner Geyer. 2017. What did I ask you to do, by when, and for whom? Passion and compassion in request management. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 1009–1023.

[76] Michael Muller, Kate Ehrlich, Tara Matthews, Adam Perer, Inbal Ronen, and Ido Guy. 2012. Diversity among enterprise online communities: collaborating, teaming, and innovating through social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2815–2824.

[77] Michael Muller and Justin Weisz. 2022. Frameworks for Collaborating Humans and AIs: Sequence and Sociality in Organizational Applications. In *CHIWORK*.

[78] Michael Muller, Justin D Weisz, and Werner Geyer. 2020. *Mixed Initiative Generative AI Interfaces: An Analytic Framework for Generative AI Applications*. Retrieved 13-Jul-2021 from https://computationalcreativity.net/workshops/cocreative-iccc20/papers/Future_of_co-creative_systems_185.pdf

[79] Michael J Muller. 2001. Layered participatory analysis: New developments in the CARD technique. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 90–97.

[80] Michael J Muller and Allison Druin. 2012. Participatory design: The third space in human–computer interaction. In *The Human–Computer Interaction Handbook*. CRC Press, 1125–1153.

[81] Michael J Muller, Werner Geyer, Beth Brownholtz, Eric Wilcox, and David R Millen. 2004. One-hundred days in an activity-centric collaboration environment based on shared objects. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 375–382.

[82] Michael J Muller and Daniel M Gruen. 2005. Working together inside an email-box. In *ECSCW 2005*. Springer, 103–122.

[83] Michael J Muller and Sarah Kuhn. 1993. Participatory design. *Commun. ACM* 36, 6 (1993), 24–28.

[84] Michael J Muller, Daniel M Wildman, and Ellen A White. 1992. Taxonomy of participatory design practices: a participatory poster. In *Posters and short talks of the 1992 SIGCHI conference on Human factors in computing systems*. 34–34.

[85] John Murawski. 2019. Mortgage Providers Look to AI to Process Home Loans Faster. *Wall Street Journal* (18 March 2019). Retrieved 07-Sep-2021 from https://www.wsj.com/articles/mortgage-providers-look-to-ai-to-process-home-loans-faster-11552899212

[86] Mahsan Nourani, Chiradeep Roy, Jeremy E Block, Donald R Honeycutt, Tahrima Rahman, Eric Ragan, and Vibhav Gogate. 2021. Anchoring Bias Affects Mental Model Formation and User Reliance in Explainable AI Systems. In *26th International Conference on Intelligent User Interfaces*. 340–350.

[87] Mohammad Nuruzzaman and Omar Khadeer Hussain. 2018. A survey on chatbot implementation in customer service industry through deep neural networks. In *2018 IEEE 15th International Conference on e-Business Engineering (ICEBE)*. IEEE, 54–61.

[88] Leysia Palen, Sarah Vieweg, Jeannette Sutton, Sophia B Liu, and Amanda Hughes. 2007. Crisis informatics: Studying crisis in a networked world. In *Proceedings of the Third International Conference on E-Social Science*. 7–9.

[89] DJ Pangburn. 2019. Schools are using software to help pick who gets in. What could go wrong? *Fast Company* (17 May 2019). Retrieved 07-Sep-2021 from https://www.fastcompany.com/90342596/schools-are-quietly-turning-to-ai-to-help-pick-who-gets-in-what-could-go-wrong

[90] Raja Parasuraman, Thomas B Sheridan, and Christopher D Wickens. 2000. A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans* 30, 3 (2000), 286–297.

[91] Yara Rizk, Abhishek Bhandwalder, Scott Boag, Tathagata Chakraborti, Vatche Isahagian, Yasaman Khazaeni, Falk Pollock, and Merve Unuvar. 2020. A unified conversational assistant framework for business process automation. *arXiv preprint arXiv:2001.03543* (2020).

[92] Tariq Samad, Paul McLaughlin, and Joseph Lu. 2007. System architecture for process automation: Review and trends. *Journal of Process Control* 17, 3 (2007), 191–201.

[93] James Schaffer, John O'Donovan, James Michaelis, Adrienne Raglin, and Tobias Höllerer. 2019. I can do better than your AI: expertise and explanations. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 240–251.

[94] Kjeld Schmidt. 2008. Taking CSCW Seriously: Supporting Articulation Work (1992). In *Cooperative Work and Coordinative Practices*. Springer, 45–71.

[95] Alexander Seeliger, Benedikt Schmidt, Immanuel Schweizer, and Max Mühlhäuser. 2016. What Belongs Together Comes Together: Activity-centric Document Clustering for Information Work. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*. 60–70.

[96] Shilad Sen, Werner Geyer, Michael Muller, Marty Moore, Beth Brownholtz, Eric Wilcox, and David R Millen. 2006. FeedMe: a collaborative alert filtering system. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*. 89–98.

[97] Thomas B Sheridan and William L Verplank. 1978. *Human and computer control of undersea teleoperators.* Technical Report. Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab.

[98] Ben Shneiderman. 2020. Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction* 36, 6 (2020), 495–504.

[99] Ben Shneiderman. 2022. *Human-Centered AI.* Oxford University Press.

[100] Keng Siau and Weiyu Wang. 2018. Building trust in artificial intelligence, machine learning, and robotics. *Cutter business technology journal* 31, 2 (2018), 47–53.

[101] Nikash Singh, Martin Tomitsch, and Mary Lou Maher. 2013. A time and place for preparatory methods in email. In *Proceedings of the 14th Annual ACM SIGCHI_NZ conference on Computer-Human Interaction.* 1–8.

[102] Angie Spoto and Natalia Oleynik. 2017. *Library of Mixed-Initiative Creative Interfaces.* Retrieved 19-Jun-2021 from http://mici.codingconduct.cc/

[103] Sarath Sreedharan, Tathagata Chakraborti, Yara Rizk, and Yasaman Khazaeni. 2020. Explainable Composition of Aggregated Assistants. *arXiv preprint arXiv:2011.10707* (2020).

[104] Susan Leigh Star and Anselm Strauss. 1999. Layers of silence, arenas of voice: The ecology of visible and invisible work. *Computer supported cooperative work (CSCW)* 8, 1 (1999), 9–30.

[105] Brian Still and John Morris. 2010. The blank-page technique: Reinvigorating paper prototyping in usability testing. *IEEE Transactions on Professional Communication* 53, 2 (2010), 144–157.

[106] Lucy Suchman. 1993. Do categories have politics? The language/action perspective reconsidered. In *Proceedings of the Third European Conference on Computer-Supported Cooperative Work 13–17 September 1993, Milan, Italy ECSCW'93.* Springer, 1–14.

[107] Gareth Terry, Nikki Hayfield, Victoria Clarke, and Virginia Braun. 2017. Thematic analysis. *The SAGE handbook of qualitative research in psychology* 2 (2017), 17–37.

[108] Sian Townson. 2020. AI Can Make Bank Loans More Fair. *Harvard Business Review* (6 November 2020). Retrieved 07-Sept-2021 from https://hbr.org/2020/11/ai-can-make-bank-loans-more-fair

[109] Johanne R Trippas, Damiano Spina, Falk Scholer, Ahmed Hassan Awadallah, Peter Bailey, Paul N Bennett, Ryen W White, Jonathan Liono, Yongli Ren, Flora D Salim, et al. 2019. Learning about work tasks to inform intelligent assistant design. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval.* 5–14.

[110] Lav R Varshney. 2020. Respect for Human Autonomy in Recommender Systems. *arXiv preprint arXiv:2009.02603* (2020).

[111] Dakuo Wang, Justin D Weisz, Michael Muller, Parikshit Ram, Werner Geyer, Casey Dugan, Yla Tausczik, Horst Samulowitz, and Alexander Gray. 2019. Human-AI collaboration in data science: Exploring data scientists' perceptions of automated AI. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–24.

[112] Jonas Wanner, Lukas-Valentin Herm, Kai Heinrich, Christian Janiesch, and Patrick Zschech. 2020. White, Grey, Black: Effects of XAI Augmentation on the Confidence in AI-based Decision Support Systems.. In *ICIS.*

[113] Justin D Weisz, Michael Muller, Stephanie Houde, John Richards, Steven I Ross, Fernando Martinez, Mayank Agarwal, and Kartik Talamadupula. 2021. Perfection Not Required? Human-AI Partnerships in Code Translation. In *26th International Conference on Intelligent User Interfaces.* 402–412.

[114] Steve Whittaker. 2005. Supporting collaborative task management in e-mail. *Human–Computer Interaction* 20, 1-2 (2005), 49–88.

[115] Steve Whittaker, Victoria Bellotti, and Jacek Gwizdka. 2006. Email in personal information management. *Commun. ACM* 49, 1 (2006), 68–73.

[116] Steve Whittaker, Victoria Bellotti, and Jacek Gwizdka. 2007. Email and PIM: Problems and possibilities. *Commun. ACM* (2007).

[117] Wikipedia contributors. 2021. Robotic process automation — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Robotic_process_automation&oldid=1041208310. [Online; accessed 7-September-2021].

[118] Anbang Xu, Zhe Liu, Yufan Guo, Vibha Sinha, and Rama Akkiraju. 2017. A new chatbot for customer service on social media. In *Proceedings of the 2017 CHI conference on human factors in computing systems.* 3506–3510.

[119] Ping Zhang and U Yeliz Eseryel. 2005. Task in HCI research in the management information systems (MIS) literature: A critical survey. In *Proceedings of the 11th International Conference on Human-Computer Interaction.* Citeseer.

A CO-DESIGNING FOR PROCESS CONSEQUENCE

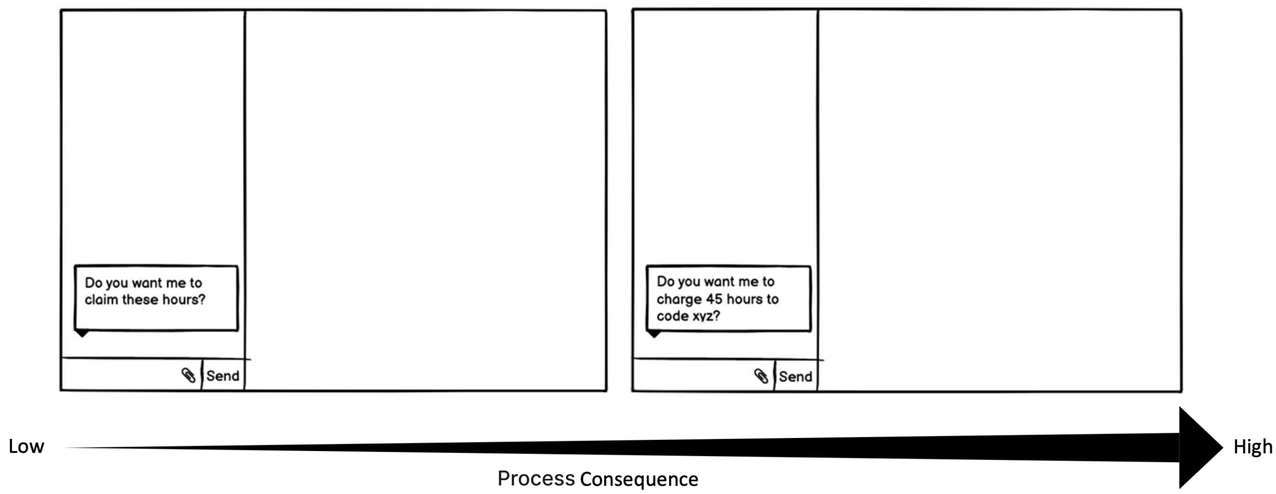

Figure 1: P203's designs for low vs. high process consequence. P203 designed two versions of a confirmation message for an automated time-sheet task. In a high consequence scenario (right), they expected the message to include specific details such as the number of hours and charging code. In a low consequence scenario (left), a less detailed confirmation sufficed. Alt-text: Two co-designed wireframes representing high and low process consequence scenarios. The left design is for a low process consequence version of the task and includes a message from the assistant in the CUI portion of the UI that says "Do you want me to claim these hours?" The right design is for a high process consequence scenario and has the same layout as the left design, but the message says "Do you want me to charge 45 hours to code xyz?".

B CO-DESIGNING FOR SOCIAL CONSEQUENCE

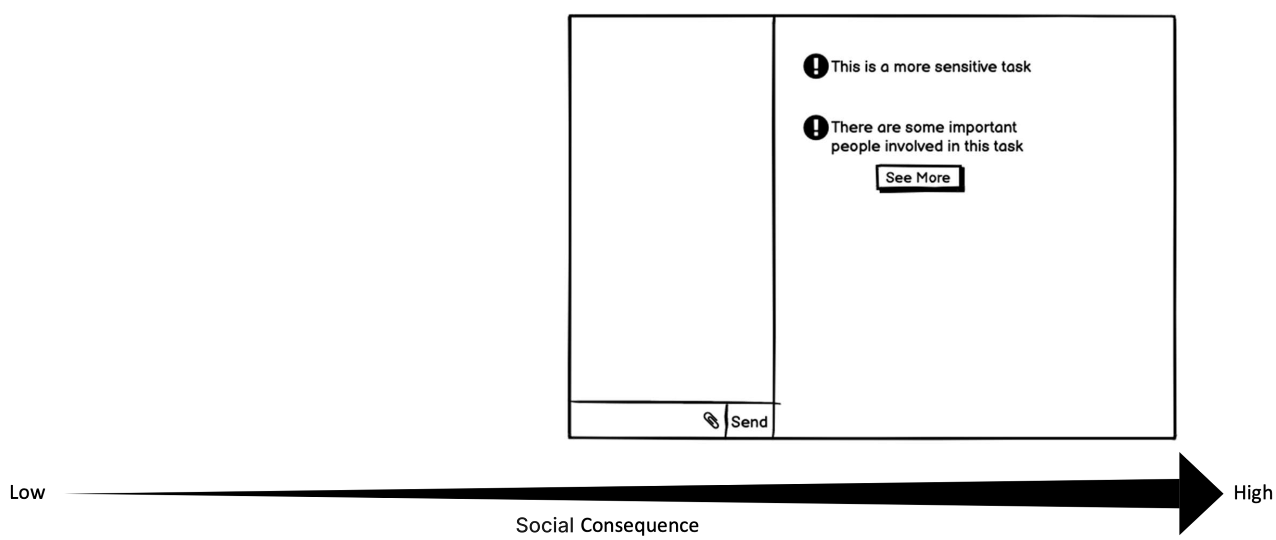

Figure 2: P215's designs for a task with high social consequence. P215 designed an example of a notification message that would give them more transparency into who is involved in an automated task, alerting them of the possible social risks associated with it. Alt-text: A co-designed wireframe on the high end of the social consequence spectrum, showing a notification with an exclamation icon and text that says "There are some important people involved in this task."

C CO-DESIGNING FOR TASK FAMILIARITY

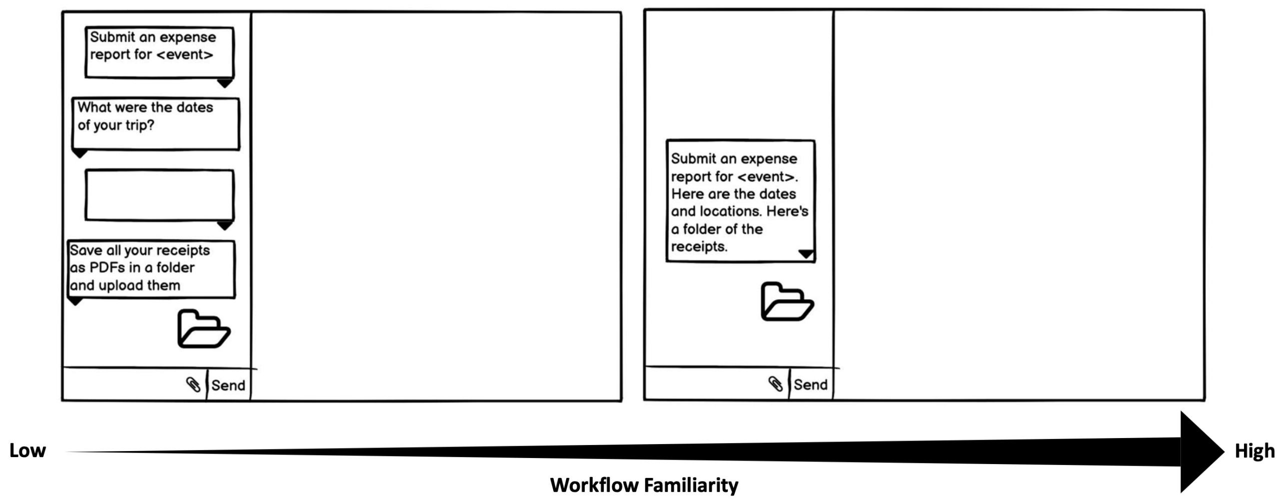

Figure 3: P212's designs for low vs. high familiarity scenarios. P212 co-designed interactions for working with the assistant on an expense approval task. When they are new to the task (left), they hoped the assistant could provide guidance by requesting inputs one by one. Once they become familiar with the task (right), they wanted to task support reduced so that they could initiate the task and provide all inputs in one step for greater efficiency. Alt-text: Two co-designed wireframes representing low and high familiarity scenarios. The left wireframe is for a low familiarity scenario and shows a conversation between the assistant and the user. At the top of the CUI, the user initiates the task with a message that says "Submit an expense report for <event>. Underneath this message, the assistant asks the user, "What were the dates of your trip?" There is a blank placeholder message bubble underneath from the user to represent their response. Below this placeholder, the assistant then instructs, "Save all your receipts as PDFs in a folder and upload them." There is a folder icon underneath this message to represent the user's uploaded files. The low familiarity scenario interaction ends here. The right wireframe is for a high scenario. In the CUI, the user sends a message that says, "Submit an expense report for <event>. Here are the dates and locations. Here's a folder of the receipts". There are is no guidance from the assistant.

D CO-DESIGNING FOR TASK COMPLEXITY

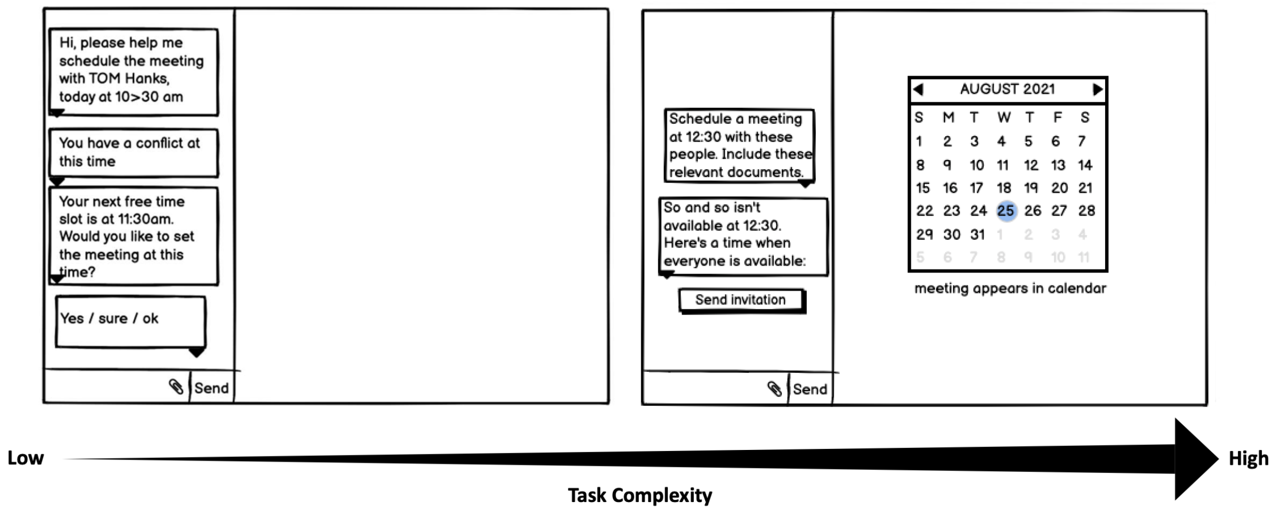

Figure 4: Two participants' designs for low vs. high task complexity scenarios. On the left, P203 designed a conversational interaction to schedule a meeting when the task is simple (i.e. the date, time, and attendees are known). When a meeting scheduling task becomes more complicated (i.e. someone has a time conflict), P217 showed that they wanted to open a calendar in the GUI to more efficiently view availability and select a new time (right). Alt-text: Two co-designed wireframes representing low and high task complexity scenarios. The wireframe on the left is for a low complexity meeting scheduling task and has all interactions in the CUI, showing a seies of messages between the assistant in the user to find a time to meet with one other person. The wireframe on the right is for a higher complexity meeting scheduling task where one person isn't available at the proposed meeting time. The assistant says "Here's a time when everyone is available" and displays a calendar with a date highlighted in the GUI.