Informing Group Informatics System Design: Balancing the Benefits and Concerns of Data-Driven Collaboration Feedback

Ryan Chan*
ryan.chan@uwaterloo.ca
University of Waterloo
Waterloo, Ontario, Canada

Devon Kisob*
dkisob@uwaterloo.ca
University of Waterloo
Waterloo, Ontario, Canada

Sharon Ferguson
sharon.ferguson@uwaterloo.ca
University of Waterloo
Waterloo, Ontario, Canada

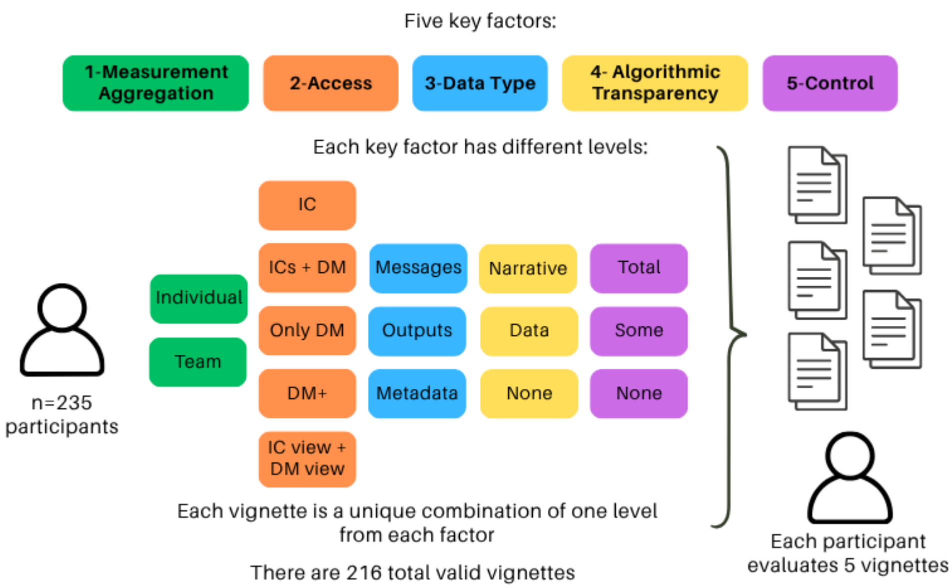

Figure 1: The comprehensive study design which included five key factors with their own levels. A total of n=216 valid vignettes, where each of the n=235 participants answered five. IC = Indivdual Contributor, and DM = Direct Manager.

Abstract

Remote work has increased reliance on online collaboration platforms, which has enabled more data-driven feedback on workplace collaboration. However, this data collection raises surveillance concerns, prompting users to weigh privacy losses against team benefits. This balance becomes asymmetric when benefits to individuals may limit benefits to teams. Studying this asymmetry, we investigate the trade-off between the benefits and concerns of *group informatics systems* that measure collaboration effectiveness from digital traces. This paper employs a survey with 216 vignettes that evaluate how five hypothetical design factors (measurement aggregation, access, data type, algorithmic transparency, and control) affect this trade-off. We find that design factors influence intention to use primarily through their effect on benefits and concerns; the relative importance of design factors shifts in group contexts; and we identify five design tensions that need to be managed in group informatics systems. We provide exploratory instantiations of contextual integrity and privacy calculus in group settings.

CCS Concepts

• **Human-centered computing → Empirical studies in collaborative and social computing**; *Collaborative and social computing systems and tools.*

Keywords

Group Informatics, Collaboration Feedback, Digital Trace Data, Benefits and Concerns, Privacy Calculus, Contextual Integrity

ACM Reference Format:
Ryan Chan, Devon Kisob, and Sharon Ferguson. 2026. Informing Group Informatics System Design: Balancing the Benefits and Concerns of Data-Driven Collaboration Feedback. In *Proceedings of the 5th Annual Symposium on Human-Computer Interaction for Work (CHIWORK '26), June 22–25, 2026, Linz, Austria.* ACM, New York, NY, USA, 25 pages. https://doi.org/10.1145/3808045.3808075

*Both authors contributed equally to this research.

CHIWORK '26, Linz, Austria

ACM ISBN 979-8-4007-2429-9/2026/06
https://doi.org/10.1145/3808045.3808075

1 Introduction

Collaboration is a pivotal component of modern knowledge work, as workers are in twice as many collaborative teams as they were five years ago [123]. Today, employees collaborate across time and space, relying on in-person communication, virtual collaboration platforms [48, 137, 138], email, shared documents, and project management tools. With collaboration fragmented across mediums and subgroups, collaboration effectiveness is impacted, which can, in turn, influence team performance [41, 93, 97]. Many collaboration characteristics important for team success, such as psychological safety [40], are comprised of collections of behaviours (e.g., mutual respect, admitting to mistakes, sharing wild ideas [100]), which can be hard for individuals to mentally keep track of.

To monitor and present patterns in behaviour, the field of personal informatics (PI) uses technology to collect relevant information for self-reflection [79]. For example, Di Lascio et al. [33] utilized a multi-sensor system to track when knowledge workers take breaks, and Epstein et al. [44] integrated break-taking data into reflection systems. We posit that there is an opportunity to investigate how this concept might extend to teams, particularly in the workplace—referred to here as *group informatics (GI)*. We define group informatics as systems that present metrics and/or summaries of important characteristics of team collaboration, which are reported at a group level or reported for group reflection. These systems may improve collaboration effectiveness—past work has shown that when team members are self-aware of their contributions to collaboration, team performance improves [34].

HCI is just beginning to explore these systems. For example, researchers have used surveys to understand which collaborative metrics would be of interest [84], or have made initial efforts toward designing algorithms [49, 100]. Other work has used design fictions to imagine the challenges that may arise in implementing GI systems in a workplace [76]. Yet the proposed design of these systems lags behind, with a noted lack of empirical studies [54]. While researchers have developed real-time systems that use trace data to help teams stay on track during meetings [25], HCI researchers have yet to study how we might design systems that report on key collaboration constructs. At the same time, simple GI systems are actively entering the marketplace. For example, Microsoft Viva's Insight functionality claims to "analyze collaboration, meetings, wellbeing, and coaching trends to uncover improvements" using data from Microsoft products, and the Pulse feature aggregates employee surveys to report team health [96]. These systems may present concerns of workplace surveillance and resulting harms to career progression [90] or performance management [98].

In general, researchers have described a trade-off [9, 51, 132] between the personalized benefits and the privacy violations arising from data collection, showing that these benefits and drawbacks are more complex than simple binaries. Individuals may undergo a "privacy calculus" to determine how comfortable they are with sharing information based on what they will receive in return [61]. This calculus becomes more challenging in group settings, where benefits are often collective (team efficiency) but risks are individual (personal privacy violations). For instance, requirements for individualized consent [27] mean that opting out can serve individual needs but harm groups, as it can degrade the accuracy of group measurements. Recent work has focused on an individual's privacy calculus, where the same person takes the risks and gets the benefits [75]; multi-party privacy, where there are conflicting privacy preferences but little consideration for benefits [125]; and workplace surveillance technology, which is used to monitor or evaluate employees, but does not often aim to provide benefit to the group [36, 61]. This leaves a critical gap in our understanding of how workers navigate privacy decisions when their personal privacy is at odds with their team's performance.

To design for this asymmetry, we need to consider privacy's contextual nature. In the design of GI systems, these contextual parameters can be concretized into design choices, and past work has shown that these choices influence perceptions [57, 90, 98, 134]. These design features can reconfigure the asymmetry—for instance, displaying group-level outputs makes those who opt out invisible. Privacy perceptions may depend on design choices such as how the data are aggregated [127], who has access to the data [134], what data are collected [60, 134], whether data collection and analysis is transparent [6, 56], and whether users have the chance to provide non-coerced consent [27].

While past work shows that users weigh privacy risks and benefits when deciding to use a technology, and contextual integrity [101] shows how context matters, we lack an understanding of how concrete design choices in the context of GI systems influence this privacy calculus in an asymmetric scenario. To explore this gap, we conducted a pilot study (32 vignettes, n=106) and a comprehensive study (216 vignettes, n=235), systematically demonstrating the effects of five hypothetical design characteristics of GIs (measurement aggregation, access, data types, algorithmic transparency, and control) on perceptions of benefits, concerns, and intentions to use.

Specifically, we ask: **(RQ1) How do design characteristics of GI systems affect workers' perceptions of the trade-off between concerns, benefits, and intention to use?** We revisit these findings in the context of a complete GI system to ask: **(RQ2) What design tensions emerge from these trade-off perceptions?**

We found that design factors influenced intention to use by affecting perceived benefits and concerns, validating the privacy calculus framing of this problem, even when benefits are collective. Specifically, participants noted that GI design characteristics, such as displaying data individually, data transparency, and opt-out control, would make the system hypothetically more beneficial, whereas restricting access to only leaders was perceived as concerning. The relative importance of design factors in this asymmetric context does not always mirror that in individual contexts [134], with data type and access levels showing minimal effects on concerns and benefits. Instead, design choices that instantiate transmission principles (rules relating to how data flows [101]) emerged as most influential. Evaluating the concerns and benefits of whole GI systems, we identified five initial design tensions that must be thoughtfully managed. We identify tensions between individual agency and group accuracy; personal utility and team cohesion; system transparency and interpersonal trust; motivation and surveillance; and algorithmic consistency and contextual fairness, reflecting the fundamental asymmetry in which design must balance individual privacy protection with collective benefit.

We contribute 1) an initial baseline understanding of user concerns for an emerging, asymmetric technology, GI, identifying design factors that may affect adoption, in the absence of real-world examples. 2) We identify which hypothetical contextual design features influence perceptions of collective benefit and individual risk, building on the idea of privacy calculus in an asymmetric setting. 3) We contribute five preliminary design tensions, offering concrete framings for the trade-offs between individuals and teams, as well as early design implications for easing these tensions.

2 Background

In this section, we review the balance of risks and benefits in technology adoption, and focus on asymmetric scenarios such as workplace surveillance. We build on privacy calculus, by studying this trade-off in groups. We also operationalize contextual integrity, showing how context manifests in design choices in GI systems. We end with an overview of common informatics design choices.

2.1 Personal and Group Informatics

PI systems exist in various forms to collect, summarize, and present individual-level data to promote reflection and behaviour change. In a review of PI in the workplace, Kersten-van Dijk et al. [71] classified these systems into two categories: systems for self-monitoring and reflection, and systems that provide actionable feedback. PI-enabled reflection is beneficial. For instance, Sefidgar et al. [113] found that reflecting on personal data improved self-awareness, which was crucial for achieving work-life balance. PI systems can also provide users with contextual insights, suggestions, and feedback. Jung and Lee [68] employs counterfactual explanations in PI systems to suggest contextual changes when the user is coping with stress.

While the predominant focus of informatics literature is still providing feedback to a single individual, some studies have started to consider technology to support reflection in pairs [78, 139], families [77], and ad hoc, online teams [105]. This informatics literature still primarily focuses on summarizing individual data, even if reflected on in groups. Yet, GI systems could also be designed to summarize and share information about group-level attributes, such as psychological safety or shared understanding. Moving to a group level introduces asymmetries in which individual decisions can affect the benefits of entire groups, in terms of perceptions of surveillance, alignment, and privacy beliefs.

2.2 Concerns with GI Systems

As with any underexplored area of research, many open-ended questions arise. Challenges from PI systems may extend to GI [11]: such as alignment [126], where misalignment between team members occurs when they differ on what "optimal" entails for a specific metric [131]. GI metrics may also be used for unintended purposes, such as to prevent career progressions [90], or for evaluation and performance management [98]. Workers are also concerned with how different data types (e.g., meeting data) will be shared with others [126]. Any productivity-related systems have implications for trust between employees and the organization [90]. Data collection can trigger the Hawthorne Effect, in which workers behave differently because they know they are being monitored [31]. Monitoring systems may also create workplace competition [70, 126].

2.2.1 *Workplace Surveillance.*

Concerns with GI systems become outsized when we consider the power imbalances in a workplace. Any workplace data collection can lead to perceptions of workplace surveillance [28]. For example, Oz et al. [104] identified a disconnect between supervisors and non-supervisors and their goals with surveillance systems, resulting in tensions. Surveillance perceptions can also be influenced by the type of data collected. For example, Corvite et al. [29] studied workers' perceptions of the benefits and risks of emotional AI in the workplace, finding a variety of risks (e.g., harm to their wellbeing, harm to the most marginalized) that overshadowed potential benefits. Researchers have identified a gap between the intentions of workplace tracking programs and their actual impact on individuals, suggesting that merely incentivizing something does not guarantee meaningful progress. For example, Adler et al. [3] studied perceptions of hypothetical stress-sensing technologies for resident physicians, finding that participants were concerned that they would be used to reinforce power asymmetries.

Workplace surveillance technologies generally do not aim to benefit those they surveil—instead, they benefit organization leaders. The asymmetry in GI systems means that although the benefits are intended to be collective, surveillance concerns may be perceived differently across individuals, leading to varying acceptance.

2.3 Privacy Frameworks

We review three privacy frameworks that may influence the benefit-risk balance that workers' perceive in GI systems.

2.3.1 *Privacy Calculus.*

When considering any form of data collection, researchers describe a trade-off that users reason through [2, 9, 51, 132] by balancing a loss of privacy with the potential benefits of personalization—solving the 'calculus' of whether a system is worth using. For example, Seberger et al. [112] examined the tension between well-being apps that propose better living and the negative affective implications due to surveillance. This calculus can be nuanced and contradictory, as tracking apps may be simultaneously perceived as "creepy" yet still beneficial [112]. This highlights asymmetry in the ongoing calculus, where the varying perceptions from individual to individual in differing situations [82] is compounded when the benefit is collective. Asymmetry can also mean that people may not always act in their own self-interest, and that the trade-offs can be quite complex to weigh [2].

2.3.2 *Multi-Party Privacy.*

This asymmetry has been studied in social media research, yielding open challenges regarding multi-party privacy. Researchers have highlighted the complexity that arises when privacy involves multiple parties, as well as the conflicting privacy preferences of different individuals [10, 67, 124, 130]. Such and Criado [125] note a tension where having more fine-grained privacy approaches could make the end product less useful. Research in this domain has also taken a more computational perspective, such as proposing cryptographic protocols that preserve privacy for multi-party access control [114], group-based stochastic modelling to identify potential privacy conflicts [144], and introducing a mediator into existing protocols without compromising privacy or accuracy [116]. Yet, the multi-party privacy literature primarily focuses on conflicting preferences rather than asymmetric benefits.

2.3.3 Contextual Integrity. This privacy calculus that users undergo is also extremely contextual. Nissenbaum [101] introduced the theory of privacy as contextual integrity, which states that privacy perceptions are affected by the context from which the data is collected, the actors involved in the interaction, the attributes of the data, and the transmission principles (stipulations that shape/constrain the flow of information) [101]. This theory has been applied in HCI to understand the privacy needs of online groups [26], observe employee attitudes towards workplace surveillance [134], study comfort with using social media data for research [57], and understand privacy in extended reality [60]. Becker [9] argued that characterizing and detailing the trade-offs between benefits and concerns further strengthens the contextual integrity framework. Thus, context is crucial in understanding this asymmetry. In informing the design of GI systems, we can operationalize this context in concrete design choices. We now review design choices that are suggested to be important for privacy perceptions.

2.4 Design Choices for Informatics Systems

We summarize six common design choices from PI literature (data transparency, data aggregation, access levels, control, data type, and use cases) that inform our vignette designs.

A common design guideline is to increase transparency into how data is collected, aggregated, and analyzed [2, 32, 54]. For example, Das Swain and Saha [31] suggest that current large language models (LLMs) will be able to provide contextualized explanations for metrics. Meyer et al. [95] suggest that systems should provide quick, high-level statistics with the option to drill down into more detail when needed, and offer natural-language explanations in place of visuals. In this work, we investigate how the benefit-concern trade-off is evaluated when increased transparency potentially risks others' privacy rather than your own.

Researchers have also investigated how informatics should be displayed [4], particularly how data should be aggregated (e.g., by groups or individuals). For example, Mathur et al. [91] successfully mitigated surveillance concerns in an informatics system by anonymizing data sharing. Lushnikova et al. [83] also suggests that aggregating data at a high-level visualization could be a straightforward approach to making privacy-sensitive design choices. Here, we analyze the effect of two forms of aggregation: individual measurements for collaborative reflection and group-level measures.

Who has access to these measurements is another important design decision. Power asymmetry has been discussed in the literature, whether it takes the form of the generalized employer [62] or a specific managing individual [30]. Vitak and Zimmer [134] stratified the workplace into six levels, including one's direct team, direct supervisor, and senior executives. Their modelling found that participants were less comfortable sharing data with members of their direct team or with coworkers on an anonymous dashboard. We analyze how individuals decide on the appropriate level of access for an entire group.

Control over personal data in a workplace setting has been extensively discussed in the literature [2, 19, 20, 32, 38, 82, 104, 107]. The interview study by Roemmich et al. [107] showed that the issue was not necessarily that workers wanted to conceal information, but rather that they were losing autonomy to manage their own information. Participants noted that opting in was important, as they would be choosing to provide information. Thus, the degree of control individuals have over the adoption of different kinds of information systems affects their comfort. Given the asymmetry in which having more personal control may increase personal benefit but harm collective benefit, we evaluate this trade-off in our work.

The types of data being used in these systems also matter, with past work using textual information and individually-produced artifacts [16, 52], emails, chats, calls and video calls [16, 19], and metadata, like who emails whom and when [94]. The literature also investigates much more invasive types of data, including geolocation [94] and biometric and health data [107]. With a focus on workplace use, we analyze how communication, work products, or metadata affect perceived benefits and concerns.

Since PI systems can be used for both actionable feedback and self-reflection [55, 71, 108, 119, 131, 140], different scenarios should be considered when designing informatics systems. In self-reflection, designing opportunities for self-discovery [11] and deliberately slowing down [140] have been suggested. For collaborative use cases, the literature emphasizes conversational components [11], especially in a workplace [19]. Evaluation of employee performance is also a common scenario in the literature on informatics in the workplace [19, 29, 38]. While designers may not have control over all use cases of a tool, given the influence of use on system acceptance, we also investigate the role of use cases.

2.5 Summary

Despite the promise of GI systems, we lack an empirical understanding of how workers perceive the asymmetrical trade-offs inherent when data is either collected, shared, or displayed at a group level. Current models of privacy calculus assume symmetric concern-benefit trade-offs, and current design choices from PI focus on individual agency and benefit. Privacy perceptions are highly contextual. Toward the design of GI systems, we operationalize this contextual nature (that designers have control of) through concrete design choices. To elucidate the tensions in making concern-benefit calculations for group systems, we investigate how the design characteristics of GI systems affect users' perceptions of benefits, concerns, and intention to use.

3 Pilot Study

As GIs are still emerging [76, 84], we first conducted a pilot study to prioritize the factors that influence perceptions of benefits and concerns.

We employ a factorial vignette methodology. Factorial vignettes are a survey method that present short scenarios which are systematically varied to introduce different contextual factors [135]. Vignettes allowed us to vary key design choices for GI systems to understand how these high-level factors relate to concerns and benefits and, in turn, how they reveal tensions between individuals and groups. Vignettes additionally allow us to incorporate multiple, co-occurring factors that influence perceptions of privacy [101], blending survey and experimental design [57, 135]. Factorial vignettes have been used across HCI [7, 37, 65, 72, 73, 80], and are particularly heavily used in surveillance and privacy research [1, 5, 15, 22, 43, 57, 60, 63, 89, 99, 128, 134].

3.1 Methods

We overview our vignette design and survey measures, followed by participant recruitment and data analysis.

3.1.1 Vignette Design. We designed a set of vignettes accompanied by background information applicable to all hypothetical scenarios. Keeping vignette scenarios simple allowed us to study an emerging area where the technology is not yet fully realized (e.g., Lutz and Tamò-Larrieux [85]), without participants fixating on a specific implementation (or interface), and instead focusing on high-level design factors. By keeping scenarios broad, we encouraged participants to imagine the system in their own context [110], thereby exposing the underlying reasons for perceived concerns and benefits. We chose not to specify collaboration metrics (e.g., focusing on a system for shared understanding) because GI systems can be used for various elements of collaboration, with some being more or less important depending on one's work experience, or too complex to fully comprehend in a short survey.

Our vignette factors were selected to systematically test three core design decisions identified by prior research as critical to privacy perceptions in workplace informatics. We grounded our factor decisions in the theory of privacy as contextual integrity [87, 101] and past research on workplace informatics tools. Previous research has shown that anonymity and aggregation influence perception [132]. Aggregation directly affects the information type or data subject in contextual integrity [87, 101], as displaying individual data can raise privacy concerns compared to displaying team-level summaries. Hierarchy and access also influence comfort with informatics systems [104, 126, 134], directly aligning with the actor or recipient in the context of contextual integrity [87, 101]. Lastly, workplace surveillance literature indicates that workers are most concerned about the misuse of these systems [19, 38, 90, 107, 134], representing use cases, a component of contextual integrity that is missing from the original model [87]. We began our exploratory pilot study with these categories. Other factors (data type, transparency mechanisms) were held constant in the pilot (e.g., public data only) to avoid overwhelming the design space at the outset.

Each key factor includes different levels (e.g. "Aggregation Level: Individual"). The first and last authors iteratively brainstormed to identify these levels, informed by the workplace surveillance [104, 132, 134] and informatics literature [45, 76, 83]. Some levels were more straightforward than others, such as the individual vs. group-level measurements. A more difficult decision was the use-case key factor, where we collaboratively ideated a set of scenarios that covered both beneficial ways the tool could be used, and harmful uses discussed in prior work [19, 38, 90, 107, 134].

A summary of all factors and levels for the pilot vignettes is shown in Table 1. The length and level of detail of the vignettes are comparable with past vignette research [13, 134]. Among these possible combinations, we filtered out eight impossible combinations, as recommended by Jasso [66], such as displaying a team-wide measurement that was only visible to an individual contributor. After this filtering, we were left with 32 vignettes. To implement the vignettes in Qualtrics, we modified a Python script [92].

3.1.2 Survey Design. Participants provided consent, were shown the instructions and background statement (Appendix A), and were told that it applied to all vignettes. Participants were required to pass a comprehension check based on the background information. Participants then saw five of the 32 vignettes, randomly assigned, with a standard Prolific attention check after the third vignette. We also collected standard demographic information.

3.1.3 Measures. Due to the evidence of "privacy calculus" [9, 30, 51, 61, 112, 132, 143], we hypothesize that perceived concern and benefit will be mediating variables of intention to use (dependent variable). This means that in each vignette, participants will evaluate the perceived benefits and concerns of this technology, and whether they intend to use it.

Dependent and Mediating Variables: We created multi-item scales for perceived benefits and concerns, based on similar scales [53, 133]. The goal was to capture a breadth of concerns and benefits that participants may have in mind when evaluating the hypothetical system. To address perceived concerns, we adapted a workplace surveillance scale from Furnham and Swami [53], selecting five items relevant to GI systems. For perceived benefits, we adapted the "Perceived Usefulness" section of Venkatesh and Davis [133], taking all four of their original items. The items in both scales were adapted to better reflect GI. All questions were measured on a seven-point Likert scale (strongly disagree/strongly agree - Appendix C). The pilot data yielded Cronbach's alphas of 0.85 for concern and 0.88 for benefit. Following best practices, we included open-ended questions [8], asking about additional concerns or benefits.

The dependent variable, intention to use, was a single statement ("Given my ratings of the statements above, I would want to use this tool in my team.") from Venkatesh and Davis [133], and measured on a seven-point Likert scale. The order of the statements for perceived benefits and concerns was randomized for each participant, with the intention to use statement always shown last.

Controls: We controlled for age, managerial level, gender, race, country of residence, and years worked with their current team. Participants were also asked to fill out a privacy segmentation scale, suggested by Kumaraguru and Cranor [74] and based on Westin [136], to capture their baseline privacy disposition.

3.1.4 Participants and Recruitment. We recruited participants through Prolific, filtering for English fluency, employment status (full-time or part-time), and an approval rate of >99%. Each participant was paid £3.35 (GBP), with a median completion time of 27 minutes (approximately £7.50 per hour). We aimed to recruit 100 valid participants, with a final sample of $n = 106$. The study received clearance through the University of Waterloo Research Ethics Board (#47337).

The pilot study participants were evenly split between men and women (47.2% each), with the majority in the 25-34 age range (44.3%) (Appendix B). Participants had primarily 1–3 years or 10+ years of experience (30.2% each). The majority of our participants were Black (63.2%), and predominantly managers, including both one-level and multi-level managers (78.3% total). Some samples of these job titles include Account Manager, Project Manager, Operations Manager, and IT Manager.

3.1.5 Data Analysis. We removed participants who failed the comprehension check or the attention check ($n = 8$), did not complete the privacy segmentation scale ($n = 3$), or identified as unemployed ($n = 1$), leaving us with $n = 106$ responses from the original $n = 118$.

Table 1: Overview of Pilot Study Vignette Factors

Factor	Levels	Description in Vignette
Aggregation	Individual Measurement	"an individual, personalized measurement"
	Team Measurement	"a single, team-wide measurement for the whole team"
Access	Only the individual contributor (IC)	"by each individual, to see only their own measurements"
	Only direct manager (DM) and their upwards reporting managers	"only by your direct manager and their upwards reporting managers (including executives)"
	All ICs + DM on the team	"by all of your team members and your direct manager"
	Only DM	"only by your direct manager, but not by you and your team"
	Everyone in the organization	"by everyone in your company"
Scenarios	Investigating a specific collaborative instance	"to look up a specific collaboration situation, such as to understand why the last meeting ended poorly"
	Scheduled team check-ins	"to check-in on the team regularly, such as during regular reviews after projects conclude"
	External evaluation	"for external evaluations, such as performance reviews or project reviews"
	Individual reflections (not within the team)	"for you to individually reflect on and decide how you can help your team improve"

As each participant rated five vignettes, each vignette received 16 ratings, more than past work [134].

Our data analysis plan consisted of three steps: first, we used a linear mixed model (LMM) to model how each key factor influenced perceived benefit and perceived concern, controlling for random participant effects (since each participant rated five vignettes), similar to the analysis in Hadan et al. [60] and Vitak and Zimmer [134]. Second, we used a cumulative link mixed model (CLMM) to predict intention to use based on key factors, perceived concerns and perceived benefits. To show mediation, we used bootstrapping. These last two steps are described in the Comprehensive Study section.

We determined the baseline for each factor based on what we perceive as the most "common" use of GI systems. For measurement aggregation, the baseline was team measurement, since we are interested in groups. For access, our baseline was all individual contributors and their managers, since the typical use case would be within a team (where the entire team has access). For the scenario, we used scheduled team check-ins because we believe this would provide the greatest benefit during the team's day-to-day activities.

3.2 Results

The regression tables (Table 9 and Table 10) are shown in Appendix D. The concern model (marginal R^2 = .18, conditional R^2 = .60) revealed that only individual access was associated with less concern (p<0.001). Restricting access to ICs significantly reduced concerns, suggesting that participants view horizontal data access as fundamentally safer than hierarchical access. Other design factors did not show strong effects in this initial model, indicating that while access matters, other, missing design levers might be dominating user perceptions. The benefits prediction model (marginal R^2 = .15, conditional R^2 = .50) similarly revealed that aggregation, access, and scenario do not dominate user benefit perceptions as currently measured. Together, both models suggest that demographics such as race and age may be related to concern and benefit perceptions, possibly reflecting varying expectations of privacy.

While the pilot confirmed that access levels influence concern, the lack of broader significance across other factors suggested our model was underspecified. Thus, we did not complete the second and third analysis steps. The lack of findings motivated the team to revisit the design of the key factors and dependent measures, specifically by analyzing the open-ended responses from the pilot study, as described in the following section.

4 Comprehensive Study

We first describe how we revised the vignette survey, describe the new survey, and present our results.

4.1 Revising the Study Design

To revise our study design, we conducted open coding [109] on the open-ended responses from the pilot dataset. We coded all mentions of specific concerns and benefits (to inform our scale measure), and all mentions of the system features that affect participant perceptions (to inform our key factor design).

4.1.1 Revisiting Key Factors. Participants described several characteristics of hypothetical GI systems that influence their intention to use them. They confirmed that the measurement aggregation (information type/data subject equivalent [87, 101]) influenced perceptions of concerns and benefits, some stating that team-wide measurements provide *"less risk of individual team members feeling scrutinized,"* while others claimed displaying IC performance *"could be helpful to balance the workload amongst team members."* Access (recipient or actor equivalent [87, 101]) also factored into participants' decision making, with participants concerned with members having *"free access to your manager's numbers/activities"* and seeing benefit in *"..this tool...[which] can be accessed by each individual in the team."* Participants were hesitant to assert any specific benefits or concerns tied to specific scenarios (purpose or use equivalent [87, 101]). Adoption was dependent on conditions, such as *"if it's part of a broader feedback system that values both data and human judgment,"* claimed some respondents, while others proposed their own desired scenario, such as *"a summary of tasks*

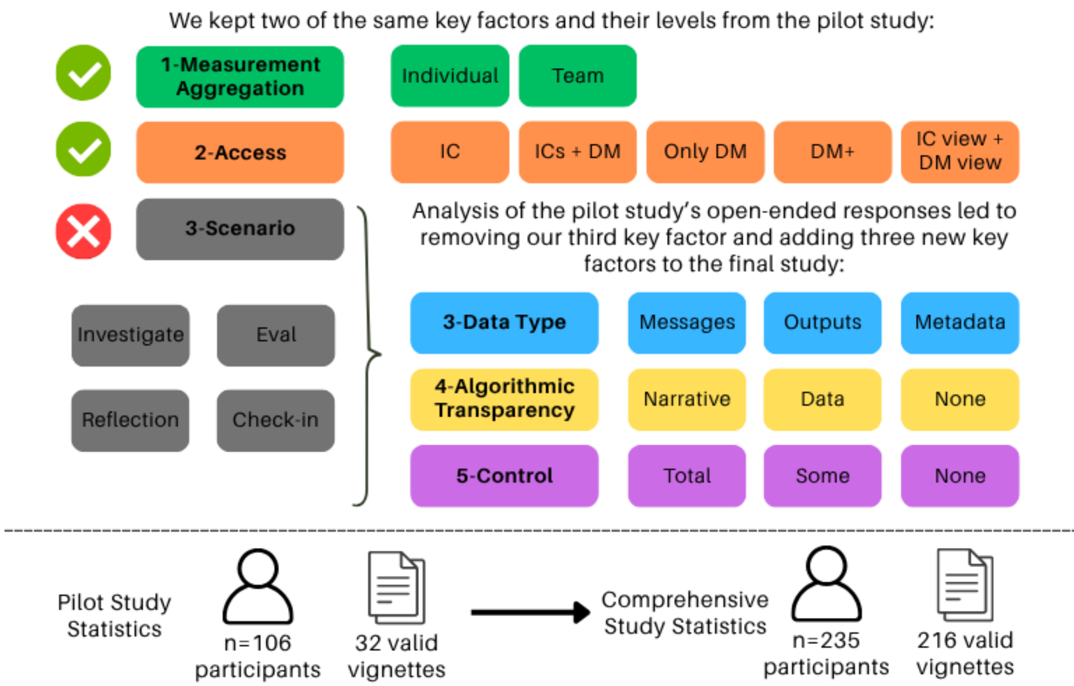

Figure 2: We kept the first two key factors and their levels, removed the third factor from the pilot, and identified three new factors for the comprehensive study. We also increased the number of participants (n=106 to n=235) and total valid vignettes (n=32 to n=216) from the pilot to the comprehensive study. IC = Individual Contributor, DM = Direct Manager

and information shared...to have all collaboration information in one place and show who is responsible for which task." We acknowledge that the ultimate use of the tool is largely outside the control of the technology designer [22, 81]. While studying the intended and actual use of the technology is a critical research step, we believe this evaluation is better suited once a more concrete design for GI systems is developed [19, 76].

Numerous participants raised concerns about algorithmic transparency, data type, and opt-out control. Participants discussed how data is processed and collected. As one respondent stated, concerns that they *"have no clue about how it does its appraisal"* highlight the salience of algorithmic transparency. Studies have shown that users are more comfortable with AI technology when they understand what the algorithms are doing [30, 62]. Algorithmic transparency is also an example of a transmission principle in contextual integrity, as described by Malkin [87], as it is a rule that dictates data can only be shared if it is appropriately explained. Participants repeatedly described how data types influence their perceptions. Some explicitly stated they *"would worry about the type of information that it really collects,"* while others mentioned specific data they would be concerned about, such as *"private things like email."* This factor maps directly to the "data type" parameter in contextual integrity [87], and was not accounted for in our pilot vignettes. Data type has been shown to influence perceptions of privacy and surveillance in prior empirical work, as some data types are considered to be more private than others [16, 19, 52, 94]. Others expressed strong concerns that the tool involved *"private information being collected without consent,"* indicating the importance of consent and opt-out control. As explicitly listed in Malkin [87], consent is an example of a transmission principle within contextual integrity. Much HCI work has stressed the importance of providing robust consent mechanisms when personal data is collected [19, 32, 38, 104, 107].

Taken together, these insights motivated a refocusing of our vignette design to better align with our research questions and with the contextual integrity framework as it applies to GI systems. Specifically, we removed scenarios as a key factor and introduced three new key factors: data type, algorithmic transparency, and opt-out control, which more accurately reflected the transmission principles and information sensitivities that participants described. A summary of the changes is shown in Table 2.

4.1.2 Revisiting Measures. Our open-ended responses identified perceived benefits and concerns that largely aligned with our measures. There were several comments about the tools' potential to increase efficiency, performance, and productive collaboration: *"...increase likelihood of fun light-hearted competition..".* Meanwhile, fear that the tool *"can cause misunderstandings between team members"* is an example of a misuse captured by our initial measures. There were also several reported benefits and concerns that our multi-item measure did not capture. One participant wrote that these systems *"can help with accountability, goal setting, and identifying areas for improvement,".* Due to the breadth of open-ended responses, we determined that the diversity of benefit and concern themes would make it difficult to develop comprehensive multi-item scales. Therefore, following precedent from past vignette studies

[1, 5, 22, 57, 60, 63, 89, 99, 128, 134], we decided to reduce the benefit and concerns measures to single items, each, which allowed for broad interpretation. This allows participants to reach conclusions organically about perceived benefits and concerns, while also reducing respondent fatigue.

To assess the validity of switching to single-item measures, we examined predictive validity as a proxy for convergent validity (an established approach for validating single-item measures [12]). Both concerns and benefits correlated strongly with the intention to use question ($r = -0.34, p > 0.001$ for concerns, and $r = 0.74, p < 0.001$ for benefits). In the comprehensive study, the single-item concern measure showed a stronger correlation with intention to use ($r = -0.56, p > 0.001$) as did the single-item benefit measure ($r = 0.82, p < 0.001$). This consistency in direction and magnitude of correlations across both measurement approaches, combined with precedent from past vignette studies [57, 89, 134], supports the use of single-item measures. As correlations are stronger in the comprehensive study, this may reflect the broader single-item framing better capturing the full range of concerns and benefits.

4.2 Methods

Much of the study design remains similar to the pilot; thus, in this section, we only describe key differences.

4.2.1 Vignette Design. The pilot study's open-ended responses and additional literature informed the key factor changes and their corresponding levels. We designed the levels to reflect what could easily be used in GI systems, particularly in workplace contexts. In the case of data types with varying levels of privacy, we defined the following based on Section 2.4: team messages and meeting transcripts (most personal) [16, 19], completed project outputs [16, 52], and metadata about activity (least personal) [94]. For algorithmic transparency, we considered a range from least to most transparent. Drawing from examples that already exist in the workplace context, such as AI summaries of meetings [96, 145], this led us to include: no transparency, a narrative description (explanation) [31], or direct links to important input/source data (e.g., "the score is lower this week due to Wednesday's meeting, see transcript here.") [95]. For control, we defined: no opt-out control, some opt-out control, and total opt-out control [19, 38, 82, 104, 107]. Lastly, we removed the "entire organization" level of Access, as all open-ended responses in the pilot highlighted that it would be unrealistic and untrustworthy for the entire organization to access a specific team's dashboard. Instead, participants brought up the benefit of both having individual access, and for the manager to have access to data about their team. As such, we combined these two options into a single level, where managers can access a team dashboard, and each individual can access their personal dashboard. This is in line with the literature in Section 2.4 [30, 62, 134].

Table 2 summarizes the key factors and levels included in the final study. Team-level display is incompatible with the two levels of the access factor that are personalized to individuals. Thus, 54 vignettes were eliminated, leaving 216 for our comprehensive study.

4.2.2 Survey Design. The general survey design remained identical to the pilot study. Due to the additional key factors, we reduced the background information, as more variables were varied rather than fixed (Appendix E). The same comprehension checks and attention checks were used, and participants still rated five randomly selected vignettes with an open-ended question after each one.

4.2.3 Measures. We followed past vignette studies [57, 60, 134] in using a single-item measure for each concern and benefit: *I imagine that this tool could have negative consequences for my team.*, and *I imagine that this tool could positively benefit my team*, measured on a seven-point Likert scale. We kept the same intention to use measure and control variables.

4.2.4 Participant Recruitment and Demographics. We recruited participants through Prolific, adding to the existing filters a filter for at least a community college education level, to scope our sample to those who may be more likely to be knowledge workers, where virtual collaboration tools and teamwork may be more common. Each participant was paid £2.67 (fewer questions reduced the completion time to 23 minutes, resulting in approximately £7/hour). We aimed to recruit at least five ratings per vignette, requiring 216 participants. After filtering participants via the exclusion criteria outlined in the pilot study, we were left with 235 participants.

Our comprehensive study had slightly more women (50.6%) than men (47.2%). Similarly to our pilot survey, most participants were aged 25-34 (42.2%), Black (57.4%), and managers at some level (72.7% total). Respondent team experience was mostly dispersed across 1-3 (29.4%), 3-5 (23.0%), and 5-10 (37.4%) years (Table 3).

4.2.5 Data Analysis. The data analysis plan was largely similar to the pilot study. With n=985 valid vignettes, we used CLMMs to predict perceived concerns, benefits, and intention to use. As benefits and concerns were now measured using single-item ordinal measures, a CLMM was more appropriate than the LMMs used in the pilot study. The proportional odds assumption underlying CLMMs was assessed using Brant tests [18] on equivalent fixed-effects models. Vignette key factors satisfied the assumption in all but one case (narrative transparency in the benefits model), with violations concentrated in sparse control cells (Appendix F).

The intention to use model retains the same multi-level structure as the benefits and concerns model, including participant fixed effects, and all the same controls. We set the baseline for some controls (e.g., years worked, age, manager level) equal to the most populous group, as suggested in Field et al. [50]. To account for demographic variation while avoiding sparse-category bias, we included binary indicator variables for focal demographic groups with sufficient sample size (e.g., race and country categories). Each indicator contrasts membership in the focal group with all other participants. This choice aligns with best practices for collecting demographic data as it avoids collapsing small groups into an "other" category, and does not present results for small sample sizes [21, 102]. A participant may contribute to multiple indicators.

To assess mediation effects, we first used a likelihood-ratio test to determine whether adding mediators (perceived concern and benefit) significantly improved model fit in the intention-to-use model. We then we used non-parametric bootstrapping [86] to determine whether the vignette factors influenced intention to use exclusively by influencing perceived benefits and concerns. The mediation model is shown in Figure 3.

Table 2: Overview of comprehensive study vignette factors

Factor	Levels	Description in Vignette
Measurement Aggregation	Individual Measurement	"a labelled score for each team member"
	Team Measurement	"one score for your whole team"
Access	Each individual only sees themselves	"only yourself, to see your own measurements"
	Manager and upwards reporting line and their upwards reporting managers	"only your direct manager and their upwards reporting managers (including executives)"
	All ICs and the DM	"all of your team members and your direct manager"
	Only DM	"only your direct manager, not by you or any of your team members"
	IC sees self; manager sees everyone	"your manager, although you can also see an individualized dashboard based on your own data"
Data Type	Messages and meeting transcripts	"messages and meeting transcripts"
	Completed project outputs	"completed project outputs like documents or code"
	Metadata about your activity	"activity timing and patterns (e.g., when you sent a message, but not what you said)"
Algorithmic Transparency	None	"does not provide explanations for the values that are shown"
	A narrative description	"provides a textual explanation of where the values came from"
	Pointing to important input/source data	"links to raw data, so that you can see the messages, transcripts, work products and/or metadata that the value refers to"
Control	Total opt out control	"able to opt out of all data collection"
	Some opt out control	"able to opt out of some data collection but not all"
	No opt out control	"not able to opt out of any data collection"

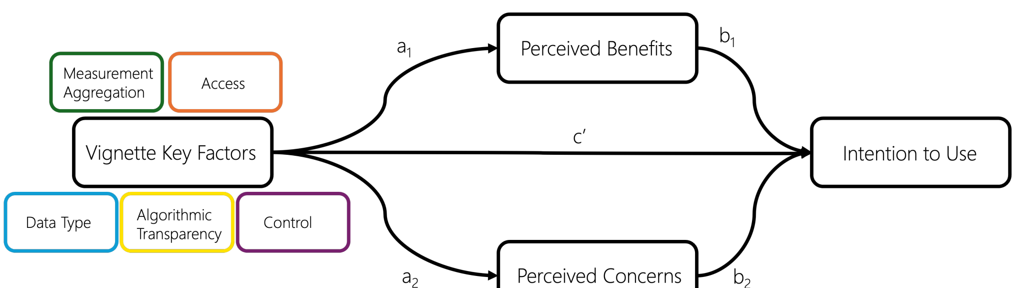

Figure 3: Mediation model. a_1, a_2 **represent the effect of key factors on mediators,** b_1, b_2 **represent the effect of mediators on intention to use,** c' **represents the direct effect of key factors on intention to use, controlling for the mediators.**

For the open-ended responses, we conducted open coding [109] to identify specific concerns and benefits participants considered when deciding (not) to use the system, as well as any comments regarding key factors. Given our open-ended quantitative measure of concerns and benefits, we felt it was essential to capture the specific concerns and benefits to inform GI system design. The first author began by tagging each open-ended response with a specific concern or benefit, if mentioned, resulting in the initial set of codes. Then, in collaboration with the last author, concerns and benefits codes were grouped into axial codes.

To derive design tensions, the three authors collaboratively assessed each finding against one another. To do this, we placed all quantitative findings (e.g., significant key factors in the benefits and concerns models) and qualitative codes (referring to a key factor) on sticky notes on a virtual whiteboard. Systematically iterating over each quantitative finding, we compared them to every other sticky note, discussing if these goals were in alignment or in tension. Then, we iterated through the qualitative findings. With a candidate list of tensions, we collaboratively grouped and prioritized these tensions based on how many pieces of evidence supported the tension (e.g., multiple key factors, or quantitative and qualitative results), and how critical the tension was to resolve, based on the impact it would have on workers.

4.3 Results

Across all vignettes, participants reported an average concern of 4.02/7 (std = 1.86), an average benefit of 4.58/7 (std = 1.82), and an average intention to use of 4.41/7 (std = 1.92). This suggests that participants are unsure whether the benefits outweigh the risks of the proposed GI tool and are not very confident in their desire to use it. The large standard deviations suggest that variation exists across participants and key vignette factors.

Table 3: Participant demographics (N = 235).

Age	**n (%)**	**Gender**	**n (%)**
18–24	39 (16.6%)	Man	111 (47.2%)
25–34	99 (42.2%)	Woman	119 (50.6%)
35–44	59 (25.1%)	Non-binary	3 (1.3%)
45–54	21 (8.9%)	Other	1 (0.4%)
55–64	12 (5.1%)	Prefer not to answer	1 (0.4%)
65–74	5 (2.1%)		
Race	**n (%)**	**Years worked**	**n (%)**
White	72 (30.6%)	<1	3 (1.2%)
Black	135 (57.4%)	1–3	69 (29.4%)
Latine	11 (4.7%)	3–5	54 (23.0%)
South Asian	10 (4.3%)	5–10	88 (37.4%)
East Asian	3 (1.3%)	10+	21 (8.9%)
Middle Eastern	2 (0.9%)		
Southeast Asian	1 (0.4%)		
Prefer not to say	1 (0.4%)		
Job Type	**n (%)**	**Job level**	**n (%)**
Education	25 (10.6%)	Not a manager	57 (24.3%)
General Management	70 (29.8%)	One-level mgr.	84 (35.7%)
Sales/Marketing	11 (4.7%)	Multi-level mgr.	87 (37.0%)
Finance/Accounting	18 (7.7%)	Unsure	5 (2.1%)
Healthcare	18 (7.7%)	Prefer not to say	2 (0.9%)
IT	52 (22.1%)		
Engineering	12 (5.1%)		
Other	29 (11.1%)		

Country of Residence	**n (%)**
South Africa	121 (51.5%)
United Kingdom	39 (16.6%)
United States	13 (5.5%)
Poland	11 (4.7%)
Other (Europe)	21 (8.9%)
Other (N. America)	14 (6.0%)
Other (Africa)	6 (2.6%)
Other (Asia)	5 (2.1%)
Other (S. America)	4 (1.7%)
Other (Oceania)	1 (0.4%)

We conducted a sensitivity analysis to determine the minimum possible partial effect size we could detect under the complex modelling parameters (Appendix G). The minimal detectable effect size with 80% power at $\alpha = 0.05$ is $r_{partial} = 0.10$ in the benefits model (0.18 points on a 1-7 scale) and $r_{partial} = 0.10$ on the concerns model (0.19 points on a 1-7 scale). As such, we can detect a substantially smaller effect than a conventional "small" effect.

Due to the overrepresentation of managers in our sample, who may feel differently about the balance of benefits and surveillance concerns [104], we re-ran all of the models presented below, including the interaction of each level of each key factor and manager status, which did not significantly improve model fit (Appendix H). Largely, benefits and concerns were not driven by manager status. The benefits model showed no significant interactions, whereas the concern model showed only one significant interaction (Manager × access - DM and upward reporting line), indicating that managers find this more concerning. In the intent model, managers responded less positively to total opt-out control, and different views for managers and ICs.

The $R^2_{marginal}$ values ranged from 0.15-0.17, and $R^2_{conditional}$ values from 0.28-0.29, suggesting that individual differences in participants' privacy perceptions account for much of these ratings. This is consistent with privacy research showing that people have strong baseline privacy ideals [74]. The absolute $R^2_{conditional}$ values suggest that individual-level contextual factors beyond baseline privacy orientation likely shape perceptions of GI systems in ways a hypothetical vignette study cannot fully capture. Prior research has proposed that privacy calculus is also influenced by macro factors such as the national culture, social norms, and legal provisions [14]. Organizational culture also plays a role [24], and we can expect that how safe one feels being vulnerable in their team, or their psychological safety [40], could also influence how beneficial they see the technology. Personal factors such as ethical orientation and prior experience with surveillance technology further shape how individuals reason about monitoring systems [14]. These factors are highly situational and person-specific, and capturing their influence would require in-situ or longitudinal study designs, which we identify as important future work.

We now present detailed results for each key factor.

4.3.1 Measurement Aggregation. Measurement aggregation had two levels: a team line (baseline), and a labelled individual line. As shown in Table 4, aggregating measurements at the individual level was a significant predictor of perceived benefit relative to group-level reporting, but not of perceived concern. Participants expressed that individual-level reporting would feel less cooperative than team-level reporting, which would create a more encouraging environment: *"When scores are reported at the individual level, the tool feels like surveillance or performance evaluation. This can discourage collaboration, create stress, and damage trust...group-level measurement would feel more supportive."* Other participants, however, noted that individual-level reporting would make the feedback more actionable: *"I would be more open to the tool when it balances personal accountability with team-level reporting to managers, I think people thrive when you focus on individual strengths and weaknesses..."* Thus, individual-level aggregation creates benefits through actionability, but creates ambiguous surveillance concerns—which is supported by the lack of statistical relationship with perceived concerns. This finding suggests that other design choices, in combination with measurement aggregation, may help to reduce some concerns.

4.3.2 Access. Access had five levels (Figure 2). As shown in Table 5, the only significant level within this key factor was when managers and their upward reporting line were the only groups with access, which was a positive predictor of concern. Participants indicated that this hierarchical access level increased feelings of surveillance. They noted that this may actually harm collaboration: *"[the] big boss watching you could discourage informal exchanges that are essential for my team cohesion."* Another noted that *"[if] none of the team members have access to data, we may end up not putting [in] any more effort,"* suggesting that individual access would be what led to behaviour changes. Some participants trusted the technology to *"protect my privacy,"* and noted that the pressure that came from multiple people being able to access the data, which *"may help every*

Table 4: Comprehensive Study: Benefits Model results. DM = Direct Manager, IC = Individual Contributor. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Each binary indicator (for race, gender, and country) uses non-members as the reference group

Predictor	Reference Category	Comparison Category	β (unstandardized)	Std. Error	Sig.
Measurement Aggregation	Team	Individual	0.28	0.14	*
Access	Team and DM	Only DM	0.21	0.17	
Access	Team and DM	Manager and upwards reporting line	-0.32	0.17	
Access	Team and DM	IC	0.31	0.21	
Access	Team and DM	Views for ICs and DM	0.01	0.23	
Data Type	Metadata	Messages	-0.28	0.15	
Data Type	Metadata	Project Outputs	-0.22	0.15	
Algorithmic Transparency	None	Source Data	0.91	0.15	***
Algorithmic Transparency	None	Narrative Description	1.00	0.15	***
Control Level	None	Some opt out control	0.33	0.15	*
Control Level	None	Total opt out control	0.64	0.15	***
Gender	All Others	Woman	-0.17	0.18	
Race	All Others	White	1.08	0.46	*
Race	All Others	Black	1.91	0.64	***
Race	All Others	Latine	0.52	0.71	
Race	All Others	South Asian	1.21	0.73	
Country	All Others	Canada	-0.89	0.54	
Country	All Others	India	-0.13	0.91	
Country	All Others	Italy	-0.29	0.64	
Country	All Others	Kenya	-1.29	0.73	
Country	All Others	Mexico	1.22	0.77	
Country	All Others	Poland	-0.48	0.51	
Country	All Others	Portugal	-0.96	0.62	
Country	All Others	South Africa	-0.56	0.49	
Country	All Others	United Kingdom	0.05	0.40	
Country	All Others	United States	0.22	0.46	
Age	25-34	18-24	-0.10	0.27	
Age	25-34	35-44	-0.13	0.21	
Age	25-34	45-54	-1.14	0.33	***
Age	25-34	55-64	-0.22	0.37	
Age	25-34	65-74	-1.06	0.72	
Manager Level	Multi-Level Manager	Not Manager	0.01	0.22	
Manager Level	Multi-Level Manager	One-Level Manager	0.17	0.20	
Years Worked	4-5	<1	2.94	1.48	*
Years Worked	4-5	1–3	-0.12	0.22	
Years Worked	4-5	10+	0.29	0.35	
Years Worked	4-5	5-10	-0.44	0.24	
PSS Score	-	PSS Measure	-0.24	0.13	

$R^2_{marginal} = .17$, $R^2_{conditional} = .29$; this indicates that vignette factors explained about 17% of the variance, while including participant-level random effects increased the explained variance to 29%

individual to perform better." Others highlighted concerns regarding competition that may arise, such as *"some members of staff thinking they're better than another member...who could just be having a difficult week/month."* The significance of manager-only access but non-significance of other levels suggests that the critical threshold for surveillance is vertical data flows to senior management. Other access levels are not universally interpreted, as participants believe that external competition could be helpful and harmful.

4.3.3 Data Type. Data type had three levels: metadata (baseline), messages and meeting transcripts, and completed work outputs. No data type was found to be a significant predictor of either perceived concern or perceived benefit. Open-ended responses highlighted mixed reactions. Some participants described that message monitoring could *"create a sense of surveillance among my peers."* Others, however, expressed appreciation for rich message-based assessments: *"[I] like the fact [that] the content of the messages is being assessed, not just the number of messages."* Perspectives on completed work were similarly divided. Some participants valued this emphasis on work-specific data, saying that they appreciate when the tool *"focuses on evaluating completed work rather than personal communications...more relevant and fair."* Yet others focused on the term "completed," noting that *"scores based only on completed work might undervalue other contributions (like mentoring, ideation, or support work), which could be unfairly reflected..."* For Metadata, which is commonly used in commercial analytics platforms [96], one participant suggested potential collaborative benefits that came

Table 5: Comprehensive Study: Concern Model results. DM = Direct Manager, IC = Individual Contributor. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Each binary indicator (for race, gender, and country) uses non-members as the reference group

Predictor	Reference Category	Comparison Category	β (unstandardized)	Std. Error	Sig.
Measurement Aggregation	Team	Individual	-0.08	0.14	
Access	Team and DM	Only DM	0.17	0.17	
Access	Team and DM	DM and upwards reporting line	0.38	0.17	*
Access	Team and DM	IC	-0.11	0.21	
Access	Team and DM	Views for ICs and DM	0.07	0.23	
Data Type	Metadata	Messages	0.21	0.15	
Data Type	Metadata	Project Outputs	0.11	0.15	
Algorithmic Transparency	None	Source Data	-0.46	0.14	**
Algorithmic Transparency	None	Narrative Description	-0.56	0.15	***
Control Level	None	Some opt out control	-0.28	0.15	
Control Level	None	Total opt out control	-0.44	0.15	**
Gender	All Others	Woman	0.39	0.18	*
Race	All Others	White	-0.75	0.47	
Race	All Others	Black	-1.85	0.56	***
Race	All Others	Latine	-1.05	0.72	
Race	All Others	South Asian	-0.01	0.75	
Country	All Others	Canada	1.09	0.57	
Country	All Others	India	-0.44	0.96	
Country	All Others	Italy	-0.43	0.67	
Country	All Others	Kenya	1.17	0.77	
Country	All Others	Mexico	0.28	0.76	
Country	All Others	Poland	-0.18	0.53	
Country	All Others	Portugal	0.61	0.66	
Country	All Others	South Africa	0.91	0.51	
Country	All Others	United Kingdom	-0.21	0.43	
Country	All Others	United States	0.17	0.48	
Age	25-34	18-24	0.12	0.27	
Age	25-34	35-44	-0.13	0.22	
Age	25-34	45-54	0.94	0.34	**
Age	25-34	55-64	0.83	0.38	*
Age	25-34	65-74	0.87	0.74	
Manager Level	Multi-Level Manager	Not Manager	-0.11	0.23	
Manager Level	Multi-Level Manager	One-Level Manager	-0.39	0.21	
Years Worked	4-5	<1	-3.30	1.52	*
Years Worked	4-5	1–3	0.42	0.23	
Years Worked	4-5	10+	-0.91	0.35	**
Years Worked	4-5	5-10	0.48	0.25	
PSS Score	–	PSS Measure	0.43	0.14	**

$R^2_{marginal} = .15$, $R^2_{conditional} = .28$; this indicates that vignette factors explained about 15% of the variance, while including participant-level random effects increased the explained variance to 28%

from knowing when and how someone was working, noting *"as a team we are able to help each other when we are all aware of the team's activity patterns."* Others worried about the limitations of the data, as it did not contain the content of the collaboration: *"measuring only activity patterns...doesn't reflect the quality of collaboration or the actual value of contributions..."* The lack of significant effects, despite rich qualitative disagreement, suggests that the fairness of a data source is not an inherent property, but is contextual based on privacy, completeness, and relevance to 'real' collaboration.

4.3.4 Algorithmic Transparency. Algorithmic transparency had three levels: no transparency (baseline), links to raw data, and a narrative explanation. Both the presence of a narrative description and links to specific data emerged as strong positive predictors of benefits and strong negative predictors of concern, as displayed in Tables 4 and 5. Regardless of the type of transparency, open-ended responses were overwhelmingly positive. Participants stressed that transparency would make the tool more actionable *"the written explanation makes the feedback more useful,"* and *"the inclusion of links provides data transparency...[to] see [what] contributed to the score"*. In contrast, when transparency was not provided, participants noted it as *"...useless for improvement"*. This response reveals that participants are not considering the negative consequences of transparency, either because they do not anticipate them, or because, in team collaboration (where much data is already public), the benefit of understanding the system outweighs other concerns.

4.3.5 *Control.* Control had three levels: no opt-out control (baseline), some opt-out control, and total opt-out control. As shown in Table 4, providing some level of control was a moderate predictor of benefit. Offering full control was a strong predictor of benefit and a considerable negative predictor of concern. Open-ended responses demonstrated a similar trend: the more opt-out control, the better. General consensus was that *"the ability to opt out of data collection—fully or partially—is essential,"* because *"mandatory participation undermines individual agency and may foster resistance or disengagement."* The absence of control was consistently described as problematic. As one participant put it, *"lack of choice reduces my sense of autonomy and increases discomfort with the tool."* Emerging as the single strongest factor in our study, participants' strong preference for control suggests they view it as a trust-building mechanism—when they choose to participate, they perceive the resulting data as more legitimate.

4.3.6 *Specific Concerns and Benefits.* In their open-ended responses, participants identified several broad categories of benefits. Participants described that the system may help improve time management: *"gaining insights into individual productivity patterns..."* Generally, participants believed the system would improve team performance by providing targeted feedback at the team level. Participants also noted that increased *"fairness and efficiency"* would result from a system that used consistent data and a systematic algorithm to assess collaboration.

Many participants brought up concerns regarding the idea of being "scored" in any way: *"a scoring system in the workplace is something extremely inappropriate"*, one participant wrote, which is similar to multiple open-ended comments that disapprove of any numerical result. A more specific concern was the fear that *"scores are too simplistic or miss important contributions (like collaboration, problem-solving, or mentoring)"*. Due to this risk of misinterpretation and missing important nuances, participants worried that the tool was *"likely to create distrust, stress, and unfair judgments"*.

Several participants discussed how the specific way a key factor is implemented will affect their perception of the tool. For example, a participant cited that they would perceive benefit from using metadata if activity patterns could be *"[uniquely] configured and if a manager could input the data that should be monitored"*, suggesting that customization is important. These findings suggest that participants do not object to GI systems broadly, but to inflexible numeric measures that cannot adapt to their team's context. Participants may accept the risks of quantification if they perceive they have agency in configuring the system to reflect their work practices.

4.3.7 *Tradeoff of Concerns and Benefits in Intention to Use.* We next determine whether participants reasoned about their intention to use the system by balancing the concerns and benefits—or, in other words, whether benefits and concerns mediated intention to use. We evaluated mediation using two complementary approaches. First, we compared a CLMM predicting intention to use, including the mediators (perceived benefits and concerns), against a model excluding them. A likelihood ratio test indicated that including perceived benefits and concerns significantly improved model fit compared to a reduced model that included only vignette factors, $\chi^2(2) = 1132.10, p < .001$. As shown in Table 6, we found that both benefits and concerns were significant predictors of intention to use ($p < 0.001$), where benefit was positively associated, and concern was negatively associated. Only a single key factor level remained significant, where having total control over the ability to opt out was positively related to intention to use ($p < 0.05$). Comparing the direct effects for each key factor before and after adding mediators, manager and upward reporting line access, datatype messages, and both types of transparency (linking to source data and narrative explanations) demonstrated full mediation, as their direct effects were no longer significant after adding the mediators. Total opt-out control exhibited partial mediation, with both significant indirect and direct effects.

Second, we used nonparametric bootstrapping (1,000 resamples) to estimate bias-corrected 95% confidence intervals for each indirect path (via perceived benefits and concerns). Indirect effects were considered significant when their confidence intervals excluded zero. We also report standardized indirect effects to communicate the effect size of these relationships. The results from this bootstrapping largely confirmed significant indirect effects (β_{std}) for the same key factors identified as mediated in the model comparison, including a significant negative indirect effect of only managers and upward reports having access, a significant negative indirect effect of messages as data type, and significant positive effects of transparency via data or narrative explanations, and total opt-out control. Additionally, bootstrapping revealed a significant indirect effect for some opt-out control that was not detected in the model comparison. Both types of transparency had the largest mediating influence on intention to use, with total opt-out control and some opt-out control showing moderate indirect effects, and manager-only access and messages as a data type showing smaller negative indirect effects. Together, these findings suggest that design factors influence intention to use primarily by shaping how users perceive benefits and concerns, rather than through direct paths. Details are shown in Appendix I.

4.4 Design Tensions

To address RQ2, we consider the quantitative and qualitative findings regarding a complete GI system within the context of asymmetric benefits and concerns in GI settings. We derive five preliminary design tensions that highlight the unique challenges with GIs.

Tension #1: Balancing Individual Agency and Representational Accuracy Users view opt-out control as beneficial. This aligns with privacy calculus theory, where perceived control is a critical mediator between risk perception and intention [17]. However, certain collaboration elements, such as psychological safety [40], exist only at a group level. Providing granular opt-out control to individuals would limit the system's ability to make accurate estimates for the entire team. This is similar to the "paradox of control" often discussed in privacy literature, where users believe that they want control over their data, but this control often leaves them more exposed in the end [17].

Tension #2: Balancing Individual Utility vs. Collective Cohesion Participants noted that displaying feedback on an individual level was more beneficial, with open-ended responses specifically noting this would be more actionable. In fact, past work that asked participants to list collaboration metrics they'd be interested in

Table 6: Cumulative link mixed model predicting Intention to Use (Path B). DM = Direct Manager, IC = Individual Contributor. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Each binary indicator (for race, gender, and country) uses non-members as the reference group

Predictor	Reference Category	Comparison Category	β (unstandardized)	Std. Error	Sig.
Measurement Aggregation	Team	Individual	-0.05	0.15	
Access	Team and DM	Only DM	0.05	0.18	
Access	Team and DM	DM and upwards reporting line	-0.04	0.18	
Access	Team and DM	IC	-0.05	0.23	
Access	Team and DM	Views for ICs and DM	0.26	0.24	
Data Type	Metadata	Messages	-0.15	0.16	
Data Type	Metadata	Project Outputs	-0.07	0.16	
Algorithmic Transparency	None	Source Data	-0.04	0.16	
Algorithmic Transparency	None	Narrative Description	0.06	0.16	
Control Level	None	Some opt out control	0.01	0.15	
Control Level	None	Total opt out control	0.39	0.16	*
Benefit	–	Benefit Measure	1.73	0.08	***
Concern	–	Privacy Concern Measure	-0.29	0.05	***
Gender	All Others	Woman	-0.09	0.17	
Race	All Others	White	-0.10	0.46	
Race	All Others	Black	0.55	0.54	
Race	All Others	Latine	1.01	0.70	
Race	All Others	South Asian	0.01	0.72	
Country	All Others	Canada	-0.03	0.55	
Country	All Others	India	0.08	0.90	
Country	All Others	Italy	1.25	0.67	
Country	All Others	Kenya	-1.52	0.75	*
Country	All Others	Mexico	-1.25	0.74	
Country	All Others	Poland	0.20	0.51	
Country	All Others	Portugal	0.68	0.64	
Country	All Others	South Africa	0.73	0.49	
Country	All Others	United Kingdom	-0.22	0.40	
Country	All Others	United States	-0.29	0.47	
Age	25-34	18-24	0.12	0.26	
Age	25-34	35-44	-0.09	0.21	
Age	25-34	45-54	-0.08	0.33	
Age	25-34	55-64	-1.09	0.38	**
Age	25-34	65-74	0.14	0.73	
Manager Level	Multi-level Manager	Not a Manager	-0.30	0.22	
Manager Level	Multi-level Manager	One-Level Manager	-0.00	0.19	
Years Worked	4-5	<1	1.67	1.41	
Years Worked	4-5	1–3	-0.29	0.22	
Years Worked	4-5	3–5	-0.26	0.23	
Years Worked	4-5	10+	-0.05	0.34	
PSS Score	–	PSS Measure	-0.45	0.13	***

McFadden's $R^2 = .44$ and Nagelkerke's $R^2 = .46$.

measuring showed that users still list characteristics that are measured individually, such as mood [84]. While some characteristics of collaboration are first measured at an individual level and then aggregated, such as turn-taking [111], other team-wide measures may be more than the sum of their parts—such as collective intelligence [106]. In these cases, it may be impossible to provide individual feedback. Furthermore, individually reported data made some participants concerned about surveillance. Thus, individual data is both high-utility (actionable), and high-risk.

Tension #3: Balancing Privacy and Transparency Participants consistently rated algorithmic transparency methods, such as narrative explanations and pointing to raw data, as making the system more beneficial and less concerning, suggesting that explanations are essential for trust. However, past work suggests that explainable AI methods may leak specific points in the training dataset [117]. Narrative explainable AI methods often employ extractive methods, in which specific input data (e.g., communication messages, recordings) are summarized into an explanation, which can result in data leakage [59]. When placed in the context of GI systems, we can imagine that this transparency may result in "placing blame," particularly because users have access to much of the source data to investigate themselves. While transparency mechanisms that point to the raw data are likely to increase trust in the tool, they could also reduce trust within teams.

Tension #4: Balancing Performance and Policing Open-ended responses revealed that participants had conflicting views on who should have access to the GI system. While it was generally believed to be more concerning when only managers and senior leaders accessed the data, other access levels were not significant in either model. Some participants were concerned that data could be misinterpreted if users lacked context about a specific team. Other participants, however, noted that they may be motivated to collaborate more effectively if they knew that others could see their performance. We observe significant personal preferences that should be considered in these design decisions, with some workers motivated by 'friendly competition' and others concerned about privacy. This suggests a thin line between motivation (facilitated by social comparison) and policing (facilitated by surveillance).

Tension #5: Balancing Algorithmic Consistency and Contextual Fairness While not significant in our concerns and benefits models, in open-ended responses, some participants noted that using completed work as input data would be fair and objective. Others worried that finished products would not capture "invisible labor" such as mentoring and ideation, which may not be present in digital traces. While automated metrics offer a form of procedural fairness, applying the same measure to all, they may lack contextual fairness by failing to account for nuanced contributions not digitally captured, which may be unequally distributed across roles.

5 Discussion

The goal of this work was to investigate how the concern-benefit trade-off manifests in GI systems, and how contextual design factors affect these trade-offs when there exists an asymmetry between individual and group benefits. Through an exploratory vignette study, we identified initial perceptions of GI design: individually presented data creates benefits through actionability, though with ambiguous concerns across users; there exists a critical threshold of access (managers and executives) that raises concerns; concerns regarding data types depend on how private, complete, and realistic the data is perceived to be; transparency is not seen as concerning, perhaps due to team open-sharing cultures; GI is seen as more legitimate if users choose to participate, even if accuracy is reduced, and; participants reject inflexible quantification in GI systems. We extend the work on informatics for pairs [78, 139], families [77], and ad-hoc teams [105] to start thinking about group informatics as a permanent fixture in knowledge work teams. We take the first step in answering the call from Gal et al. [54] for the empirical study of workplace informatics.

We identify five preliminary tensions when balancing individuals and groups, which act as a form of "intermediate knowledge" to bring empirical findings closer to design theory [64]. These tensions provide concrete instantiations of the frameworks of contextual integrity and privacy calculus in the context of GI systems, thereby exploring the specific mechanisms through which group-level asymmetry manifests in design. We've introduced the idea of GI systems that measure high-level collaboration constructs, such as psychological safety, moving beyond the simple metrics (e.g., number of messages sent, time in collaboration) that are currently implemented in commercial tools such as Microsoft Viva.

5.1 Asymmetric Balance in Individual Risk and Collective Benefit

Our design tensions illuminate how collective action dynamics reconfigure privacy calculus in group settings. Complicating the traditional understanding of privacy calculus as "if I share data, I will get something in return" [36], we identify the asymmetric effects across concerns and benefits.

We found that individuals perceive something as more beneficial when they have control, enabling them to protect their privacy by opting out. Yet, this creates a free-rider problem: individual concern is reduced without recognizing that this may make the entire system less accurate. This complicates traditional privacy calculus [36], as opting out of data collection creates a false sense of safety for the individual user, while actually degrading the system's outcomes and, in turn, team collaboration. We also found that users prefer actionable feedback presented at the individual level, contradicting prior work that argues data should be shared only anonymously [91] or in aggregate [84]. While these methods can protect privacy, they inherently also challenge accountability in a group setting, potentially resulting in the "bystander effect," where if group-level feedback is provided, everyone might consider it someone else's responsibility. Both of these findings suggest that users perceive collaboration effectiveness as the "sum of its parts," a view that many group-level constructs contradict [106].

As Seberger et al. [112] note, users seem to weigh benefits and concerns when deciding whether to use a system, but these trade-offs are not always straightforward. Not all factors were both positive benefits and negative concerns (or vice versa), as we might expect if all design factors were simply positive or negative. For example, having some control had a positive relationship with perceived benefit, but was not a significant predictor of concern. This suggests that there may be a more universal understanding of how different design choices lead to team benefits, and greater variation in how these choices may be perceived as concerning. This finding provides further evidence that it may be challenging for users to balance largely individual concerns (e.g., someone may see and misinterpret my performance), with collective benefits (e.g., the team may be able to reflect on and improve performance).

We identified key differences and similarities between our findings and the personal informatics, group informatics, and workplace surveillance literature. First, data type may be less important in reasoning about the concerns and benefits of a GI system. While we hypothesized that communication data would be sensitive in a workplace context (similar to emotional data [29]), data type was not a significant predictor of concerns or benefits — unlike in personal informatics [31, 126]. Open-ended responses revealed that it was too sensitive for some, but appropriately contextual for others. Second, granular access levels may have a limited effect on benefits and concerns in GI systems, contradicting past work [134]. It may be less important who on the team accesses the data, as long as team members also can. This may be due to the semi-public nature of most communication and work products in teams (consider public Slack channels and shared Google Drives). Third, past work has found that surveillance technologies are generally described as useful for leaders but not for workers [104]. Our concerns and benefits models showed little difference in perception between managers and non-managers, which aligns with the goal of the GI systems we have proposed, where they are used primarily as a collaborative tool to support teamwork.

We add evidence to findings from PI literature that transparency increases intention to use a system [11, 139, 140]. Both narrative and raw data transparency are perceived as beneficial. Building on Das Swain et al. [30], this suggests that LLMs would likely be useful for these narrative explanations. If the risks of hallucination or data privacy concerns (e.g., non-local models) were too high, the system could instead link directly to the raw data. We also add evidence to PI findings that show that being in control of what one chooses to share, whether with another individual [78, 139] or the system itself [23, 62], can support self-reflection. Control being the strongest predictor aligns with theories of algorithmic agency, which suggest that users will trust AI-based systems when they are perceived as active agents rather than data subjects [115]. We also echo some concerns from group informatics literature, such that these systems may create mistrust [90]. Extending this, we show that people resist the idea of being quantified in any way, and think that a lot of the work they do that is “soft” e.g., mentoring, would be missed by numbers, and that flexibility in how these features are configured, so that they could align with team norms, was critical.

5.2 Contextual Integrity in Group Settings

We extend the discussion of contextual integrity in HCI [26, 57, 60] by operationalizing the elements of the theory [87, 101], and adding to the discussion of which elements may be most important in the asymmetric group context. Our findings show that the data subject (i.e., the individuals or teams the information concerns) in the contextual integrity framework is important for achieving benefits in the GI context. The most significant relationships with perceived concerns and benefits were algorithmic transparency and control, which we argue are both examples of the transmission principle element of contextual integrity. Transmission principles control whether workers can reduce their own concern while still benefiting. Measurement aggregation obscures individual opt-out, and transparency reveals how individual data become team metrics (potentially revealing missing contributors). These features reconfigure the free-rider problem, and suggest that for GI systems, we may need to be more creative in creating rules for when data can be shared and with whom. Our findings show that, in terms of the actors (or recipients) of the data, there may not be a universally agreed-upon way to segment who can access it and who cannot. Due to the semi-public nature of much of the team-level data types we analyzed in this work, we recognize that users may not consider the data type an important factor when deciding whether to use a GI system. However, we hypothesize that if these systems collect more invasive data (e.g., keystrokes, private/direct messages, webcam surveillance), then data type could become more important.

5.3 Design Implications

All the design tensions mentioned in Section 4.4 are speculative in this exploratory study. Validation of these preliminary design tensions is part of future work, which will need to thoughtfully and meaningfully address these tensions and their implications for GI system design. We enumerate through each tension in Section 4.4. We recognize that each GI system and its goals may be different, so we do not aim to be prescriptive, but rather, use examples to encourage those designing GI systems to consider these tensions.

First, we acknowledge that even when GI systems are designed with collaborative intent, employees may interpret them through the lens of existing organizational power dynamics [29, 104]. Despite the most thoughtful design, organizations may still deploy GI systems for oversight or productivity measurement while framing them as collaboration tools, and employees' awareness of this gap shapes their perceptions of concerns. This concern extends to opt-out control—our finding assumes that opting out is a genuinely free choice. In practice, employees may fear that opting out signals distrust or non-compliance. Thus, we must not only consider the design of these systems, but the power dynamics, organizational culture, and implementation practices that will influence effectiveness as much as design does. Future work that qualitatively investigates how these systems are integrated into work practices, and when they work or fail, will contribute the organizational requirements alongside the design requirements explored here.

Tension #1: When users lack control, the perceived concerns often outweigh the potential benefits. By granting control, the system lowers the perceived privacy cost. We argue that in GI, this control is a prerequisite for accurate systems. Without the safeguards provided by control mechanisms, users may game the system, rendering its data useless regardless of its volume. To prevent perceptions of a loss of agency at work [134], we suggest designers enable GI systems to be configurable, allowing leaders to work with employees to identify which feedback (and, thus, required data) is most critical. For example, a team may select a few measures among many possibilities to maximize measurement accuracy where individuals are most comfortable. Alternatively, designers could specify a threshold of participation for accuracy (e.g., 80% of the team). If too many users opt out, the system should suppress the metric entirely rather than reporting inaccurately. Future work can also look to legal and philosophical solutions for group-level privacy (e.g., Taylor and Floridi [129]) that could inspire design choices.

Tension #2: Even though GI systems focus on group-level metrics, there seems to be some need for personalized, individual feedback. Future work should investigate this challenge by each group-level metric, as some may be able to be broken down to provide individual feedback, while others may not. If possible, we can consider developing GI systems that allow each user to interact with the data in their own way, such as a personalized visual encoding, which can then connect to actionable items and the broader picture [140]. Broadly, participants rejected the idea of a numerical “score" for their collaboration. Thus, while some constructs may exist in theory only at the group level, by focusing on narrative feedback rather than quantification, GI systems may be able to tailor team-level feedback to each individual based on their specific collaboration style. For example, if a team regularly experiences misunderstandings and one team member is rarely involved in these conversations, their feedback may ask them to communicate more explicitly as an example to the team.

Tension #3: Participants noted that transparency was critical to use the system. Yet, many transparency mechanisms risk exposing raw data that may single out team members. Both tensions

2 and 3 may require a technical solution, similar to the computational approaches to multi-party privacy [39, 46, 69, 142]. In GI, this could involve investigating different algorithmic transparency methods, particularly given research on LLMs and explainable AI [32, 42, 47, 141]. To address the risk of data leakage through transparency, we recommend progressive disclosure of explanations [122]. Given that both narrative and raw data explanations were rated as increasingly beneficial, GI systems may first provide narrative explanations of the metric. If raw data is needed for trust, designers can use synthetic explanations by generating representative, anonymized examples that illustrate the pattern without exposing real interactions between specific colleagues. Although not in a workplace setting, Spotify Wrapped is a commercial example of blended explanations and visualizations that address both personal and group data, as well as narrative and raw data [121].

Tension #4: Participants seemed to have differing opinions on the impact of data being shared with those outside of their team—with some viewing this as healthy competition, and others seeing this as policing. Since sharing data to motivate employees is separated from policing employees by a thin line of context, GI systems may temporally bound access. For example, GI dashboards may only be available during specific events, such as a sprint retrospective on a software team or when employees are reflecting on their yearly performance. This might encourage employees to use these systems to benefit collaboration, or internally motivate themselves, without risking anyone becoming overly focused on a metric, and preventing perceptions that one is "always being watched." By bounding the "surveillance" to a specific, agreed-upon time, the system may be seen as a specific tool for a specific task, rather than an always-on monitor.

Tension #5: Participants were worried that much of the labour they do at work is invisible to automated algorithms. Any systems that rely on automated measures are likely to miss any "invisible work", or even work that is simply "digitally invisible", even if it is visible in person—ignoring principles of data feminism [35]. To address this fairness gap, GI systems may need to consider hybrid metrics, where users can input their own context into the system (e.g., record mentoring or emotional support activities) that augment any digital traces. One view of the dashboard could explicitly highlight the discrepancy between digital-trace-based and human-reported measures, forcing the system and its users to acknowledge blind spots around invisible labour.

Lastly, many of these design decisions will depend on the collaboration elements measured by the system. There are numerous ways to classify successful collaboration [88, 118]—some are bound to be more important than others, more easily measured using digital traces than others, and more easily (dis)aggregated than others. In future work, we aim to bridge the gap between teamwork literature and HCI by understanding which, if any, of these elements can be translated for GI systems. A better understanding of what GI systems can measure will help inform detailed design guidelines.

5.4 Limitations

Due to the contextual nature of privacy, Barkhuus [8] argues that imagined scenarios may not always accurately capture privacy perceptions. Thus, participants' survey perceptions may not align with how they would feel using the system. Additionally, we deliberately left the specifics of the GI system vague—wanting to gather broad perceptions of GI systems as a whole rather than specific reactions to a single measurement, which likely also affected the participants' ability to imagine themselves in this scenario. As such, we present these findings as an early exploration of the potential design space, and not as final guidelines for future development. To overcome these limitations, we plan to hone in on a specific implementation of GI, and use an in-depth qualitative approach (using a design probe) to enable more realistic and nuanced design feedback. Our vignette study framed GI systems as tools designed for team benefit, which may have led participants to evaluate them more charitably than they would in a real deployment, where organizational motives are ambiguous and can be perceived as surveillance-oriented [29]. Future qualitative studies will also ask participants to consider how the tool might be used "correctly" or "incorrectly" across use cases.

We are also limited by the study's participant pool. Using crowdsourcing platforms meant that our participant pool reflected those who were on the platform at the time of the study launch, which in our case, meant that we had a large percentage of participants from South Africa. It has been noted that attitudes towards privacy can vary between cultures [2]. Further, our sample contained a larger percentage of managers. While past work has shown that managers perceive workplace surveillance risks differently than employees [104], our sensitivity analysis showed that our findings were not largely driven by managers. Due to the increasing length of the survey, our average payment value fell below the ideal Prolific amount. While in line with other work [58, 103, 120], this lower amount may have introduced a self-selection bias among participants who completed our study.

6 Conclusions

To analyze the trade-offs between perceptions of benefits and concerns in group informatics systems, an asymmetric context where individual concerns are weighed against team benefits, we conducted a pilot and a comprehensive factorial vignette study. We analyzed the influence of measure aggregation, access levels, algorithmic transparency, data types, and control on perceptions of concerns, benefits, and intentions to use these systems, finding that design factors, specifically algorithmic transparency and user control, may influence intention to use primarily through benefits and concerns. We reveal five preliminary design tensions that must be carefully managed in the design of group informatics systems, and we extend theories of privacy calculus and contextual integrity to describe privacy decisions in group contexts.

Acknowledgments

We are grateful for the support of a Microsoft AI and the New Future of Work grant which enabled this work.

The authors wish to acknowledge the use of the generative AI tool Perplexity in the writing of this paper. Perplexity was used in brainstorming phrasing and sentence structure. The paper remains an accurate representation of the authors' underlying work and novel intellectual contributions.

References

[1] Noura Abdi, Xiao Zhan, Kopo M. Ramokapane, and Jose Such. 2021. Privacy Norms for Smart Home Personal Assistants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–14. doi:10.1145/3411764.3445122

[2] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2015. Privacy and human behavior in the age of information. *Science* 347, 6221 (Jan. 2015), 509–514. doi:10.1126/science.aaa1465

[3] Daniel A. Adler, Emily Tseng, Khatiya C. Moon, John Q. Young, John M. Kane, Emanuel Moss, David C. Mohr, and Tanzeem Choudhury. 2022. Burnout and the Quantified Workplace: Tensions around Personal Sensing Interventions for Stress in Resident Physicians. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 430 (Nov. 2022), 48 pages. doi:10.1145/3555531

[4] Yoana Ahmetoglu, Duncan Brumby, and Anna Cox. 2024. Bridging the Gap Between Time Management Research and Task Management App Design: A Study on the Integration of Planning Fallacy Mitigation Strategies. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work*. 1–14.

[5] Noah Apthorpe, Yan Shvartzshnaider, Arunesh Mathur, Dillon Reisman, and Nick Feamster. 2018. Discovering Smart Home Internet of Things Privacy Norms Using Contextual Integrity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (July 2018), 1–23. doi:10.1145/3214262

[6] Zahra Atf and Peter R Lewis. 2025. Is trust correlated with explainability in AI? A meta-analysis. *IEEE Transactions on Technology and Society* (2025).

[7] Christiane Atzmüller and Peter M. Steiner. 2010. Experimental Vignette Studies in Survey Research. *Methodology* 6, 3 (Jan. 2010), 128–138. doi:10.1027/1614-2241/a000014

[8] Louise Barkhuus. 2012. The mismeasurement of privacy: using contextual integrity to reconsider privacy in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 367–376.

[9] Marcel Becker. 2019. Privacy in the digital age: comparing and contrasting individual versus social approaches towards privacy. *Ethics and Information Technology* 21, 4 (Dec. 2019), 307–317. doi:10.1007/s10676-019-09508-z

[10] Rim Ben Salem, Esma Aïmeur, and Hicham Hage. 2022. Aegis: An Agent for Multi-party Privacy Preservation. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Oxford United Kingdom, 68–77. doi:10.1145/3514094.3534134

[11] Marit Bentvelzen, Paweł W. Woźniak, Pia S.F. Herbes, Evropi Stefanidi, and Jasmin Niess. 2022. Revisiting Reflection in HCI: Four Design Resources for Technologies that Support Reflection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (March 2022), 1–27. doi:10.1145/3517233

[12] Lars Bergkvist and John R. Rossiter. 2007. The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research* 44, 2 (2007), 175–184. doi:10.1509/jmkr.44.2.175

[13] Jaspreet Bhatia and Travis D. Breaux. 2018. Empirical Measurement of Perceived Privacy Risk. *ACM Transactions on Computer-Human Interaction* 25, 6 (Dec. 2018), 1–47. doi:10.1145/3267808

[14] Devasheesh P. Bhave, Laurel H. Teo, and Reeshad S. Dalal. 2020. Privacy at Work: A Review and a Research Agenda for a Contested Terrain. *Journal of Management* 46, 1 (Jan. 2020), 127–164. doi:10.1177/0149206319878254

[15] August Bourgeus, Laurens Vandercruysse, and Nanouk Verhulst. 2024. Understanding contextual expectations for sharing wearables' data: Insights from a vignette study. *Computers in Human Behavior Reports* 15 (Aug. 2024), 100443. doi:10.1016/j.chbr.2024.100443

[16] Karen L. Boyd and Nazanin Andalibi. 2023. Automated Emotion Recognition in the Workplace: How Proposed Technologies Reveal Potential Futures of Work. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 95 (April 2023), 37 pages. doi:10.1145/3579528

[17] Laura Brandimarte, Alessandro Acquisti, and George Loewenstein. 2013. Misplaced Confidences: Privacy and the Control Paradox. *Social Psychological and Personality Science* 4, 3 (May 2013), 340–347. doi:10.1177/1948550612455931 Publisher: SAGE Publications Inc.

[18] Rollin Brant. 1990. Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression. *Biometrics* 46, 4 (1990), 1171–1178. doi:10.2307/2532457

[19] Teshan S Bunwaree, Katarzyna Stawarz, Philippa Collins, and Sandy JJ Gould. 2025. Boss is aWare—Are you? Employee comprehension and legal awareness of workplace monitoring. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–19.

[20] Dan Calacci. 2022. Organizing in the End of Employment: Information Sharing, Data Stewardship, and Digital Workerism. In *Proceedings of the 1st Annual Meeting of the Symposium on Human-Computer Interaction for Work (CHIWORK '22)*. Association for Computing Machinery, New York, NY, USA, 1–9. doi:10.1145/3533406.3533424

[21] Christine C. Call, Kristen L. Eckstrand, Steven W. Kasparek, Cassandra L. Boness, Lorraine Blatt, Nabila Jamal-Orozco, Derek M. Novacek, and Dan Foti. 2023. An Ethics and Social Justice Approach to Collecting and Using Demographic Data for Psychological Researchers. *Perspectives on psychological science : a journal of the Association for Psychological Science* 18, 5 (Sept. 2023), 979–995. doi:10.1177/17456916221137350

[22] Jiaxun Cao, Hiba Laabadli, Chase H Mathis, Rebecca D Stern, and Pardis Emami-Naeini. 2024. " I Deleted It After the Overturn of Roe v. Wade": Understanding Women's Privacy Concerns Toward Period-Tracking Apps in the Post Roe v. Wade Era. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–22.

[23] Aishwarya Chandrasekaran, London Bielicke, Diya Shah, Harisha Janakiraman, and Matthew Louis Mauriello. 2025. "I spent 14 hours debugging just one assignment": Toward Computer-Mediated Personal Informatics for Computer Science Student Mental Health. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–19. doi:10.1145/3706598.3713269

[24] Shuchih Ernest Chang, Anne Yenching Liu, and Sungmin Lin. 2015. Exploring privacy and trust for employee monitoring. *Industrial Management & Data Systems* 115, 1 (Feb. 2015), 88–106. doi:10.1108/IMDS-07-2014-0197

[25] Xinyue Chen, Lev Tankelevitch, Rishi Vanukuru, Ava Elizabeth Scott, Payod Panda, and Sean Rintel. 2025. Are We On Track? AI-Assisted Active and Passive Goal Reflection During Meetings. In *CHI 2025*. ACM. https://www.microsoft.com/en-us/research/publication/are-we-on-track-ai-assisted-active-and-passive-goal-reflection-during-meetings/

[26] Madiha Zahrah Choksi, Ero Balso, Frauke Kreuter, and Helen Nissenbaum. 2024. Privacy for Groups Online: Context Matters. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW2 (Nov. 2024), 406:1–406:23. doi:10.1145/3686945

[27] Shreya Chowdhary, Anna Kawakami, Mary L Gray, Jina Suh, Alexandra Olteanu, and Koustuv Saha. 2023. Can workers meaningfully consent to workplace wellbeing technologies?. In *Proceedings of the 2023 ACM conference on fairness, accountability, and transparency*. 569–582.

[28] Marios Constantinides and Daniele Quercia. 2022. The future of hybrid meetings. In *Proceedings of the 1st Annual Meeting of the Symposium on Human-Computer Interaction for Work*. 1–6.

[29] Shanley Corvite, Kat Roemmich, Tillie Ilana Rosenberg, and Nazanin Andalibi. 2023. Data Subjects' Perspectives on Emotion Artificial Intelligence Use in the Workplace: A Relational Ethics Lens. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–38.

[30] Vedant Das Swain, Lan Gao, William A Wood, Srikruthi C Matli, Gregory D. Abowd, and Munmun De Choudhury. 2023. Algorithmic Power or Punishment: Information Worker Perspectives on Passive Sensing Enabled AI Phenotyping of Performance and Wellbeing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–17. doi:10.1145/3544548.3581376

[31] Vedant Das Swain and Koustuv Saha. 2024. Teacher, trainer, counsel, spy: How generative AI can bridge or widen the gaps in worker-centric digital phenotyping of Wellbeing. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work*. 1–13.

[32] Thomas Deacon and Mark D. Plumbley. 2024. Working with AI Sound: Exploring the Future of Workplace AI Sound Technologies. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work (CHIWORK '24)*. Association for Computing Machinery, New York, NY, USA, 1–21. doi:10.1145/3663384.3663391

[33] Elena Di Lascio, Shkurta Gashi, Juan Sebastian Hidalgo, Beatrice Nale, Maike E Debus, and Silvia Santini. 2020. A multi-sensor approach to automatically recognize breaks and work activities of knowledge workers in academia. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–20.

[34] Erich C Dierdorff, David M Fisher, and Robert S Rubin. 2019. The power of percipience: Consequences of self-awareness in teams on team-level functioning and performance. *Journal of Management* 45, 7 (2019), 2891–2919.

[35] Catherine D'Ignazio and Lauren F. Klein. 2020. *Data Feminism*. The MIT Press. doi:10.7551/mitpress/11805.001.0001

[36] Tamara Dinev and Paul Hart. 2006. An Extended Privacy Calculus Model for E-Commerce Transactions. *Information Systems Research* 17, 1 (March 2006), 61–80. doi:10.1287/isre.1060.0080 Publisher: INFORMS.

[37] Verena Distler, Tamara Gutfleisch, Carine Lallemand, Gabriele Lenzini, and Vincent Koenig. 2022. Complex, but in a good way? How to represent encryption to non-experts through text and visuals – Evidence from expert co-creation and a vignette experiment. *Computers in Human Behavior Reports* 5 (March 2022), 100161. doi:10.1016/j.chbr.2021.100161

[38] Kimberly Do, Maya De Los Santos, Michael Muller, and Saiph Savage. 2024. Designing Gig Worker Sousveillance Tools. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 384, 19 pages. doi:10.1145/3613904.3642614

[39] Wenliang Du and Mikhail J. Atallah. 2001. Secure multi-party computation problems and their applications: a review and open problems. In *Proceedings of the 2001 workshop on New security paradigms (NSPW '01)*. Association for

Computing Machinery, New York, NY, USA, 13–22. doi:10.1145/508171.508174
[40] Amy Edmondson. 1999. Psychological safety and learning behavior in work teams. *Administrative science quarterly* 44, 2 (1999), 350–383.
[41] Amy C Edmondson and Derrick P Bransby. 2023. Psychological Safety Comes of Age: Observed Themes in an Established Literature. *Annual Review of Organizational Psychology and Organizational Behavior* (2023). doi:10.1146/annurev-orgpsych-120920-055217
[42] Upol Ehsan and Mark Riedl. 2024. Explainable AI Reloaded: Challenging the XAI Status Quo in the Era of Large Language Models. In *Proceedings of the Halfway to the Future Symposium*. ACM, Santa Cruz CA USA, 1–8. doi:10.1145/3686169.3686185
[43] Pardis Emami Naeini, Martin Degeling, Lujo Bauer, Richard Chow, Lorrie Faith Cranor, Mohammad Reza Haghighat, and Heather Patterson. 2018. The influence of friends and experts on privacy decision making in IoT scenarios. *Proceedings of the ACM on human-computer interaction* 2, CSCW (2018), 1–26.
[44] Daniel A Epstein, Daniel Avrahami, and Jacob T Biehl. 2016. Taking 5: Work-breaks, productivity, and opportunities for personal informatics for knowledge workers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 673–684.
[45] Daniel A Epstein, Clara Caldeira, Mayara Costa Figueiredo, Xi Lu, Lucas M Silva, Lucretia Williams, Jong Ho Lee, Qingyang Li, Simran Ahuja, Qiuer Chen, et al. 2020. Mapping and taking stock of the personal informatics literature. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–38.
[46] Qi Feng, Debiao He, Zhe Liu, Huaqun Wang, and Kim-Kwang Raymond Choo. 2020. SecureNLP: A System for Multi-Party Privacy-Preserving Natural Language Processing. *IEEE Transactions on Information Forensics and Security* 15 (2020), 3709–3721. doi:10.1109/TIFS.2020.2997134
[47] Sharon Ferguson, Paula Akemi Aoyagui, Rimsha Rizvi, Young-Ho Kim, and Anastasia Kuzminykh. 2024. The Explanation That Hits Home: The Characteristics of Verbal Explanations That Affect Human Perception in Subjective Decision-Making. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW2 (Nov. 2024), 1–37. doi:10.1145/3687056
[48] Sharon Ferguson, Kimberly Lai, James Chen, Safa Faidi, Kevin Leonardo, and Alison Olechowski. 2022. "Why couldn't we do this more often?": exploring the feasibility of virtual and distributed work in product design engineering. *Research in engineering design* 33, 4 (2022), 413–436.
[49] Sharon A Ferguson, Georgia Van de Zande, and Alison Olechowski. 2024. No risk, no reward: Towards an automated measure of psychological safety from online communication. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–7.
[50] Andy Field, Jeremy Miles, and Zoë Field. 2012. *Discovering Statistics Using R*. SAGE. Google-Books-ID: Q9GCAgAAQBAJ.
[51] Daniel Franzen, Claudia Müller-Birn, and Odette Wegwarth. 2024. Communicating the Privacy-Utility Trade-off: Supporting Informed Data Donation with Privacy Decision Interfaces for Differential Privacy. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1 (April 2024), 32:1–32:56. doi:10.1145/3637309
[52] Thomas Fritz, Alexander Lill, André N. Meyer, Gail C. Murphy, and Lauren Howe. 2023. Cultivating a Team Mindset about Productivity with a Nudge: A Field Study in Hybrid Development Teams. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2 (Oct. 2023), 335:1–335:21. doi:10.1145/3610184
[53] Adrian Furnham and Viren Swami. 2015. An Investigation of Attitudes toward Surveillance at Work and Its Correlates. *Psychology* 06, 13 (2015), 1668. doi:10.4236/psych.2015.613163 Publisher: Scientific Research Publishing.
[54] Uri Gal, Tina Blegind Jensen, and Mari-Klara Stein. 2020. Breaking the vicious cycle of algorithmic management: A virtue ethics approach to people analytics. *Information and Organization* 30, 2 (June 2020), 100301. doi:10.1016/j.infoandorg.2020.100301
[55] Cornelia Gerdenitsch, Till Bieg, Myriam Gaitsch, Philip Schörpf, Manfred Tscheligi, and Simone Kriglstein. 2023. Tracking to Success? A Critical Reflection on Workplace Quantified-Self Technologies from a Humanistic Perspective. In *Proceedings of the 2nd Annual Meeting of the Symposium on Human-Computer Interaction for Work*. 1–7.
[56] Katherine Gibbard, Harjinder Gill, Deborah Powell, and Peter A Hausdorf. 2025. Explain it to me like I'm five: harnessing the power of explanations to increase trust in workplace generative AI. *Behaviour & Information Technology* (2025), 1–19.
[57] Sarah Gilbert, Jessica Vitak, and Katie Shilton. 2021. Measuring Americans' comfort with research uses of their social media data. *Social Media+ Society* 7, 3 (2021), 20563051211033824.
[58] Akshit Gupta, Debadeep Basu, Ramya Ghantasala, Sihang Qiu, and Ujwal Gadiraju. 2022. To Trust or Not To Trust: How a Conversational Interface Affects Trust in a Decision Support System. In *Proceedings of the ACM Web Conference 2022*. ACM, Virtual Event, Lyon France, 3531–3540. doi:10.1145/3485447.3512248
[59] Sai Gurrapu, Ajay Kulkarni, Lifu Huang, Ismini Lourentzou, and Feras A. Batarseh. 2023. Rationalization for explainable NLP: a survey. *Frontiers in Artificial Intelligence* 6 (Sept. 2023). doi:10.3389/frai.2023.1225093 Publisher: Frontiers.
[60] Hilda Hadan, Derrick M Wang, Lennart E Nacke, and Leah Zhang-Kennedy. 2024. Privacy in immersive extended reality: Exploring user perceptions, concerns, and coping strategies. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–24.
[61] Kristina Hall, Berit Helmus, and Torsten Eymann. 2024. How to Balance Privacy and (Health) Benefits: Privacy Calculus and the Intention to Use Health Tracking at the Workplace. *International Journal of Human–Computer Interaction* (2024), 1–18.
[62] Rie Helene (Lindy) Hernandez, Qiurong Song, Yubo Kou, and Xinning Gui. 2024. "At the end of the day, I am accountable": Gig Workers' Self-Tracking for Multi-Dimensional Accountability Management. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–20. doi:10.1145/3613904.3642151
[63] Roberto Hoyle, Luke Stark, Qatrunnada Ismail, David Crandall, Apu Kapadia, and Denise Anthony. 2020. Privacy Norms and Preferences for Photos Posted Online. *ACM Trans. Comput.-Hum. Interact.* 27, 4 (Aug. 2020), 30:1–30:27. doi:10.1145/3380960
[64] Kristina Höök and Jonas Löwgren. 2012. Strong concepts: Intermediate-level knowledge in interaction design research. *ACM Trans. Comput.-Hum. Interact.* 19, 3 (Oct. 2012), 23:1–23:18. doi:10.1145/2362364.2362371
[65] Sarah Janboecke, Diana Loeffler, and Marc Hassenzahl. 2020. Using Experimental Vignettes to Study Early-Stage Automation Adoption. doi:10.48550/arXiv.2004.07032 arXiv:2004.07032 [cs].
[66] Guillermina Jasso. 2006. Factorial survey methods for studying beliefs and judgments. *Sociological Methods & Research* 34, 3 (2006), 334–423.
[67] P. Jayaprabha and K. Paulose Jacob. 2025. Critical Review of Different Approaches of Multiparty Privacy Protection Methods and Effectiveness on Social Media. *Transactions on Emerging Telecommunications Technologies* 36, 4 (April 2025), e70130. doi:10.1002/ett.70130
[68] Gyuwon Jung and Uichin Lee. 2025. CounterStress: Enhancing Stress Coping Planning through Counterfactual Explanations in Personal Informatics. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–20. doi:10.1145/3706598.3713730
[69] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. 2015. Secure Multi-party Differential Privacy. In *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett (Eds.), Vol. 28. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2015/file/a01610228fe998f515a72dd730294d87-Paper.pdf
[70] Anna Kawakami, Shreya Chowdhary, Shamsi T Iqbal, Q Vera Liao, Alexandra Olteanu, Jina Suh, and Koustuv Saha. 2023. Sensing wellbeing in the workplace, why and for whom? envisioning impacts with organizational stakeholders. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–33.
[71] Elisabeth T Kersten-van Dijk, Joyce HDM Westerink, Femke Beute, and Wijnand A IJsselsteijn. 2017. Personal informatics, self-insight, and behavior change: a critical review of current literature. *Human–Computer Interaction* 32, 5-6 (2017), 268–296.
[72] Jane Paik Kim and Hyun-Joon Yang. 2024. A novel experimental vignette methodology: SMART vignettes. *Methodological Innovations* 17, 2 (June 2024), 111–118. doi:10.1177/20597991241240081 Publisher: SAGE Publications Ltd.
[73] Maina Korir, Sharon Slade, Wayne Holmes, Yingfei Héliot, and Bart Rienties. 2023. Investigating the dimensions of students' privacy concern in the collection, use and sharing of data for learning analytics. *Computers in Human Behavior Reports* 9 (March 2023), 100262. doi:10.1016/j.chbr.2022.100262
[74] Ponnurangam Kumaraguru and Lorrie Faith Cranor. 2005. Privacy indexes: a survey of Westin's studies. (2005).
[75] Lin Kyi, Sushil Ammanaghatta Shivakumar, Cristiana Teixeira Santos, Franziska Roesner, Frederike Zufall, and Asia J Biega. 2023. Investigating deceptive design in GDPR's legitimate interest. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.
[76] Carine Lallemand, Alina Lushnikova, Joshua Dawson, and Romain Toebosch. 2024. Trinity: A Design Fiction to Unravel the Present and Future Tensions in Professional Informatics and Awareness Support Tools. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work*. 1–15.
[77] Hyunsoo Lee, Yugyeong Jung, Youwon Shin, Hyesoo Park, Woohyeok Choi, and Uichin Lee. 2024. FamilyScope: Visualizing Affective Aspects of Family Social Interactions using Passive Sensor Data. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (April 2024), 1–27. doi:10.1145/3637334
[78] Kwangyoung Lee, Yeohyun Jung, Gyuwon Jung, Xi Lu, and Hwajung Hong. 2025. Peerspective: A Study on Reciprocal Tracking for Self-awareness and Relational Insight. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–17. doi:10.1145/3706598.3713404
[79] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 557–566.
[80] Tianyi Li, Mihaela Vorvoreanu, Derek Debellis, and Saleema Amershi. 2023. Assessing Human-AI Interaction Early through Factorial Surveys: A Study on

the Guidelines for Human-AI Interaction. *ACM Transactions on Computer-Human Interaction* 30, 5 (Oct. 2023), 1–45. doi:10.1145/3511605

[81] Tony W Li, Arshia Arya, and Haojian Jin. 2024. Redesigning Privacy with User Feedback: The Case of Zoom Attendee Attention Tracking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–14. doi:10.1145/3613904.3642594

[82] Zhan Liu, Jialu Shan, Riccardo Bonazzi, and Yves Pigneur. 2014. Privacy as a Tradeoff: Introducing the Notion of Privacy Calculus for Context-Aware Mobile Applications. In *2014 47th Hawaii International Conference on System Sciences*. 1063–1072. doi:10.1109/HICSS.2014.138 ISSN: 1530-1605.

[83] Alina Lushnikova, Kerstin Bongard-Blanchy, Vincent Koenig, and Carine Lallemand. 2023. Eliciting Meaningful Collaboration Metrics: Design Implications for Self-Tracking Technologies at Work. In *Human-Computer Interaction – INTERACT 2023*, José Abdelnour Nocera, Marta Kristín Lárusdóttir, Helen Petrie, Antonio Piccinno, and Marco Winckler (Eds.). Vol. 14144. Springer Nature Switzerland, Cham, 643–664. doi:10.1007/978-3-031-42286-7_36 Series Title: Lecture Notes in Computer Science.

[84] Alina Lushnikova, Kerstin Bongard-Blanchy, and Carine Lallemand. 2022. What aspects of collaboration are meaningful to you? informing the design of self-tracking technologies for collaboration. In *Adjunct Proceedings of the 2022 Nordic Human-Computer Interaction Conference*. 1–5.

[85] Christoph Lutz and Aurelia Tamò-Larrieux. 2021. Do Privacy Concerns About Social Robots Affect Use Intentions? Evidence From an Experimental Vignette Study. *Frontiers in Robotics and AI* 8 (April 2021). doi:10.3389/frobt.2021.627958 Publisher: Frontiers.

[86] David P MacKinnon, Chondra M Lockwood, and Jason Williams. 2004. Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate behavioral research* 39, 1 (2004), 99–128.

[87] Nathan Malkin. 2023. Contextual Integrity, Explained: A More Usable Privacy Definition. *IEEE Security & Privacy* 21, 1 (Jan. 2023), 58–65. doi:10.1109/MSEC.2022.3201585

[88] Michelle A Marks, John E Mathieu, and Stephen J Zaccaro. 2001. A temporally based framework and taxonomy of team processes. *Academy of management review* 26, 3 (2001), 356–376.

[89] Kirsten E. Martin. 2012. Diminished or Just Different? A Factorial Vignette Study of Privacy as a Social Contract. *Journal of Business Ethics* 111, 4 (Dec. 2012), 519–539. doi:10.1007/s10551-012-1215-8

[90] Wendy Martinez, Johann Benerradi, Serena Midha, Horia A Maior, and Max L Wilson. 2022. Understanding the ethical concerns for neurotechnology in the future of work. In *Proceedings of the 1st Annual Meeting of the Symposium on Human-Computer Interaction for Work*. 1–19.

[91] Akhil Mathur, Marc Van den Broeck, Geert Vanderhulst, Afra Mashhadi, and Fahim Kawsar. 2015. Tiny habits in the giant enterprise: understanding the dynamics of a quantified workplace. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 577–588.

[92] Helena Mc Elhinney, Marlene Sinclair, and Brian Taylor. 2016. The administration of an online factorial survey using Qualtrics software. In *International Symposium on Factorial Survey 2016*. Brunel University London.

[93] Sara A McComb. 2007. Mental model convergence: The shift from being an individual to being a team member. In *Multi-level issues in organizations and time*. Vol. 6. Emerald Group Publishing Limited, 95–147.

[94] Tobias Mettler. 2024. The connected workplace: Characteristics and social consequences of work surveillance in the age of datification, sensorization, and artificial intelligence. *Journal of Information Technology* 39, 3 (Sept. 2024), 547–567. doi:10.1177/02683962231202535 Publisher: SAGE Publications Ltd.

[95] Andre N Meyer, Gail C Murphy, Thomas Zimmermann, and Thomas Fritz. 2017. Design recommendations for self-monitoring in the workplace: Studies in software development. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–24.

[96] Microsoft. 2025. Microsoft Viva. https://www.microsoft.com/en-us/microsoft-viva. Accessed: 2026-01-23.

[97] Susan Mohammed, Lori Ferzandi, and Katherine Hamilton. 2010. Metaphor no more: A 15-year review of the team mental model construct. *Journal of management* 36, 4 (2010), 876–910.

[98] Phoebe Moore, Lukasz Piwek, and Ian Roper. 2018. The quantified workplace: A study in self-tracking, agility and change management. *Self-Tracking: Empirical and philosophical investigations* (2018), 93–110.

[99] Suvadeep Mukherjee, Verena Distler, Gabriele Lenzini, and Pedro Cardoso-Leite. 2024. Balancing The Perception of Cheating Detection, Privacy and Fairness: A Mixed-Methods Study of Visual Data Obfuscation in Remote Proctoring. In *Proceedings of the 2024 European Symposium on Usable Security*. ACM, Karlstad Sweden, 337–353. doi:10.1145/3688459.3688474

[100] Marjan Naghshbandi, Sharon Ferguson, and Alison Olechowski. 2025. Two Sides to Every Story: Exploring Hybrid Design Teams' Perceptions of Psychological Safety on Slack. *Proceedings of the ACM on Human-Computer Interaction* 9, 7 (2025), 1–35.

[101] Helen Nissenbaum. 2004. Privacy as contextual integrity. *Wash. L. Rev.* 79 (2004), 119.

[102] British Columbia Ministry of Citizens' Services. 2023. *Guide on Using Categorical Race & Ethnicity Variables*. Technical Report. https://www2.gov.bc.ca/assets/gov/british-columbians-our-governments/multiculturalism-anti-racism/anti-racism/anti-racism-hub/anti-racism-stats-and-research/arda_categoricalraceethnicityvariables2023.pdf

[103] Jonas Oppenlaender, Kristy Milland, Aku Visuri, Panos Ipeirotis, and Simo Hosio. 2020. Creativity on Paid Crowdsourcing Platforms. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3313831.3376677

[104] Effy Oz, Richard Glass, and Robert Behling. 1999. Electronic workplace monitoring: what employees think. *Omega* 27, 2 (1999), 167–177.

[105] Soya Park and Chinmay Kulkarni. 2023. Retrospector: Rapid collaborative reflection to improve collaborative practices. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (Sept. 2023), 1–20. doi:10.1145/3610084

[106] Christoph Riedl, Young Ji Kim, Pranav Gupta, Thomas W Malone, and Anita Williams Woolley. 2021. Quantifying collective intelligence in human groups. *Proceedings of the National Academy of Sciences* 118, 21 (2021), e2005737118.

[107] Kat Roemmich, Florian Schaub, and Nazanin Andalibi. 2023. Emotion AI at Work: Implications for Workplace Surveillance, Emotional Labor, and Emotional Privacy. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–20. doi:10.1145/3544548.3580950

[108] Koustuv Saha and Shamsi T Iqbal. 2023. Focus Time: Effectiveness of Computer Assisted Protected Time for Wellbeing and Work Engagement of Information Workers. In *Proceedings of the 2nd Annual Meeting of the Symposium on Human-Computer Interaction for Work*. 1–13.

[109] Johnny Saldaña. 2021. The coding manual for qualitative researchers. (2021).

[110] Helen Sampson and Idar Alfred Johannessen. 2020. Turning on the tap: the benefits of using 'real-life' vignettes in qualitative research interviews. *Qualitative Research* 20, 1 (Feb. 2020), 56–72. doi:10.1177/1468794118816618 Publisher: SAGE Publications.

[111] Samiha Samrose, Daniel McDuff, Robert Sim, Jina Suh, Kael Rowan, Javier Hernandez, Sean Rintel, Kevin Moynihan, and Mary Czerwinski. 2021. MeetingCoach: An Intelligent Dashboard for Supporting Effective & Inclusive Meetings. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–13. doi:10.1145/3411764.3445615

[112] John S. Seberger, Hyesun Choung, Jaime Snyder, and Prabu David. 2024. Better Living Through Creepy Technology? Exploring Tensions Between a Novel Class of Well-Being Apps and Affective Discomfort in App Culture. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1 (April 2024), 22:1–22:39. doi:10.1145/3637299

[113] Yasaman S. Sefidgar, Matthew Jörke, Jina Suh, Koustuv Saha, Shamsi Iqbal, Gonzalo Ramos, and Mary Czerwinski. 2024. Improving Work-Nonwork Balance with Data-Driven Implementation Intention and Mental Contrasting. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1 (April 2024), 74:1–74:29. doi:10.1145/3637351

[114] Mina Sheikhalishahi, Gamze Tillem, Zekeriya Erkin, and Nicola Zannone. 2019. Privacy-Preserving Multi-Party Access Control. In *Proceedings of the 18th ACM Workshop on Privacy in the Electronic Society (WPES'19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3338498.3358643

[115] Don Donghee Shin. 2023. *Algorithms, humans, and interactions: How do algorithms interact with people? Designing meaningful AI experiences*. Taylor & Francis.

[116] Erez Shmueli and Tamir Tassa. 2020. Mediated Secure Multi-Party Protocols for Collaborative Filtering. *ACM Trans. Intell. Syst. Technol.* 11, 2 (Feb. 2020), 15:1–15:25. doi:10.1145/3375402

[117] Reza Shokri, Martin Strobel, and Yair Zick. 2021. On the Privacy Risks of Model Explanations. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Virtual Event USA, 231–241. doi:10.1145/3461702.3462533

[118] Aaron De Smet, Gemma D'Auria, Liesje Meijknecht, Maitham Albaharna, Anaïs Fifer, and Kim Rubenstein. 2024. Go, Teams: When Teams Get Healthier, the Whole Organization Benefits. https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/go-teams-when-teams-get-healthier-the-whole-organization-benefits McKinsey Quarterly.

[119] Sowmya Somanath, Bahare Bakhtiari, and Regan L Mandryk. 2024. Design Fiction on Capturing, Amplifying, and Instilling Happiness in Work. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work*. 1–17.

[120] Jaeyoon Song, Zahra Ashktorab, Qian Pan, Casey Dugan, Werner Geyer, and Thomas W. Malone. 2025. Interaction Configurations and Prompt Guidance in Conversational AI for Question Answering in Human-AI Teams. *Proceedings of the ACM on Human-Computer Interaction* 9, 7 (Oct. 2025), 1–27. doi:10.1145/3757486

[121] Spotify. 2025. Spotify 2025 Wrapped. https://newsroom.spotify.com/2025-wrapped/

[122] Aaron Springer and Steve Whittaker. 2020. Progressive Disclosure: When, Why, and How Do Users Want Algorithmic Transparency Information? *ACM Transactions on Interactive Intelligent Systems* 10, 4 (Dec. 2020), 1–32. doi:10.1145/3374218

[123] Inside Track staff. 2025. A foundation for modern collaboration: Microsoft 365 bolsters teamwork. https://www.microsoft.com/insidetrack/blog/a-foundation-for-modern-collaboration-microsoft-365-bolsters-teamwork/

[124] Jose M. Such and Natalia Criado. 2016. Resolving Multi-Party Privacy Conflicts in Social Media. *IEEE Transactions on Knowledge and Data Engineering* 28, 7 (July 2016), 1851–1863. doi:10.1109/TKDE.2016.2539165

[125] Jose M. Such and Natalia Criado. 2018. Multiparty privacy in social media. *Commun. ACM* 61, 8 (July 2018), 74–81. doi:10.1145/3208039

[126] Jina Suh, Javier Hernandez, Koustuv Saha, Kathy Dixon, Mehrab Bin Morshed, Esther Howe, Anna Kawakami, and Mary Czerwinski. 2023. Towards successful deployment of wellbeing sensing technologies: Identifying misalignments across contextual boundaries. In *2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, 1–8.

[127] Jennifer Jiyoung Suh, Miriam J. Metzger, Scott A. Reid, and Amr El Abbadi. 2018. Distinguishing Group Privacy From Personal Privacy: The Effect of Group Inference Technologies on Privacy Perceptions and Behaviors. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 168 (Nov. 2018), 22 pages. doi:10.1145/3274437

[128] Nathaniel Swinger, Cynthia M. Baseman, Myeonghan Ryu, Saeed Abdullah, Christopher W. Wiese, Andrew M. Sherrill, and Rosa I. Arriaga. 2025. There's No "I" in TEAMMAIT: Impacts of Domain and Expertise on Trust in AI Teammates for Mental Health Work. *Proc. ACM Hum.-Comput. Interact.* 9, 2 (May 2025), CSCW019:1–CSCW019:36. doi:10.1145/3710917

[129] Linnet Taylor and Luciano Floridi. 2017. Group Privacy: New Challenges of Data Technologies. *Group Privacy* (2017).

[130] Kurt Thomas, Chris Grier, and David M. Nicol. 2010. unFriendly: Multi-party Privacy Risks in Social Networks. In *Privacy Enhancing Technologies*, Mikhail J. Atallah and Nicholas J. Hopper (Eds.). Vol. 6205. Springer Berlin Heidelberg, Berlin, Heidelberg, 236–252. doi:10.1007/978-3-642-14527-8_14 Series Title: Lecture Notes in Computer Science.

[131] Roy Van Den Heuvel and Carine Lallemand. 2023. Personal Informatics at the Office: User-Driven, Situated Sensor Kits in the Workplace. In *Proceedings of the 2nd Annual Meeting of the Symposium on Human-Computer Interaction for Work*. 1–13.

[132] Carlota Vazquez Gonzalez, Timothy Neate, and Rita Borgo. 2025. Trusting Tracking: Perceptions of Non-Verbal Communication Tracking in Videoconferencing. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–25. doi:10.1145/3706598.3714306

[133] Viswanath Venkatesh and Fred D. Davis. 2000. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science* 46, 2 (Feb. 2000), 186–204. https://www.proquest.com/docview/213174901/abstract/248604F2988C4200PQ/1 Num Pages: 19 Place: Linthicum, United States Publisher: Institute for Operations Research and the Management Sciences.

[134] Jessica Vitak and Michael Zimmer. 2023. Surveillance and the future of work: exploring employees' attitudes toward monitoring in a post-COVID workplace. *Journal of Computer-Mediated Communication* 28, 4 (2023), zmad007.

[135] Lisa Wallander. 2009. 25 years of factorial surveys in sociology: A review. *Social science research* 38, 3 (2009), 505–520.

[136] Alan F. Westin. 2003. Social and Political Dimensions of Privacy. *Journal of Social Issues* 59, 2 (2003), 431–453. doi:10.1111/1540-4560.00072 _eprint: https://spssi.onlinelibrary.wiley.com/doi/pdf/10.1111/1540-4560.00072.

[137] Microsoft WorkLab. 2023. In the Changing Role of the Office, It's All about Moments That Matter. https://www.microsoft.com/en-us/worklab/in-the-office-it-is-all-about-moments-that-matter

[138] Tianna Xu, Advait Sarkar, and Sean Rintel. 2023. Is a Return To Office a Return To Creativity? Requiring Fixed Time In Office To Enable Brainstorms and Watercooler Talk May Not Foster Research Creativity. In *Proceedings of the 2nd Annual Meeting of the Symposium on Human-Computer Interaction for Work*. 1–12.

[139] Di Yan, Jacky Bourgeois, Yen-Chia Hsu, and Gerd Kortuem. 2025. PAIRcolator: Pair Collaboration for Sensemaking and Reflection on Personal Data. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–20. doi:10.1145/3706598.3713332

[140] Di Yan, Chenge Tang, Senthil Chandrasegaran, and Gerd Kortuem. 2025. Reciportrait: a Data Humanism Approach for Collaborative Sensemaking of Personal Data. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–21. doi:10.1145/3706598.3713300

[141] Qi Yang, Marlo Ongpin, Sergey Nikolenko, Alfred Huang, and Aleksandr Farseev. 2023. Against Opacity: Explainable AI and Large Language Models for Effective Digital Advertising. In *Proceedings of the 31st ACM International Conference on Multimedia (MM '23)*. Association for Computing Machinery, New York, NY, USA, 9299–9305. doi:10.1145/3581783.3612817

[142] Benxin Yin, Hansong Xu, Xinliang Li, Jun Sun, and Hanlin Zhang. 2022. MLPSI: Multi-party Privacy Set Intersection with Linear Complexity. In *Proceedings of the 5th International Conference on Big Data Technologies (ICBDT '22)*. Association for Computing Machinery, New York, NY, USA, 339–343. doi:10.1145/3565291.3565346

[143] Zhiping Zhang, Michelle Jia, Hao-Ping (Hank) Lee, Bingsheng Yao, Sauvik Das, Ada Lerner, Dakuo Wang, and Tianshi Li. 2024. "It's a Fair Game", or Is It? Examining How Users Navigate Disclosure Risks and Benefits When Using LLM-Based Conversational Agents. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–26. doi:10.1145/3613904.3642385

[144] Haoti Zhong, Anna Squicciarini, and David Miller. 2018. Toward Automated Multiparty Privacy Conflict Detection. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM '18)*. Association for Computing Machinery, New York, NY, USA, 1811–1814. doi:10.1145/3269206.3269329

[145] Zoom. 2026. Using Meeting Summary with AI Companion. https://support.zoom.com/hc/en/article?id=zm_kb&sysparm_article=KB0058013

A Pilot Study - Background Information

Please read the following background information which applies to all future scenarios you will be shown in this survey. Imagine you work for a large software company with over 10,000 employees. In this scenario, you are an individual contributor (IC), meaning you have a direct manager who has other managers above them. Your entire company has recently adopted a new tool that uses data from your virtual collaboration tools (e.g., email, chat platforms like Slack or Microsoft Teams, video call recordings, online shared documents, etc.) to measure your team's collaboration practices, and display them over time, on a dashboard. This tool **calculates a metric on how successfully your team is collaborating**. The goal of the system is to summarize these important characteristics, **allowing individuals to reflect on, and improve, their collaboration practices**, which should ultimately lead to more efficient and effective software outputs for your customers. **Only data that is already visible to the entire team** (e.g., an email sent to the entire team, or a shared document all team members have access to) **are used in the measurement. Thus, private emails, messages, or meetings are not used by the tool. Assume that these measurements do not contain errors.**

B Pilot Study - Demographics

Table 7: Pilot Study participant demographics (N = 106).

Category	Option	n	%
Age	18–24	15	14.2%
	25–34	47	44.3%
	35–44	23	21.7%
	45–54	11	10.4%
	55–64	8	7.5%
	65–74	2	1.9%
Gender	Man	50	47.2%
	Woman	50	47.2%
	Non-binary	4	3.8%
	Two Spirit	1	0.9%
	Prefer not to answer	1	0.9%
Race	White	27	25.5%
	Black	67	63.2%
	Latine	5	4.7%
	South Asian	3	2.8%
	East Asian	2	1.9%
	Other	2	1.9%
Years worked	<1	0	0.0%
	1–3	32	30.2%
	3–5	16	15.1%
	5–10	26	24.5%
	10+	32	30.2%
Job level	Not a manager	23	21.7%
	One-level manager	42	39.6%
	Multi-level manager	41	38.7%

C Pilot Study - Concerns and Benefits Scale

Table 8: Pilot Study: Perceived Concerns and Benefits Scale. Measured on a 7-point Likert scale.

Perceived Concerns	Perceived Benefits
I am concerned about how much information could be collected by this tool.	I imagine that this tool could improve my team's performance.
I am concerned about how this tool could increase my team's level of stress.	I imagine that this tool could increase my team's efficiency.
I am concerned about how this tool could weaken trust within my team.	I imagine that this tool could improve my team's collaboration.
I am concerned about how this tool could represent a violation of privacy rights.	I imagine that this tool could increase my team's satisfaction.
I am concerned about how this tool could be misused.	

D Pilot Study - Concerns and Benefits Model Results

Table 9: Pilot Study: Benefits Model results. DM = Direct Manager, IC = Individual Contributor. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Note that reference values differ from the comprehensive study, as reference decisions were based on the most common value in each specific dataset. Each binary indicator (for race, gender, and country) uses non-members as the reference group

Predictor	Reference Category	Comparison Category	β (unstandardized)	Std. Error	Sig.
Measurement Aggregation	Team	Individual	-0.08	0.11	
Access	Team and DM	Only DM	-0.17	0.16	
Access	Team and DM	DM and upwards reporting line	-0.10	0.16	
Access	Team and DM	Everyone	-0.03	0.14	
Access	Team and DM	IC only	0.11	0.18	
Scenario	Team Reflection	External Evaluation	0.10	0.13	
Scenario	Team Reflection	Individual Reflection	-0.26	0.17	
Scenario	Team Reflection	Specific Situation	-0.23	0.13	
Gender	All Others	Male	0.12	0.21	
Race	All Others	White	-0.79	0.25	**
Race	All Others	Latine	-0.65	0.50	
Age		18-24	-0.26	0.32	
Age		35-44	0.26	0.27	
Age		45-54	-0.28	0.38	
Age		55-64	0.44	0.43	
Age		65-74	0.61	0.80	
Manager Level	Not a Manager	Multi-Level Manager	0.54	0.32	
Manager Level	Not a Manager	One-Level Manager	-0.08	0.28	
Years Worked	4-5	1–3	-0.06	0.28	
Years Worked	4-5	10+	-0.24	0.38	
Years Worked	4-5	5–10	-0.56	0.29	

Table 10: Pilot Study: Concern Model results. DM = Direct Manager, IC = Individual Contributor. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Note that reference values differ from the comprehensive study, as reference decisions were based on the most common value in each specific dataset. Each binary indicator (for race, gender, and country) uses non-members as the reference group

Predictor	**Reference Category**	**Comparison Category**	β (unstandardized)	**Std. Error**	**Sig.**
Measurement Aggregation	Team	Individual	0.121	0.104	
Access	Team and DM	Only DM	-0.0473	0.152	
Access	Team and DM	DM and upwards reporting line	-0.000799	0.152	
Access	Team and DM	Everyone	0.0802	0.138	
Access	Team and DM	IC	-0.581	0.175	***
Scenario	Team Reflections	External Evaluation	-0.161	0.130	
Scenario	Team Reflections	Individual Reflection	-0.139	0.161	
Scenario	Team Reflections	Specific Situation	0.0571	0.131	
Gender	All Others	Male	-0.231	0.241	
Race	All Others	White	0.395	0.287	
Race	All Others	Latine	1.32	0.570	*
Age	25-34	18-24	0.136	0.358	
Age	25-34	35-44	0.390	0.306	
Age	25-34	45-54	0.471	0.434	
Age	25-34	55-64	1.35	0.488	**
Age	25-34	65-74	1.16	0.907	
Manager Level	Not a Manager	Multi-Level Manager	1.06	0.361	**
Manager Level	Not a Manager	One-Level Manager	0.412	0.323	
Years Worked	4-5	1–3	0.660	0.318	*
Years Worked	4-5	10+	-0.350	0.426	
Years Worked	4-5	5–10	0.573	0.328	

E Comprehensive Study - Background Information

Imagine you work for a large software company with over 10,000 employees. In this scenario, you are an individual contributor (IC), meaning you have a direct manager who has other managers above them (see Figure). Your entire company has recently adopted a new tool that uses team data to score your team's collaboration practices, and displays them over time, on a dashboard.

F Proportional Odds Assumption Tests

To assess the proportional odds assumption underlying all three CLMMs (benefit, concern, and intention to use), we conducted Brant tests [18] on equivalent regression models fitted without random effects, as the Brant test is not available for mixed-effects ordinal models. Results are summarized below.

Intention to Use Model. All vignette key factor levels satisfied the parallel regression assumption (all $p > .07$), as did the mediating variables of perceived benefit and concern ($p = .18$ and $p = .22$ respectively) and the majority of demographic controls. Demographic violations are attributable to sparse cell sizes.

Concern Model. All vignette key factor levels satisfied the proportional odds assumption (all $p > .25$). Violations were observed only in demographic control variables due to sparse cells, including age categories, PSS score, years worked 10+, and several small country and race cells.

Benefit Model. The majority of vignette key factor levels satisfied the proportional odds assumption. One marginal violation was observed for narrative transparency ($\chi^2 = 16.34$, $p = .01$). All other key factor levels passed cleanly (all other $p > .23$). As with the concern model, violations were concentrated in demographic control variables, due to sparse cells.

Across all three models vignette key factors satisfied the assumption in all or nearly all cases, while failures were concentrated in demographic controls. This consistent pattern, combined with the numerical instability warnings across all three tests, suggests that these violations reflect sparse-cell artifacts. As the primary inferential focus of this work is on vignette key factors, and all but one key factor level satisfies the proportional odds assumption across all three models, we are confident the CLMM results are not meaningfully compromised by violations of assumptions.

G Comprehensive Study - Sensitivity Analysis

While an exact simulation-based sensitivity analysis was not possible with our mixed-effects model, we conducted an approximation by estimating the intraclass correlation for concern/benefit ratings, and computing the effective sample size that accounts for participant fixed-effects. Using this sample size, we applied multiple regression power analysis to determine the smallest partial effect size detectable with 80% power at $\alpha = 0.05$. This sensitivity analysis provides an approximate minimal detectable effect for a single fixed effect in the mixed model, assuming the estimated intraclass correlation and observed variance structure, and does not reflect power for any individual predictor. This analysis revealed that our study had 80% power to detect an effect of approximately $r_{partial} = 0.101$ in the benefits model (corresponding to 0.18 points on a 1-7 Likert scale) and $r_{partial} = 0.104$ on the concerns model (corresponding to 0.19 points on the Likert scale). As such, even small effect sizes can be detected with our sample size (i.e., substantially smaller than a conventional "small" effect).

H Comprehensive Study - Manager Sensitivity Analysis

To assess whether the overrepresentation of managers in our sample biased our main findings, we re-ran all three models on the full sample, adding interaction terms between manager status (binary: manager vs. non-manager) and each level of each vignette key factor. Table 11 reports the significant interaction terms across all three models. Managers found only manager plus access to be more concerning than non-managers, and managers were less likely to use the system when there was a view for managers and a different view for ICs, and they were also less likely to use the system when one has total control to opt out of data collection. Likelihood ratio tests showed that adding manager × key factors did not significantly improve model fit for any model ($p = 0.66$ for benefits, $p = 0.30$ for concerns, $p = 0.20$ for intention to use).

To further validate these findings, we re-ran all models stratified by managerial status. In both subgroups, benefit was a strong positive predictor of intention to use and concern was a strong negative predictor, confirming that the privacy calculus framing holds across managerial levels. The results of the stratified benefit models were overwhelming similar to the overall model—the only difference being that managers plus access was a significantly negative predictor only for managers. For the concern model, non-managers showed few significant vignette factor effects on concern (only narrative explanations), suggesting non-managers rely more on individual privacy orientation than specific design choices when evaluating concerns. Control also became less significant in the concern model for both managers and non-managers, which may be due to the smaller sample sizes.

Table 11: Significant Interaction Terms — Manager Status × Vignette Factors. $^*\ p < .05$. No significant interactions were found in the benefits model.

Model	Interaction Term	β	SE	Sig.
Concerns	Manager × DM+ Access	0.78	0.39	*
Intent	Manager × IC+DM Access	-1.13	0.55	*
Intent	Manager × Total Control	-0.79	0.39	*

I Complete Bootstrapped Mediation Path Results

Table 12: Full mediation analysis for all vignette factors. Total and direct effects are from cumulative link mixed models; indirect effects are bootstrapped (1,000 resamples).

Predictor	**Total effect**		**Direct effect**		**Indirect effect (total)**		**Std. indirect**	**Mediation**
	β_{total}	p_{total}	β_{direct}	p_{direct}	95% CI lower	95% CI upper	$\boldsymbol{\beta}_{\text{ind}}$	
Individual Aggregation	0.20	0.17	-0.05	0.73	-0.02	1.43	0.13	No
Only Manager Access	0.23	0.20	0.07	0.71	-0.41	1.39	0.07	No
Manager Plus Access	-0.37	0.04	-0.04	0.81	-1.96	-0.09	-0.15	Full
Only IC Access	0.29	0.19	-0.05	0.82	-0.23	1.98	0.10	No
IC and Manager Access	0.22	0.35	0.26	0.27	-1.20	1.31	0.00	No
Data Messages	-0.34	0.03	-0.15	0.35	-1.67	-0.01	-0.14	Full
Data Work Outputs	-0.19	0.20	-0.08	0.63	-1.38	0.28	-0.10	No
Transparency Raw Data	0.78	0.00	-0.04	0.82	1.52	3.22	0.42	Full
Transparency Narrative	0.92	0.00	0.06	0.73	1.68	3.56	0.47	Full
Some Opt-Out Control	0.23	0.13	0.01	0.94	0.12	1.80	0.16	No
Total Opt-Out Control	0.76	0.00	0.39	0.02	0.97	2.62	0.31	Partial