School of Computation, Information and Technology – Informatics

Technical University of Munich

Dissertation in Informatics

Designing People Analytics for Data Sovereigns

Valentin Zieglmeier

Technische Universität München
TUM School of Computation, Information and Technology

Designing People Analytics for Data Sovereigns

Valentin J. J. Zieglmeier

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology der Technischen Universität München zur Erlangung eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: apl. Prof. Dr. Georg Groh

Prüfende der Dissertation:

1. Prof. Dr. Alexander Pretschner
2. Prof. Dr. Jens Großklags
3. Prof. Dr. Thomas Hess

Die Dissertation wurde am 08.12.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Computation, Information and Technology am 05.07.2024 angenommen.

Danksagung

Mein größter Dank gilt meinem Doktorvater Alexander Pretschner. In den Jahren meiner Promotion habe ich ihn immer als Unterstützer und Förderer wahrgenommen. Er gab mir die Freiheit, meine Forschung und Lehre selbst zu gestalten, ohne mich bei Problemen oder Schwierigkeiten alleine zu lassen. In unseren vielen Diskussionen forderte er mich immer wieder heraus und half mir dabei, fachlich fundiert für meine Ideen zu argumentieren. Ich hatte dabei stets das Gefühl, dass er ernsthaft an meiner Forschung interessiert war. Außerdem danke ich Jens Großklags und Thomas Hess, dass sie sich bereiterklärt haben, als zusätzliche Prüfer meine Dissertation zu begutachten.

Ein weiterer wichtiger Unterstützer meiner Promotion war mein Mitarbeiter Gabriel Loyola Daiqui. Seit seiner Teilnahme an meinem ersten Seminar wusste ich Gabriel an meiner Seite und konnte bis zum Ende der Promotion auf ihn zählen. Er war eine wichtige Stütze und hat mich in kritischen Situationen entscheidend entlastet. Auch fachlich war er für mich ein wertvoller Partner.

Wichtig ist mir, explizit auch meinen Koautor:innen zu danken. Diese Dissertation basiert auf Publikationen, die ohne die Mitwirkung von Alexander Pretschner, Gabriel Loyola Daiqui, Antonia Maria Lehene und Maren Gierlich-Joas nicht existieren würden. Sie haben mich bei der Entstehung der Manuskripte entscheidend unterstützt und waren eine große Hilfe. Zusätzlich danke ich meinen Korrekturleser:innen Gabriel Loyola Daiqui, Julia Schuller, Severin Kacianka, Markus Schnappinger, Nicola Kolb und Daniel Elsner.

Darüber hinaus haben mir auch Amjad Ibrahim und Julia Schuller inhaltlich geholfen. Mein Mentor Amjad war vor allem in meiner Anfangsphase behilflich dabei, mich thematisch sinnvoll auszurichten. Dank seiner Erfahrung hat er mich außerdem strategisch beim Platzieren meiner Forschung beraten. Julia hat in ihrer Zeit am Lehrstuhl intensiv mit mir zusammengearbeitet. Sie brachte großes Wissen über mein wichtigstes Forschungsfeld mit und hat mich deshalb beim Framing und der Suche nach passenden Veröffentlichungszielen entscheidend unterstützt.

Die Promotion kann sehr isolierend sein. Deshalb waren meine aktuellen und ehemaligen Kolleg:innen am Lehrstuhl für mich nicht nur bei fachlichen Diskussionen, sondern auch als moralische Unterstützung wichtig. Dank ihnen war ich nicht alleine mit scheinbar unüberwindbaren Problemen und konnte immer mit jemandem über meine Schwierigkeiten sprechen. Besonders gilt das für Severin Kacianka, Amjad Ibrahim, Julia Schuller, Ana Petrovska, Daniel Elsner, Nicola Kolb, Claudius Jordan, Stephan Lipp, Markus Schnappinger, Florian Hauer und Patrick Stöckle.

Dank gilt auch meinen Kolleg:innen im Forschungsprojekt ‚Inverse Transparenz' für die interessante und gute Zusammenarbeit. Die Arbeit im Projekt gab mir wertvolle Impulse für meine Dissertation und half mir, den praktischen Nutzen der erzielten Ergebnisse direkt zu sehen, was sehr motivierend für mich war. Ich bin stolz auf das, was wir gemeinsam erreicht haben.

Abschließend will ich meiner Familie und meinen Freunden danken. Je tiefer ich in die Promotion eingetaucht bin, desto abstrakter und schwerer nachvollziehbar wurden meine Probleme. Trotzdem wart ihr immer für mich da und habt mir das Gefühl zu geben, nicht alleine zu sein. Danke, dass es euch gibt.

Kurzzusammenfassung

Kontext. Der Arbeitsplatz wird zunehmend digitalisiert, wodurch mehr Daten über Arbeitsabläufe verfügbar werden. Sogenannte *People Analytics* (PA) können diese Daten nutzen, um Beschäftigte besser zu verstehen und zu führen. Sie versprechen beispielsweise die Entscheidungsfindung zu erleichtern, da evidenzbasierte Erkenntnisse zugrunde gelegt werden können. Im Mittelpunkt dieser Dissertation steht dabei die Situation in Deutschland, wo der Einsatz von PA in der Regel das Einverständnis der analysierten Mitarbeitenden erfordert.

Herausforderungen. Allem voran nutzen PA-Analysen potentiell sensible Daten, die missbraucht werden können, und ihre Ergebnisse können schwerwiegende Konsequenzen für das Individuum haben. Außerdem werden PA mit zunehmendem Reifegrad immer undurchsichtiger für diejenigen, die von ihnen analysiert werden. Und schließlich zielen PA, obwohl sie Risiken für Mitarbeitende mit sich bringen, darauf ab, Vorteile für das Unternehmen zu bieten. Sie geben dem Individuum also keine hinreichenden Gründe für die Weitergabe seiner Daten.

Lösung. Wir glauben, dass das Design von PA Mitarbeitende, die analysiert werden, als zentrale Stakeholder betrachten sollte. Sie sollten als *Datensouveräne* verstanden werden, die davon überzeugt werden müssen, ihre Daten zu teilen. Konkret empfehlen wir: (1) Um Rechenschaftslegung und fundierte Datenfreigabeentscheidungen zu ermöglichen, sollte Mitarbeitenden die Nutzung ihrer Daten sichtbar gemacht werden. (2) Um Gründe für die Datenfreigabe zu schaffen, sollte die Attraktivität von PA für Mitarbeitende, die von ihnen analysiert werden, erhöht werden.

Forschungslücke. Die meisten verwandten Arbeiten betrachten ähnliche Forschungsprobleme im Konsumentenkontext, der sich vom Arbeitsplatzkontext von PA unterscheidet. Deshalb können sie zwar als Inspiration dienen, die Lösungen lassen sich aber nicht leicht übertragen. Die verbleibenden Arbeiten sind entweder nur konzeptionell, führen keine empirischen Studien zur Validierung durch, sind in Bezug auf die rechtlichen Anforderungen in Deutschland technisch limitiert oder legen keine systematischen Schritte dar, wie unser Ziel erreicht werden kann.

Beiträge. Wir stellen vier Beiträge vor, die sich mit den festgestellten Forschungslücken befassen. *Erstens*: Mit unserem Konzept von *Inverser Transparenz ‚by Design'* zeigen wir auf, wie Datennutzungstransparenz in PA erreicht werden kann. Unsere zwei empirischen Studien belegen, dass der Ansatz realisierbar ist und positiv aufgenommen wird. *Zweitens*: Mit unserer Lösung für *dezentralisierte Inverse Transparenz* zielen wir darauf ab, Transparenz auch über Datenzugriffe auf sensible Daten zu gewährleisten, die auf persönlichen Geräten gespeichert sind. In unserer Analyse zeigt sich unser Ansatz sicher gegenüber möglichen Angriffen und unsere technischen Messungen ermitteln ausreichende Leistung und lineare Skalierbarkeit. *Drittens*: Unsere Taxonomie zu Maßnahmen für

vertrauenswürdiges Benutzeroberflächendesign kann angewandt werden, um die Vertrauenswürdigkeit eines Transparenzdashboards zu erhöhen. Ein solches Dashboard macht Datennutzung für das Individuum sichtbar und ist deshalb ein wichtiger Faktor, um das Vertrauen in die gewährte Transparenz zu stärken. *Viertens*: Unsere Taxonomie zu *attraktivitätssteigernden Strategien für People Analytics* ergänzt unsere Anstrengungen, die Risiken von PA zu reduzieren, indem konkrete Wege aufgezeigt werden, wie die Attraktivität für alle Mitarbeitenden individuell erhöht werden kann. Die identifizierten Dimensionen von PA-Attraktivität decken sowohl die Design- als auch die Nutzungsphase ab. Insgesamt stellen unsere Beiträge konkrete Schritte in Richtung der Gestaltung von People Analytics für Datensouveräne dar.

Abstract

Context. The workplace is becoming increasingly digital, leading to more available data on work activities. *People analytics* (PA) can use these data to help understand and guide employees. PA promise to improve, e.g., decision-making by providing evidence-based insights. We consider the situation in Germany, where most applications of PA require the acceptance of employees under analysis.

Challenges. PA analyses may utilize potentially sensitive data, which could be misused, and their results can have severe consequences for the individual employee. Furthermore, with increasing maturity, PA become more opaque to those subjected to their analysis. Finally, even though PA implicate risks for employees, they are focused on providing organizational benefits and do not give the individual sufficient reasons to share their data.

Solution. We believe the design of PA should consider the employees under analysis as core stakeholders. They should be understood as *data sovereigns* who must be convinced to share their data. Concretely, we suggest (1) enabling accountability and informed data sharing decisions by providing employees visibility into how their data are used and (2) giving reasons to share by increasing PA appeal for employees under analysis.

Research Gap. Most related works consider similar research problems in the consumer context, which differs from the workplace context of PA. Therefore, they can serve as inspiration, but the solutions cannot be easily transferred. The remaining works are either only conceptual, lack empirical studies of their suitability, are technically limited given the legal requirements in Germany, or do not provide systematic guidance on how to achieve our goal.

Contributions. We present four contributions that address the identified research gaps. *First*, our concept of *inverse transparency by design* presents a path to achieve data usage transparency in PA. In two empirical studies, we find that it can be practical and is positively received. *Second*, with our solution for *decentralized inverse transparency*, we aim to ensure transparency over accesses to sensitive data stored on individuals' personal devices. Our analysis shows that our tool is secure against expected attacks, with our benchmarks revealing adequate performance and linear scalability. *Third*, our taxonomy of *trustworthy user interface design* identifies measures that can be applied to improve the trustworthiness of a transparency dashboard. Such a dashboard makes data usages visible to the individual, meaning it can influence if the provided transparency is trusted. *Fourth*, our taxonomy of *appeal strategies for people analytics* complements our efforts to reduce risks in PA by presenting concrete ways to also increase the appeal for individual employees. We identify appeal dimensions both for the design and the usage phase of PA. Overall, our contributions present concrete steps toward designing people analytics for data sovereigns.

Outline

Chapter 1: Introduction

This chapter introduces the topic by outlining the context, carving out problems, summarizing the identified research gaps, and presenting our contributions towards closing them.

Chapter 2: Background

This chapter provides a high-level introduction to the fundamentals and theories relevant for our thesis and defines the concepts we work with. It serves to delineate our research and ground it in theory.

Chapter 3: Related Work

This chapter discusses existing research and identifies the research gaps that we address in our contributions. Thereby, it embeds our work in the state of the art.

Chapters 4 to 7: Conceptual and Technical Solutions

These four chapters comprise our contributions, which consist of conceptual and technical solutions. As part of this publication-based thesis, each contribution was previously published as an individual research paper.

Chapter 8: Limitations

This chapter discusses overarching limitations of our thesis, including of our conceptual goal, technical approaches, and empirical studies.

Chapter 9: Summary and Outlook

This chapter concludes our thesis with a summary and avenues for future work.

Contents

List of Acronyms

AI	artificial intelligence	3
APCO	Antecedents → Privacy Concerns → Outcomes	19
API	application programming interface	25
CFIP	concern for information privacy	17 f.
GDPR	General Data Protection Legislation	4, 8, 10, 13, 15, 27, 29 ff., 70, 131
HR	human resources	3, 5 f., 10
IUIPC	Internet users' information privacy concerns	18
JAR	Java archive file	25
M&R	Markowitch and Roggeman	28, 31
PA	people analytics	vii ff., 3–10, 13–16, 18, 21, 24 ff., 33 f., 37 f., 99, 109, 127 ff., 131 f.
REST	representational state transfer	25
SGX	Software Guard Extensions	27
TETs	transparency-enhancing technologies	25
TTP	trusted third party	27–31, 69 f.
UI	user interface	8 ff., 99, 128
WPC	workplace privacy concerns	7, 18 f.

Part I.

Introduction, Background, and Related Work

1. Introduction

This chapter introduces the topic by outlining the context, carving out problems, summarizing the identified research gaps, and presenting our contributions towards closing them.

1.1. Context and Motivation

With increasing digitalization, data generation and processing have become ubiquitous. Digital tools are used for most activities, from listening to music or watching movies to communication or navigation. This change encompasses not just private life but specially the workplace. Pervasive digital tools lead to a more transparent work environment—every action becomes potentially traceable—that can increase productivity and facilitate collaboration. Most activities in companies today are automated or supported by digital tools, starting from hiring, covering internal communication and meeting organization, and increasingly including even core work activities [11, 97, 197]. In this environment, *people analytics* (PA), also known as HR analytics, are increasingly adopted to understand and guide employees. These tools enable data-driven management of the workforce. They can utilize various data sources, including internal HR data or employee surveys, but also company-external HR data, e.g. from social networks, or activity and health data, e.g. from wearable sensors [49, Tab. 4, 151, p. 294].[1] PA can offer valuable insights into work processes and have the potential to improve decision-making, productivity, and job satisfaction [90, Sec. 2.2.1, 151, p. 295]. For example, analyzing employee relationships has been claimed to help identify efficient or innovative teams [141]. Having access to such evidence-based insights from PA improves the adoption of changes, even if they just confirm conventional wisdom [93]. Accordingly, companies increasingly consider and introduce such tools [141, p. 73, 224, pp. 901 f.].

Even with many valuable use cases of PA, various risks for employees under analysis exist. Depending on the used data and generated insights, PA can lead to more invasive tracking and surveillance of employees [224, pp. 907 f.]. Their perceived objectivity threatens to oversimplify complex situations by providing seemingly objective and unambiguous numbers [90, Sec. 3.1]. Yet, PA are inherently subjective, as they are limited by the available data [see, e.g., 82, p. 8] and affected by human biases [173, p. 1136]. And especially those use cases that promise to be most powerful, such as AI-based analyses or the use of wearable sensors, can be ethically questionable [151, pp. 294 f.]. As a consequence, employees show reactance and resistance against PA projects [90, 125] and lose trust in their employer [172, 224]. This takes the form of protesting [224, p. 910], not sharing relevant data, or (if forced to share) not providing data truthfully [224, p. 909] and gaming the system [113]; [224, p. 909]. Where possible, e.g. through the workers' council,

[1] For a more detailed introduction to PA, see Section 2.2.

PA initiatives are sometimes even completely blocked [175]. In the long term, employees' commitment to the company can be reduced [34, p. 1037]; [125] and they may react with counterproductive behavior [90, p. 418], which both negatively affect productivity and competitiveness of the organization. Combined, these issues lead to reduced adoption and use of PA. A recent survey of *Insight222* found that 81% of 57 surveyed global companies considered their PA projects at risk due to data ethics and privacy concerns [175]; [224, p. 909]. Microsoft, a large provider of PA, has even started working directly with their workers' council to try to increase acceptance of their tools [76].

The legal context of this thesis is Germany, where the General Data Protection Legislation (GDPR) has been the status quo since 2018. It grants individuals additional rights to protect their privacy. These rights extend to the workplace, where they are complemented by works agreements. In these agreements, the usage of employee data for behavior control and performance assessment (covering most use cases of PA) can be limited or completely forbidden.[2] Consequently, PA projects require the acceptance of individual employees to be usable in practice. Below, we analyze the status quo from the employees' perspective and identify problems that currently limit their acceptance.

1.2. Problems

To understand why employees resist PA and how this could be improved, we consider their perspective. Thereby, we determine three concrete problems in the design of PA today that we address with our research contributions. In the following, we derive and explain each identified problem individually before providing an overarching summary.

1.2.1. Increased Risks

To start with, PA base their analyses on various data that may be considered sensitive by individuals [90, p. 417]. For example, sentiment analysis is used to identify emotions in texts produced by the employee [71, Tab. 3]. Data from wearables allow assessment of the individual's physical health and activity [149, p. 546]. And even seemingly harmless employee surveys can include questions that allow personality classification [224, pp. 905–906]. Such analyses can be easily, even unintentionally, misused to discriminate against protected groups. Beyond using sensitive data, the insights that PA produce can have significant consequences for employees. Automated recruitment tools have been found to include biases, which can mean that a fitting candidate is never invited for an interview based on, e.g., their gender [224, p. 905]. Experiencing constant tracking and surveillance during their work can lead to employees feeling controlled, impeding their autonomy [90, p. 418]. Using the wrong measures for performance or productivity can warp incentives and has been shown to reduce morale [121]. In extreme cases, the insights from PA may lead to employees losing their job [122]. Combined with their use of sensitive data that can be misinterpreted or misread, these potentially severe consequences make data misusage[3] especially problematic. We summarize this as an elevated *risk of data misusage* in PA.

[2] We analyze the legal context of PA in Germany in more detail in Section 2.2.3.

[3] We define *data misusage* as any usage of data that is *unexpected* or *unintended* by the data owner [see 1, p. 2].

1.2.2. Reduced Transparency

Yet, the data processing of PA is often opaque to those subjected to it [82, Sec. 3.1, 90, Sec. 3.4]. PA have developed in maturity from simple descriptive and predictive systems towards being prescriptive and even acting autonomously (see Section 2.2.1). Especially in more advanced PA, the insights they produce can be based on logic too complex even for their operators[4] to understand [72, p. 2, 82, p. 5]; [90, Sec. 3.4]. The seeming superiority of algorithms makes it harder to argue against recommendations from these systems [140, p. 384]. In combination with the tendency of humans to defer to automated machines [235, p. 121], recommendations or even autonomous decisions by advanced PA are less likely to be critically examined by their users. Because of this, PA operators may be naturally incentivized to reduce transparency further with increasing reliance on technology, shielding themselves from critique or oversight that may question their role [17, p. 980]. For employees under analysis, this experienced "black box" is especially paradoxical given that, while the decision-making becomes increasingly opaque to them, they are expected to be *more* transparent to enable PA [13, p. 2]. This *lack of transparency* in PA means that potential data misusage or wrong analyses may be hard to uncover. Therefore, it exacerbates the risk of data misusage and hinders the accountability of PA operators.

1.2.3. No Reasons to Share

Even though PA can entail increased risks for employees, most available tools are focused on providing organizational benefits and do not consider employee-oriented benefits [see, e.g., 223, p. 229]. The available tools predominantly address managers or analysts, providing high-level analyses and promising to support macro decisions [108, Sec. 5]. This is also how they are marketed [see, e.g., 198] and used in practice [see, e.g., 163]. Even in PA that offer features targeting the self-improvement of employees, these features are a separate function or a completely different tool from the organizational analyses [108, Sec. 5.4]; [154]. No connection is made between the organizational value of PA and the individual's contribution of data that enables it. We summarize this third issue as a *missing appeal* for employees. Thus, even if their concerns are addressed, employees are not provided reasons to contribute their data voluntarily.[5] While understandable on an individual level, this is not an ideal situation, though. PA depend on personal data to provide analyses, meaning they are less effective if individual employees choose not to share their data. Thereby, useful and beneficial applications of PA are also prevented.

1.2.4. Summary

To summarize, we identify **three problems** for employees in the current design of PA: **risk of data misusage** (**I**), **lack of transparency** (**II**), and **missing appeal** (**III**). We argue that these problems can be addressed by incorporating the employee perspective into the design of PA.

[4] We refer to all actors that utilize PA to generate insights as *PA operators*. Mostly, these are employees in general HR or management positions. Some companies have PA-specific positions, though [see 4, Tab. 1].

[5] The underlying *privacy calculus* is explained in Section 2.4.1.

1.3. Solution

To enable sensible uses of PA while reducing potential privacy concerns and preventing data misusage, we consider it necessary to rethink the design of PA, incorporating the employee perspective. This covers two main aspects: (1) enabling accountability and informed data sharing decisions (problems **I** and **II**) by providing *inverse transparency*,[6] and (2) increasing employee appeal (problem **III**) to give reasons to share. We think these steps enable individuals to act as *data sovereigns* who are self-determined with regard to how their data may be used (see Section 2.3). By systematically shifting PA design towards incorporating data sovereigns as core stakeholders, all three problems can be addressed.

> **Solution.** Design people analytics for data sovereigns by providing inverse transparency and increasing employee appeal.

1.4. Research Directions and Gaps

From our proposed solution, we derive four concrete research directions (Sections 1.4.1 to 1.4.4) in the following. We then identify six research gaps (denoted by **G1–G6**) that we address in our contributions.

1.4.1. Reducing Risks by Increasing Transparency

We think that PA should be designed to provide inverse transparency. Granting such transparency over data usages can create accountability and reduce the risk of data misusage. In our review of related literature (see Section 3.1), we find two gaps that concern this goal (**G1** and **G2**):

G1: Inverse Transparency in People Analytics. Existing research on data usage transparency mainly considers the consumer context, which differs from the workplace context of PA [e.g., 23, 27, 102], or it is only conceptual [e.g., 28, 179, 231]. We lack a solution approach to achieving inverse transparent PA.

G2: Empirical Studies of the Developer and User Perspective. To assess the suitability of inverse transparent PA, empirical studies are necessary. Importantly, they should cover both the PA developer and the employee perspective, as these are relevant stakeholders. This necessity has been identified in related works [e.g., 27, 73, 178]. Yet, no previous studies for the workplace context exist.

1.4.2. Transparency for Sensitive Data

Most data used for analysis in PA are generated or stored in company-controlled systems, e.g. internal HR data [see 49, Tab. 4] or digital communication traces [see 108, p. 246]. We consider the retroactive accountability afforded by inverse transparency sufficient to

[6] *Visibility into how one's data are used* (see Section 2.5).

protect these data, as it promises to deter misusage. Yet, a different class of data has been identified for use in PA: sensitive data such as health or genetic data, especially those collected by wearable devices [see 37, 153]. These are almost exclusively *primary data*, meaning they are gathered by the employees themselves and therefore fall under their personal governance [see 133, pp. 33 f.].

The analysis of sensitive data poses additional risks for employees [see, e.g., 14]. Consequently, employees' willingness to share decreases significantly for data perceived as more sensitive [220, Fig. 4]. Yet, in specific usage scenarios, analysis of, e.g., health data can be beneficial for the employee, such as by detecting work-related stress or reducing harmful sedentary behavior [see 153, Tab. 1]. We argue normatively: The processing of sensitive data for such cases should not require employees to give their employer control over them. Instead, we think they should keep these data on their own devices, akin to the idea of *owner-controlled information* [85]. If a PA operator wants to access sensitive data for analysis, they should have to request them from the affected employee and personally guarantee their appropriate handling.

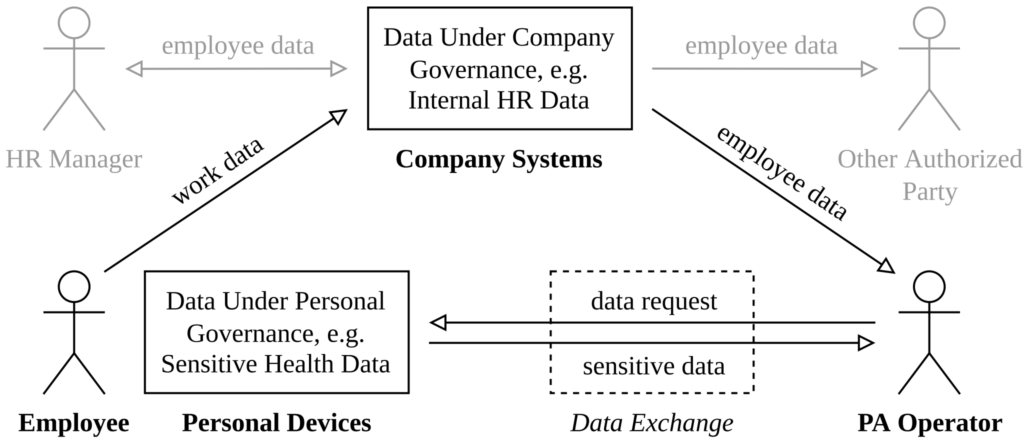


Figure 1.1.: Differentiating the flow (represented by arrows) of employee data under company governance and sensitive data under personal governance. Colored in black are the main actors and data flows as part of PA usage. Beyond these, stakeholders such as HR managers or other authorized parties can interact with employee data stored under company governance (colored in gray). We argue that sensitive data should not have to be made similarly accessible. Instead, they should only be directly shared with the PA operator on request.[7]

Sending data directly to the PA operator means that we minimize who they are exposed to. Thereby, we can reduce the risk of *improper access*[8] to employees' sensitive data. Yet, removing the intermediary introduces *plausible deniability* (see Section 2.7), allowing both parties to *repudiate* their involvement in the *data exchange*: Suppose that an employee's sensitive data were misused after they shared them, so they want to make use of inverse transparency and hold the PA operator accountable. That means they want to prove that

[7] Technically, data stored on the PA operator's system are also under company governance, depending on definition. Importantly, though, significantly fewer parties are authorized to access them.

[8] A dimension of the *workplace privacy concerns* (see Section 2.4.2).

this PA operator had been granted access to their misused sensitive data. Yet, the recipient can simply claim to have never received the data, e.g. due to a network error. Consequently, they could not have been responsible for the misusage. And as no third party witnessed their exchange, either side could believably lie. This endangers our goal of accountability. [see also 15, Sec. 2, 130, Sec. 1]

To counteract this issue, the data exchange between employee and PA operator needs to be recorded and stored securely in a way that prevents either party from repudiating their involvement. To that end, we *first* require *non-repudiable data exchange*, creating undeniable evidence both that the employee sent the data as well as that the PA operator received them.[9] And, *second*, we want to record the created evidence in a *decentralized transparency log* so that no single actor can forge or retroactively purge it. There are solutions for the underlying problems in related work (see Section 3.2), but they do not suffice. We find two gaps (**G3** and **G4**):

G3: Algorithms for Non-repudiable Data Exchange. As introduced above, to be able to prove that data were successfully shared directly between two parties even if, e.g., the recipient denies it, we require non-repudiable data exchange [see 130]. There are applicable protocols [namely 150, 155], but we lack algorithms to implement them in practice.

G4: GDPR-Compliant Decentralized Transparency Log. After the data exchange, we want to securely store a log of this exchange for later retrieval, enabling trusted transparency [78, 190]. We aim for a decentralized solution for the scenario of sensitive data so that no party can unilaterally manipulate or remove log entries. Existing approaches are not applicable to our use case of transparency logs [e.g., 96, 116], are not GDPR-compliant [e.g., 207, 221], or weaken the provided security guarantees to achieve compliance [e.g., 58, 74].

1.4.3. Fostering Trust Through User Interface Design

For inverse transparency to create accountability, employees need to make use of the technical infrastructure that provides them with transparency. Only then can there be a reasonable expectation that data misusage could be uncovered. And user trust is one important facet influencing technology acceptance and use [137]. In our other contributions, we therefore work towards providing functionally comprehensive and technically secure solutions for inverse transparency, which can foster trust in the correctness of the collected information [see 48, Sec. 4.3]. Yet, this does not suffice, as the user-facing tool that provides transparency also needs to be trusted by the individual. And, as previous studies have found, facets of user interface design can influence user trust in a software tool [e.g., 162, 188]. In our review of related literature (see Section 3.3), we find a gap (**G5**):

G5: Trustworthy User Interface Design. Existing research on trust in software provides no comprehensive guidance for the design of trustworthy user interfaces. It either does

[9] The property of *non-repudiation* and the required evidences for *non-repudiable data exchange* are more precisely defined in Section 2.8.

not contain general guidance by focusing on specific use cases [e.g., 22, 70], does not follow a systematic approach [e.g., 48, 211], or does not present concrete measures to improve user interface trustworthiness [e.g., 79, 211].

1.4.4. Motivating Employees to Share Their Data

Increasing transparency and reducing the (perceived) risks in PA can increase employees' willingness to share, but it does not suffice. The individual's privacy calculus additionally incorporates the perceived benefits of sharing (see Section 2.4.1). In some cases, intrinsic motivation [see 196] or organizational commitment [see 157] may suffice without a direct benefit. If not, appeal strategies need to be applied [e.g., 112, pp. 166 f.]. In our review of related literature (see Section 3.4), we find a gap (**G6**):

G6: Employee Appeal of PA. Most existing research on appeal strategies focuses on the consumer context [e.g., 47, 159, 229], meaning it cannot be easily transferred to the workplace. To our knowledge, no guidance on increasing the employee appeal of PA exists.

1.5. Contributions

We contribute to closing the identified gaps in four ways (see also Figure 1.2).

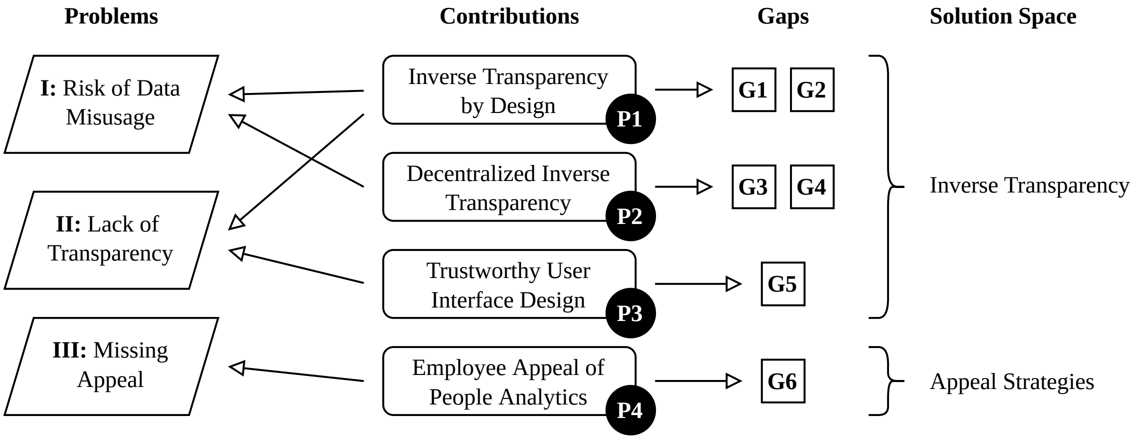


Figure 1.2.: Big picture relating problems, contributions, and gaps addressed in this thesis.

First, addressing **G1** and **G2**, we contribute the concept of *inverse transparency by design* and two empirical studies on it (Chapter 4). To attend to the risk of data misusage (problem **I**) and lack of transparency (problem **II**) in PA, we propose to rethink their data flow design. The core principle of our proposal is that PA tools should be designed so that whenever data on an individual are used, that usage is made visible to them. We describe the concept, outline the potential implications on software development, and evaluate it in two exploratory studies: one covering the developer perspective, the other the perspective of users. We find that developers could integrate the required data usage monitoring with minimal changes to their source code. Data owners perceive the concept as useful and helpful and consider it an improvement upon the status quo.

Second, addressing **G3** and **G4**, we contribute a concept, implementation, and evaluation for fully *decentralized inverse transparency* (Chapter 5). This addresses the risk of data misusage (problem **I**) and can specifically be applied to sensitive data such as health data. These data are typically excluded from PA analyses and require additional protection. To enable selected beneficial applications of sensitive data, we propose that employees keep their sensitive data on their devices and only make them directly available to the PA operator on request. Ensuring that their data exchange is recorded securely is challenging, as we do not want to depend on a potentially untrusted intermediary to facilitate the interaction. With our Kovacs data exchange and logging system, we show how data access transparency can be realized without necessitating the involvement of a third party. Furthermore, we describe an algorithm that ensures GDPR compliance regarding the right to erasure of stored logs.

Third, addressing **G5**, we contribute a taxonomy of *trustworthy user interface design* factors based on a systematic review of the literature (Chapter 6). Just providing access to a log of data usages alone is not enough to tackle the lack of transparency (problem **II**) in PA. We need to consider the trustworthiness of the provided transparency dashboard for individuals, which influences their intention to use such a tool. Therefore, we have systematically analyzed the literature and present a taxonomy of factors that influence user interface trustworthiness in general. Derived from these, we present concrete measures that can be taken when developing a tool. As an additional contribution, we apply exemplary measures to our transparency dashboard in a proof of concept implementation and evaluate it in a preliminary study. We find that the applied measures can be effective in fostering user trust towards the transparency dashboard in our context.

Fourth, addressing **G6**, we contribute a taxonomy of *appeal strategies for people analytics* (Chapter 7), encompassing types and concrete examples. These alleviate the missing employee appeal (problem **III**). Privacy calculus theory suggests that the disclosure decision is made based on perceived risks but also perceived benefits of sharing (see Section 2.4.1). While the *risks* of data sharing are attended to by our previous contributions, not all use cases of PA lead to immediate *benefits* for individuals. In cases where data benefit the company as a whole but not the individual directly, they may not be sufficiently motivated to contribute their data. To tackle this, we have researched appeal strategies to increase employees' willingness to share. Based on our literature review and interviews with HR practitioners, we present a taxonomy of concrete appeal strategies. From the found examples, we inductively derive the dimensions *values*, *benefits*, and *incentives*, covering both the design and usage phase.

1.6. Core Publications

As part of this publication-based doctoral thesis, our contributions have been published in peer-reviewed journals (**P1** and **P2**) or conference proceedings (**P3** and **P4**). The included core publications are:

P1 **Valentin Zieglmeier** and Alexander Pretschner. "Rethinking People Analytics With Inverse Transparency by Design." *Proceedings of the ACM on Human-Computer Interaction* 7.CSCW2, Article 292 (2023), 29 pages. DOI: `10.1145/3610083`.

P2 **Valentin Zieglmeier**, Gabriel Loyola Daiqui, and Alexander Pretschner. "Decentralized Inverse Transparency With Blockchain." *Distributed Ledger Technologies: Research and Practice* 2.3, Article 17 (2023), 28 pages. DOI: 10.1145/3592624.

P3 **Valentin Zieglmeier** and Antonia Maria Lehene. "Designing Trustworthy User Interfaces." In: *Proceedings of the 33rd Australian Conference on Human-Computer Interaction*. ACM. 2021, pp. 182–189 (8 pages). DOI: 10.1145/3520495.3520525.

P4 **Valentin Zieglmeier**, Maren Gierlich-Joas, and Alexander Pretschner. "Increasing Employees' Willingness to Share: Introducing Appeal Strategies for People Analytics." In: *Proceedings of the 13th International Conference on Software Business*. Lecture Notes in Business Information Processing 463. Springer, 2022, pp. 213–226 (14 pages). DOI: 10.1007/978-3-031-20706-8_15.

1.7. Related Publications

In addition to the core publications, we have published the following papers that are related to the contribution of this thesis, but are not included as part of it:

P5 **Valentin Zieglmeier**. "The Inverse Transparency Toolchain: A Fully Integrated and Quickly Deployable Data Usage Logging Infrastructure." *Software Impacts* 17, Article 100554 (September 2023), 4 pages. DOI: 10.1016/j.simpa.2023.100554.

P6 **Valentin Zieglmeier** and Gabriel Loyola Daiqui. "GDPR-Compliant Use of Blockchain for Secure Usage Logs." In: *Proceedings of the 25th International Conference on Evaluation and Assessment in Software Engineering*. ACM. 2021, pp. 313–320 (8 pages). DOI: 10.1145/3463274.3463349.

P7 **Valentin Zieglmeier**. *Appending Data to Blockchain Is Not Sufficient for Non-repudiation of Receipt.* 2023 (7 pages). arXiv: 2308.04781 [cs.CR].

P8 Maren Gierlich-Joas, **Valentin Zieglmeier**, Rahild Neuburger, and Thomas Hess. "Leading Agents or Stewards? Exploring Design Principles for Empowerment Through Workplace Technologies." In: *Proceedings of the 42nd International Conference on Information Systems*. AIS, 2021, Article 1519 (9 pages). URL: https://aisel.aisnet.org/icis2021/is_future_work/is_future_work/7

P9 Patrik Zander and **Valentin Zieglmeier**. "Data Owner Benefit-Driven Design of People Analytics." *Proceedings of the ACM on Human-Computer Interaction* 7.EICS, Article 173 (2023), 38 pages. DOI: 10.1145/3593225.

Their relationship to this thesis is as follows. *First*, to realize our research efforts for **P1**, **P2**, and **P3**, we implemented the *Inverse Transparency Toolchain* (P5) and built on its foundations. *Second*, we published our initial idea how to ensure GDPR compliance in a blockchain-based secure usage log in a concept paper (P6). The ideas were incorporated into **P2**, which is an extended version of that paper. *Third*, we show in a position paper why a common approach to achieve non-repudiation with blockchain is insufficient (P7).

The argumentation supports our chosen approach in **P2**. *Fourth*, as a complement to inverse transparency, we conceptualized how to empower individual employees with people analytics (related to **P1** and **P4**). Our ideas were incorporated into our colleagues' design science research effort (P8). *Finally*, we built on our work in **P4** to implement and evaluate our proposed *benefit* appeal strategy for people analytics (P9).

2. Background

This chapter provides a high-level introduction to the fundamentals and theories relevant for our thesis and defines the concepts we work with. It serves to delineate our research and ground it in theory.

2.1. Data Owners and Consumers

In this thesis, we refer to the participants in a data sharing transaction as the *data owner* and the *data consumer* as defined by Pretschner et al. [185]. The *data owner* "possesses the rights to the data" [185, p. 40]. When relating this concept to the GDPR, the role corresponds to the "data subject." The *data consumer*, meanwhile, is the person or program that processes, and thereby "consumes," insights that identify one or more *data owners* (see Figure 2.1) [185, p. 40]. When someone is accessing their own data, they inhabit both roles. [1]

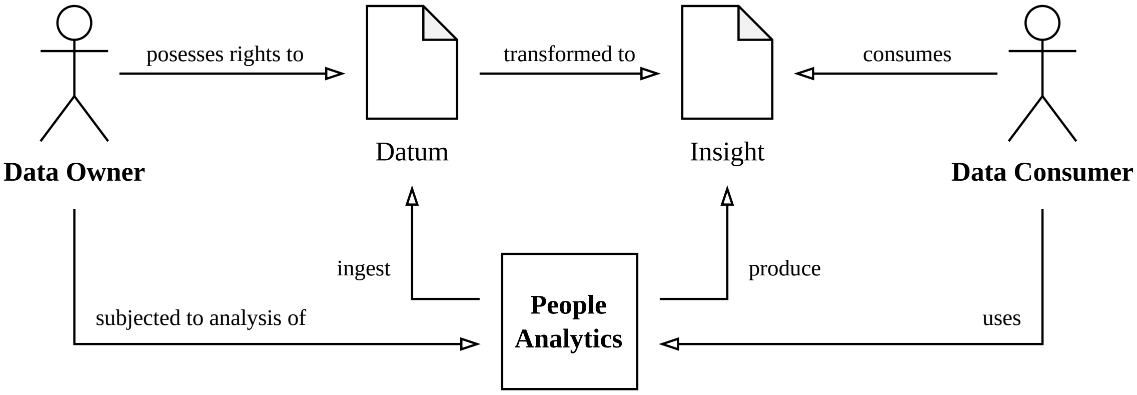


Figure 2.1.: The actors, namely *data owner* and *data consumer*, and their relationship. Arrows and their direction denote a subject → object relationship.

In addition to these actors, we additionally talk about a specific *datum*, which is transformed by PA into an *insight*. This, in turn, is consumed by the data consumer. Importantly, the data owner does not automatically possess rights to the generated insight as well, for example if their included data are anonymized [see 103, p. 153]. Yet, without their contributed data, the insight could not have been produced. [see also 9, Fig. 1]

2.2. People Analytics

With the term *people analytics* (PA), we refer to tools for data-driven management of the workforce. Various other terms are sometimes used interchangeably, among them *HR* *analytics*, *workforce analytics*, and *talent analytics* [223, Sec. 2.1]. They mainly consider behavioral data, which are traces of the individual's behavior in digital tools; for example their emails. Depending on the use case, various other data, including from employee surveys or staff master data, may be used for analysis [71, Tabs. 3–4].

> **People analytics.** Tools for data-driven management of the workforce that mainly consider employees' behavioral data.

In the following, we elaborate on the maturity levels used to classify PA, their different archetypes, and their legal context in Germany.

2.2.1. Maturity Levels

PA are typically classified based on their maturity level, which correlates with their impact but also their risks. Following Giermindl et al. [90, Tab. 4], four maturity levels can be identified. The most basic form are *descriptive* PA, which consider past data to explain how they impact the current status of the company. This level can further be differentiated by how the insights from these PA are used, differentiating between *reactive*, *proactive*, and *strategic* use [106, Fig. 2]. Yet, importantly, they do not make predictions, limiting their scope to the past. More mature PA are *predictive*, meaning they aim to forecast developments in the future. This introduces a level of uncertainty, as the models have to work based on probabilities. Their analyses allow for forward-facing scenario planning and risk mitigation [106, p. 678]. Most literature identifies the third level, *prescriptive* PA, as the most advanced form [see, e.g., 60, 106]. Beyond just forecasting, they also introduce semi-automated decision-making by applying more complex analyses and creating recommendations. They still require human operators to make the final decision, though. To define systems that go beyond this limitation, Giermindl et al. [90] introduce an additional maturity level, namely *autonomous* PA. They define them as autonomously acting systems that make decisions and only communicate their reasoning retroactively [90, p. 425]. Such capability promises to accelerate decision-making and significantly increase efficiency but can result in increased risks of erroneous or harmful decisions.

Summary. The advancement of PA applications characterizes their maturity. Four levels can be identified, namely *descriptive*, *predictive*, *prescriptive*, and *autonomous*. They describe an evolution from retroactive explanations towards autonomous decisions.

2.2.2. Archetypes

Following Hüllmann et al. [108, Sec. 5], PA can also be differentiated according to their archetype, which can be understood as their application area or core use case. The most broadly usable tools identified are *technical platforms*, which do not define own analyses, instead providing the operators instruments that can be used to design custom PA. More directed are *employee surveillance* PA, representing tools that perform invasive tracking of employees. For example, Teramind promises to find insider threats based on their analytics [218]. The archetype of *social network analysis*, then, classifies tools that consider the interaction of employees, deriving their social networks. Such analyses are also suggested by Leonardi and Contractor [141]. One example is Polinode, which claims to help, e.g., identify emerging talents [177]. Finally, the most advanced category is referred to by Hüllmann et al. as *human resources analytics*. As this is, in our view, merely a synonym for the term people analytics, we instead refer to this category as *comprehensive* people analytics. Most identified tools fall under this category, which covers diverse use cases spanning individual wellbeing [154], talent retention [225], but also productivity improvements [199].

Hüllmann et al. identify an additional archetype, namely *technical monitoring* tools [108, Sec. 5.5]. These track technical components or services, allowing, e.g., the identification of technical performance issues. In our definition, tools that do not consider the individual in their analyses are not considered PA, even if their insights theoretically allow conclusions about an individual. Therefore, we forego this class of tools.

Summary. Hüllmann et al. find five archetypes, of which four fit our definition of PA: *technical platforms*, *employee surveillance*, *social network analysis*, and *comprehensive* PA. In practice, most tools fall under the last category, allowing multi-faceted analyses.

2.2.3. Legal Context in Germany

The use of PA in Germany is regulated by multiple laws, most importantly the European GDPR and the German industrial constitution law ("Betriebsverfassungsrecht") [33, Chap. D]. In the following, we give a broad overview of how these laws could impact the use of PA in Germany. We do not consider more specific laws that are only applicable for individual use cases, but are not relevant for PA in general. Also, as this thesis is not a legal one, we refer to the comprehensive work of Blum [33] for a full analysis.

The GDPR applies only to personally identifiable information. As we define PA by their focus on analyzing employees' data, though, the GDPR is generally applicable. One exception are analyses that make use of anonymized data, for example to provide a comparison with the overall market. Note that pseudonymization, in contrast to true anonymization, is thereby not sufficient and means that the GDPR still applies [103]. Given its applicability, the first implication is that the processing entity can only use an individual's data if the processing is *necessary* or the data subject *consented* to the use [87, Art. 6]. Importantly, in the context of the workplace, if the data usage is required for core business processes, it does not require consent. For most cases, though, we can assume consent to be required [92, p. 169]; [33, p. 164]. There has been discussion if true consent is possible in the workplace, given the power asymmetry between employer and employee [see, e.g., 195, 238]. In Germany, this situation is legally clarified: Consent at the workplace is legally possible if, for example, the employee gains legal or economic benefits or their interests are aligned with the employer [119]; [33, p. 112]. We can summarize that, based just on the GDPR and its German instantiation, most applications of PA are legal if their use is *consented to* by the affected employees. Note that this does not apply for sensitive data (e.g., ethnicity or religion), though, whose processing is in many cases forbidden by the GDPR. [33, Sec. D.1]

Beyond the GDPR, the German industrial constitution law enables further restrictions of the use of PA [33, pp. 169 ff.]. It gives works councils the right to decide if PA can be introduced in the first place, as they track the behavior and performance of employees. The works council, therefore, must be informed proactively if such a tool is to be introduced and must agree to its planned use. And, importantly, the works council can even request the abolition of PA that are already in use [33, p. 182]. This gives works councils significant power over the use of PA in the workplace and makes employees' acceptance a vital aspect of their success. [33, Sec. D.2]

Summary. The use of PA in Germany can be legal but is heavily limited by data protection laws. Most use cases require the affected individual's consent. Processing of sensitive data is additionally restricted. Furthermore, through the works council, employees can decide if PA may be used in their company at all, and how.

2.3. Data Sovereigns

We have established that the successful use of PA depends on employees' acceptance. Our vision is that their role is transformed into central stakeholders in the design process of PA. We aim to enable individuals to gain *data sovereignty* [109], meaning "self-determination [...] with regard to the use of their data" [115, p. 550]. This has the potential to improve their trust to allow data usage that might benefit them. Our vision is inspired by Dabrock [56], who notes that advances in data processing can unquestionably bring benefits. Therefore, the focus of data protection should not be on preventing any data usage, but instead on providing transparency and a clear value proposition [56].

> **Data sovereigns.** Individuals that are self-determined regarding the use of their data and must be continuously convinced of giving access to it.

This vision is in contrast to the traditional input-oriented idea of privacy, which only considers the data that are fed into the processing systems and use cases that are defined in advance. Instead, the focus is shifted to be output-oriented and dynamic, allowing individuals to retain oversight while enabling them to give and revoke access to their data at any time. [56, 110]

2.4. Privacy Theories

Given that we see individuals as data sovereigns, we want to empower them to make their own decisions in which cases to allow the usage of their data. We refer to this as their *disclosure decision*, which we try to understand and potentially influence.

> **Disclosure decision.** An individual's decision whether to allow usage of their data, by whom, and for which purpose.

The individual decision factors are a black box for us, which is why we ground our concepts with established and empirically tested privacy theories [following 61, 208].

2.4.1. Privacy Calculus

One of the core privacy theories is that of the *privacy calculus*, which originates in the social sciences [see 127, 136], was transferred to the workplace context [see, e.g., 29, 213], and empirically tested in various scenarios [see, e.g., 43, 44, 54, 62]. At its core is the idea that an individual's disclosure decision represents a trade-off—a calculus—between the costs (risks) and benefits of the disclosure. [208, pp. 1001–1002]

The most comprehensive privacy calculus model for our use case is presented by Bhave et al. [29]. They focus on the workplace and incorporate antecedents and context factors as well as the organizational calculus (see Figure 2.2). Note that the calculus factors differ from the consumer context in which the privacy calculus was originally introduced. Furthermore, independently of the concrete data usage, this model assumes the various context factors to influence the calculus decision directly.

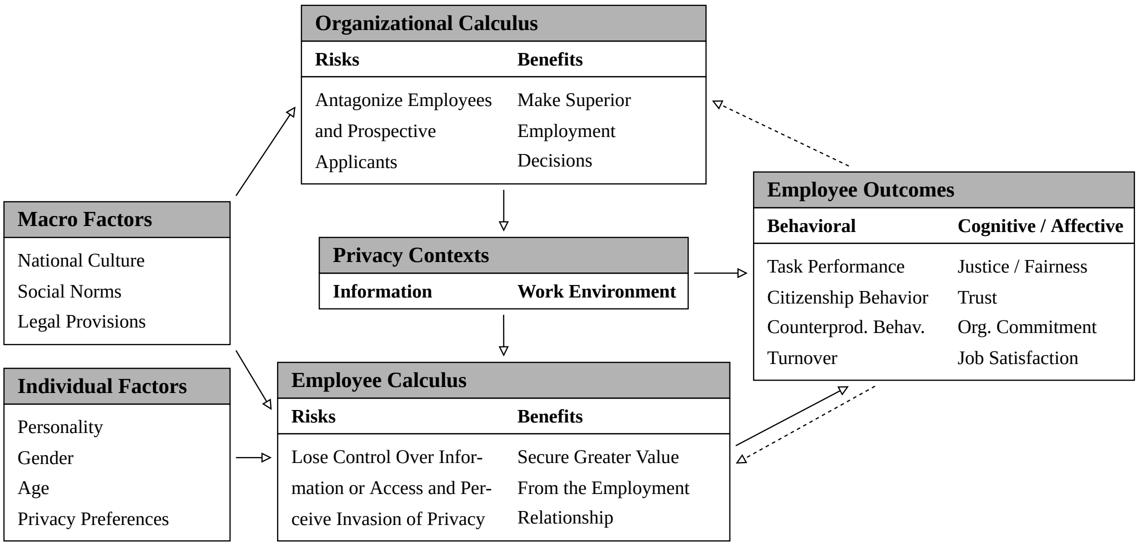


Figure 2.2.: The stakeholders' privacy calculus model [adapted from 29, Fig. 1]. Dashed arrows denote an aggregated, bottom-up effect occurring over time [29, Fig. 1].

Summary. The privacy calculus theory states that an individual's behavioral and affective outcomes are dependent on a weighing of the perceived *risks* and *benefits* of the disclosure.

2.4.2. Privacy Concerns

The second important class of privacy theories is founded in the idea that, fundamentally, the individual's *privacy concerns* are the focal point to understanding their beliefs and behaviors. The seminal model is referred to as *concern for information privacy* (CFIP) [209]. Smith et al. present and empirically validate the model, which consists of the four concerns *collection*, *errors*, *secondary use*, and *unauthorized access* [209, Tab. 5]. Building on this work, Malhotra et al. [148] introduce the influential *Internet users' information privacy* *concerns* (IUIPC) [148], which was identified as the most influential privacy theory recently [see 61, Tab. 2]. The IUIPC only identifies three concerns, namely *collection*—as in the CFIP—, *control*, and *awareness*.

The focus of the IUIPC on the consumer context and internet users means that it is not directly applicable for our scenario. Instead, we consider the more recent *workplace privacy concerns* (WPC) model [217]. It combines and adapts the CFIP and IUIPC for the workplace while adding two additional dimensions to it (see Figure 2.3).

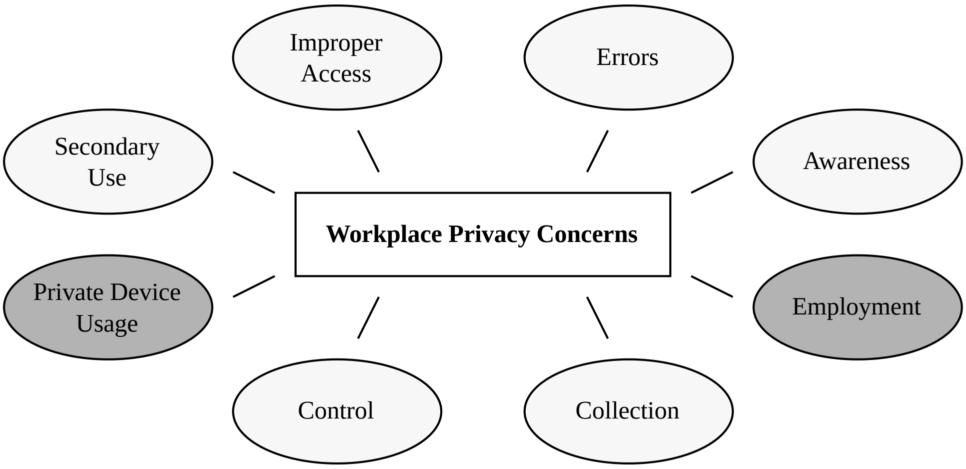


Figure 2.3.: The dimensions of workplace privacy concerns [adapted from 217, Fig. 1]. The ovals with lighter gray background are the original CFIP and IUIPC dimensions, ovals with darker gray background are the additional WPC dimensions.

For our scenario, we find that multiple WPC dimensions are relevant. The *unauthorized secondary use* can be directly linked to the *risk of data misusage* we identify. It defines an employee's concern that their data are used for purposes that they did not anticipate and agree to [209, p. 171]. Related, the concern of *improper access* relates to those usages that are technically authorized, such as access by a system administrator, but are considered inappropriate by the data owner [209, p. 173]. The final aspect relating to the *risk of data misusage* is the dimension of potential *errors*, which describes issues in the data quality or interpretation [209, p. 173]. Then, directly concerning the identified *lack of transparency*, employees are concerned that they do not have sufficient *awareness of privacy practices*, meaning how their data will be used [148, Sec. 2.2.3].

The concern regarding the *employment* is more broadly related to our scenario, describing employees' concern that their productivity is analyzed by the employer [217, p. 6666]. This can induce a general mistrust towards PA, which we may affect indirectly with our contributions. Yet, we do not consider the concrete use cases in PA and therefore do not directly address this concern. The remaining concerns are not relevant to our specific context: The *collection* of data is in most cases not directly informed by PA. The lack of *control* is addressed by the data protection laws in our context (see Section 2.2.3). Finally, the *private device usage* is not applicable, as we do not consider employees' usage of private devices, such as smartphones, for their work. These concerns are still given, of course, but we do not address them further.

Summary. The theories of privacy concerns explain individuals' beliefs and behaviors through factors that cause them concern. In our context, the WPCs of *unauthorized secondary use*, *improper access*, potential *errors*, and lacking *awareness* are relevant.

2.4.3. APCO Macro Model

Both the *privacy calculus* and *privacy concerns* are relevant for the individual's disclosure decision. Integrating them in one holistic model, Smith et al. [208] present the APCO (Antecedents → Privacy Concerns → Outcomes) macro model. It provides an integrative and abstract view that connects the aforementioned concepts (see Figure 2.4).

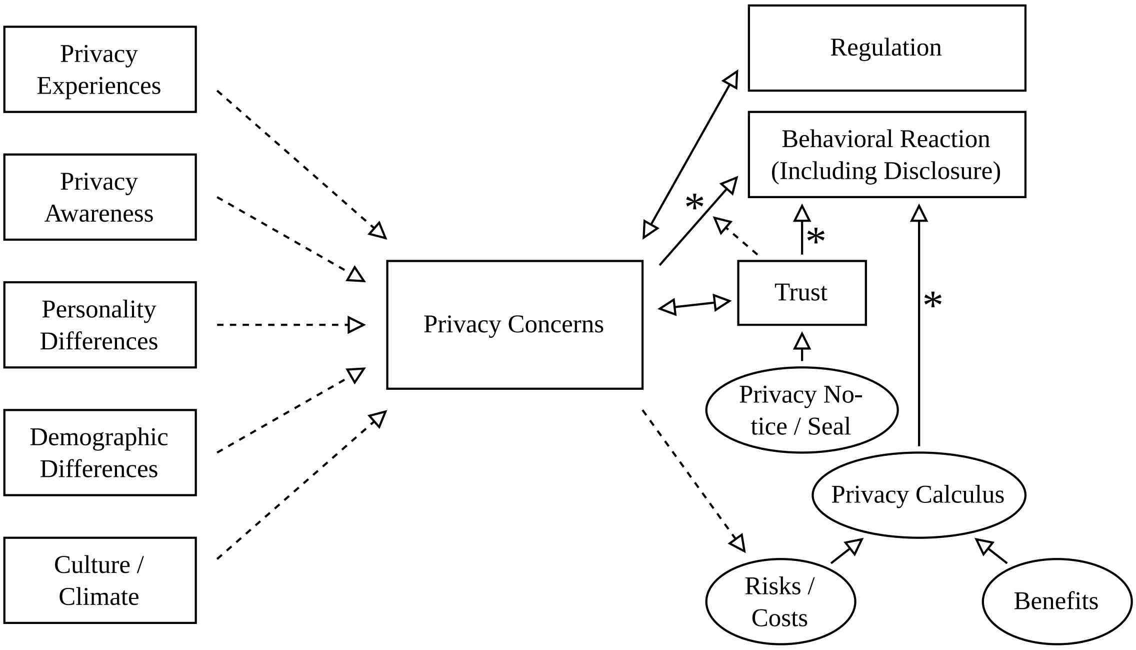


Figure 2.4.: The APCO macro model [adapted from 208, Fig. 3]. Dashed lines denote assumed relationships. Following Smith et al., we omit the potential relationship of actions on the right with the antecedents on the left. Relationships marked with a star (*) may be threatened by the privacy paradox (see Section 2.4.4), as studies on these mostly measure intentions, not actual behaviors [208, Fig. 3].

The APCO macro model shows how the privacy calculus and privacy concerns relate to each other and associated concepts, such as trust. Thereby, it can help to explain how they interact and influence each other. Notably, privacy concerns are assumed to affect the perceived risks, which are part of the privacy calculus (see Figure 2.4).

Summary. The APCO macro model integrates the privacy concerns and privacy calculus theories and thereby shows their relationship.

2.4.4. Privacy Paradox

The APCO macro model and many other models of privacy behaviors are challenged by the *privacy paradox*. This is because, to verify that such a model is representative of real-world behaviors, researchers often employ surveys [e.g., 43, Sec. 4] or interviews [e.g., 217, Sec. 3]. Such methods are reasonable, as they allow estimating the correctness of a model with relatively little resource effort. Yet, *asking* individuals instead of *observing* them will necessarily not measure actual *behaviors*, but instead behavior *intentions*.

Norberg et al. uncover that, for privacy, an intention-behavior gap exists, which threatens the validity of studies that measure intentions only [160, p. 108]. They denote it the *privacy paradox*, referring to the seemingly paradoxical relationship that can be observed: On the one hand, consumers express strong interest in data protection rights, yet, on the other hand, willingly provide personal information nonetheless [160, pp. 100 f.]. Therefore, they hypothesize that the behavioral intention of an individual may be separate from their actual disclosure behavior [160, Fig. 2]. In a two-part study, they empirically test their hypothesis and find that it holds, with participants' stated sharing intentions significantly differing from their actual sharing behaviors [160, Tab. 3].

Summary. The *privacy paradox* denotes the observable difference between individuals' stated *intentions* and their actual disclosure *behavior*. This weakens the significance of models that build only on studies of intentions.

2.5. Inverse Transparency

To understand our contributions and their underlying goal, we now consider *inverse transparency*, which is inspired by Brin [38]; [following 1]. They describe a dystopia in which citizens are continuously monitored by the police, making their lives transparent. To balance this, they propose to empower individuals by letting them watch over their watchers—providing what they call *inverse transparency* [38]. Boes et al. transferred the idea to the workplace in the research project "Inverse Transparenz," suggesting to make all usages of employees' data visible (transparent) to them [35, 36]. More recently, inverse transparency has been proposed as a new digital leadership concept, aiming to solve tensions between managers and employees [88].

> **Inverse transparency.** Visibility into how one's data are used.

Similar concepts to inverse transparency have been developed in the software development context specifically, also aiming to provide transparency to ensure accountability [e.g., 12, 231]. Note that these works consider the consumer context, though. Concretely, the idea of *hippocratic databases* by Agrawal et al. is based on giving individuals access to audit trails of databases holding their information, allowing them to detect misusage [12]. Along those lines, Weitzner et al. deliberate the potential benefits of making data usages transparent to individuals, achieving what they refer to as *information accountability* [231]. They see two main advantages of the approach: *First*, reducing individuals' mental load as they do not have to judge *ex ante* all potential usages of their data [231]. *Second*, helping uncover misusage retroactively, which enables *accountability* [231]. Beyond this, we hypothesize that this transparency could also increase the *felt accountability* [98] of data consumers, thereby deterring data misusage even before it occurs.

Based on these previous works, we propose to build PA with *inverse transparency by design* (see Chapter 4). We think that having an overview of how their data are used can empower employees to gain true data sovereignty (see Section 2.3). Beyond our primary goals of enabling accountability and reducing data misusage, the provided transparency may even be able to reduce privacy concerns. Gierlich-Joas et al. [89] theorize that this effect could counteract the negative impact of direct transparency (see Figure 2.5).

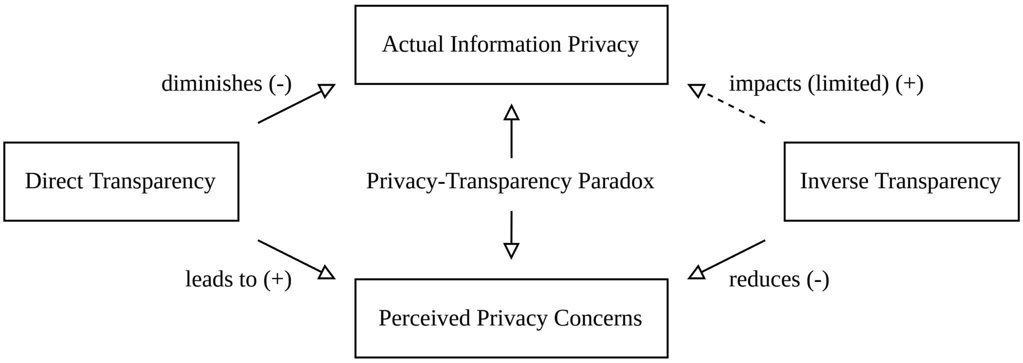


Figure 2.5.: The privacy-transparency paradox [adapted from 89, Fig. 2], showing the hypothesized influence of inverse transparency on (actual) information privacy and perceived privacy concerns.

2.6. Accountability

One of the fundamental concepts to understand the hypothesized effects of inverse transparency is that of *accountability*. A useful starting point is the definition of Weitzner et al., who state that "information accountability means the use of information should be transparent so it is possible to determine whether a particular use is appropriate under a given set of rules and that the system enables individuals and institutions to be held accountable for misuse" [231, p. 84]. To concretize how this concept is used and understood, Kacianka et al. perform a systematic mapping study to identify two important facets: They note that accountability (I) links actions to entities, and (II) that this link is used to hold the entity responsible for their actions [117, p. 210]. This relies on: (a) a complete, tamper-proof log that (b) associates actions and entities and (c) allows reasoning about occurred events [117, pp. 210 f.].[1]

> **Accountability.** Ability to hold entities responsible for their actions.

Furthermore, we can differentiate between *actual* accountability, which is technically or legally ensured, and *felt* accountability, which reflects an individual's perceptions [98, pp. 205 f.]. The latter relies only on the individual's expectation to be held accountable, which means an actual evaluation is not required. Importantly for our context, the "evaluating audience may include the actor themselves" [98, p. 206].

[1] Refer to Kacianka and Pretschner [118] for an interdisciplinary view on accountability as a concept.

2.7. Plausible Deniability

Achieving accountability requires a link between entities and their actions that can be used to hold those entities responsible. This is endangered by *plausible deniability*, which can be understood as the existence of at least one plausible, alternative explanation for an observed phenomenon that does not implicate the responsible entity [see, e.g., 132].

> **Plausible deniability.** Ability of an entity to deny an action by providing a plausible alternative explanation for an observed phenomenon.

2.8. Non-repudiation

When dealing with mistrusting entities, a sought-after property to prevent plausible deniability is *non-repudiation* [see 111, 239]. This refers to the impossibility of an action to be denied by the responsible entity. Relating this to plausible deniability and accountability, we find: Non-repudiation prevents plausible deniability and thereby ensures the accountability of an entity.

> **Non-repudiation.** Impossibility for an entity to deny their action.

In our second contribution, we specifically consider non-repudiation of data exchange, also referred to as data delivery [see 130]. Concretely, the data owner sends their sensitive data directly to the data consumer, with this exchange being provably recorded. Consequently, we want to ensure two properties: *First*, *non-repudiation of origin* ensures that the data owner cannot deny that they sent the data to the data consumer. *Second*, *non-repudiation of receipt* ensures that the data consumer cannot deny having received the data. [150, p. 2]; [239, p. 135]

3. Related Work

This chapter discusses existing research and identifies the research gaps that we address in our contributions. Thereby, it embeds our work in the state of the art. It is structured based on the four research directions that we derive in the Introduction (see Section 1.4).

3.1. Reducing Risks by Increasing Transparency

As we introduce in Section 1.2, employees face an elevated *risk of data misusage* due to advanced PA, exacerbated by a *lack of transparency* (problems **I** and **II**). Various related works aim to address these risks. They can be structured based on their approach. To begin with, *preventive* solutions only permit data usage for purposes specified in advance. They strive to prevent misusage *ex ante* [see, e.g., 25, 186]. *Detective* solutions, meanwhile, rely on self-regulation and voluntary compliance. They assume many data usages to be benign and enable them by default. Contractual agreements enable *ex post* reaction to misusage, and thereby accountability [see, e.g., 181, 183]. Finally, *transparency-enhancing* approaches are founded on the assumption that end users benefit from direct transparency over the usage of their data. Thereby, they can inform consent decisions and enable oversight [see, e.g., 101, 158].

3.1.1. Preventive Approaches

Foundational to our thinking is the assumption that some data usage is acceptable or even beneficial. This has also been noted by Cate, who observed that "many uses of personal information pose no risk of harm to individuals, while creating significant benefits for data subjects and society more broadly" [40, p. 37]. The question remains: How to enable sensible usage of data? *Preventive* approaches are a pessimistic solution for this [e.g., 23, 24, 31, 120, 129, 171, 219].

Cate proposed rethinking data protection with *use-based privacy*, which entails defining collective norms that specify acceptable uses [40]. These could explicitly exclude, e.g., discriminatory uses of data where they are obvious. These norms are meant to be enforced automatically by systems that process data. For example, Bagdasaryan et al. [23] have recently presented *Ancile*, an online privacy platform that implements use-based privacy. Python analysis scripts can be sent to the *Ancile* server for execution in a sandbox. The server expects predefined policies that it automatically enforces [23, Sec. 4].

A similar strategy to use-based privacy underlies *usage control*, which extends access control to the usage and distribution of data [168, 186]. When data are given away, *distributed usage control* aims to enforce usage policies on client systems [184, 185]. This depends on the enforceability of the policies, which encompasses both controllability and observability. As Pretschner et al. note, not all events can be fully controlled, which requires the weaker notion of observability [185, p. 41]. In this case, accountability can still be achieved by logging violating events and holding actors responsible retroactively [183, p. 3]. This approach could be considered *hybrid*, in that it does not *prevent* all policy violations but allows appropriate reactions when they are *detected* [see also 30, 169]. Fully preventive systems, meanwhile, can technically be realized [see, e.g., 124, 226] depending on the usage context and performance requirements.

Preventive approaches are promising and can serve as a minimum safeguard. We identify two issues for our context, though: *First*, we consider it impossible to know all acceptable usage patterns for data in advance, which limits the applicability of this strategy. *Second*, it can be difficult in practice to fully prevent all policy violations [see 41, 185] and, even if it is possible, can lead to challenging technical requirements [see, e.g., 226, Sec. 4.3] and performance issues [see, e.g., 124, Sec. 5.2.1].[1]

3.1.2. Detective Approaches

To solve issues with preventive policy enforcement, *detective* approaches have been proposed [e.g., 12, 32, 41, 68, 91, 135, 214, 230, 232, 233]. They also expect usage policies that are created in advance, as is done for preventive systems. Contrary to those, they still allow most usages of data, though, and are instead based on logging and retroactive enforcement [see also 185, p. 41].[2]

For example, Cederquist et al. propose *audit-based compliance control*, which could be translated as *retroactive usage control enforcement*. When data consumers access data, their actions are logged. After the fact, an auditing authority checks the logs and holds them accountable in case of misusage [41, Sec. 2]. The *APPLE* system described by Etalle and Winsborough [68] is comparable, albeit focused on the use case of sending documents. These need to be annotated with usage policies that cannot be removed. Logs of user actions are created locally on their system. Auditing authorities regularly request those logs to verify that the usage was allowed [68, Sec. 2]. More recently, solutions for cloud environments [e.g., 214] or big data analysis pipelines [e.g., 233] have been proposed. They differ in their technical implementation, adapting to the challenges of their context, but are conceptually comparable.

Independently of the concrete usage context, detective approaches can complement purely preventive enforcement in some aspects. Yet, issues remain: *First*, allowed usage policies still have to be fully specified before the analysis, which is, in our view, not feasible in all cases (see above). *Second*, they cannot address the *lack of transparency* in PA. Even authors of detective systems acknowledge that it can be useful to provide end-users with transparency over the usage of their data [see, e.g., 12, Sec. 5.8, 135, p. 239]. Therefore, we consider it preferable to provide individuals with *inverse transparency*, allowing them direct oversight over how their data are used.

[1] Refer to Feigenbaum et al. [75] for further arguments against purely preventive approaches.

[2] Note that, technically, preventive systems are capable of operating detectively as well. In some cases, this is even a feature [e.g., 124]. The approaches differ conceptually, though.

3.1.3. Transparency-Enhancing Approaches

The idea of *inverse transparency* is to enable individuals' direct oversight of how their data are used to deter data misusage and enable accountability (see Section 2.5). Many relevant approaches are referred to as transparency-enhancing technologies (TETs) [see also 101, 114, 158], which is why we summarize approaches towards this goal as *transparency-enhancing* [e.g., 18, 27, 81, 94, 95, 102, 129, 142, 143, 174, 205, 215, 236]. At their core is the goal to provide end-users with direct transparency over how their data are used.

To provide a minimum level of transparency, some researchers suggest standardizing the information provided to users about the *intended* processing of their data [e.g., 18, 53, 94, 95], for example by annotating REST APIs [95]. This can be useful but is, in our view, not sufficient, as no additional transparency beyond intentions is provided. Data misusage is, on principle, not part of intended data processing, so it cannot be captured with these solutions.

Therefore, many authors instead present contributions towards providing transparency over *actual* data usages. The most basic solution in this direction is *information flow tracking* [e.g., 81, 143, 214, 236]. Here, the goal is to follow the path that data take in a system, including transfer, modification, and access. For example, Zavou et al. [236] present *Cloudopsy*, which aims to detect data transfers across a defined boundary. This allows them to provide users with a graph view of which third parties their data were shared with [236, Fig. 2]. Similarly, Fromm and Stepa [81] develop *HDFT++*, a combination of static information flow analysis and dynamic tracking mechanisms for cloud services [81]. Contrary to *Cloudopsy*, their approach is specifically aimed at system-internal data transfers. Yet, the collected information is more technical, meaning it is meant for auditors to verify and does not address users directly [81, Fig. 1]. Solutions such as these can be useful, but are limited. *First*, by design, they are focused on simpler use cases such as ensuring that data are not transferred outside a geographical boundary [see 81, p. 333]. *Second*, the focus on the technical layer means that the *purpose* of a data access is not captured, and the systems cannot easily determine its *cause*, i.e. if it originated in a data consumer's action or is just the effect of a technical process.

Accordingly, more advanced transparency-enhancing approaches have been developed, aiming toward addressing those issues in various ways [e.g., 27, 30, 102, 142, 174, 205, 215, 240]. We discuss the most relevant examples for our context in the following [namely 27, 142, 215]. To start with, the *CIA framework* [215] works by packaging personal data with a usage logging module in Java JAR files before they are shared. Every access to the data triggers the usage logger, whose logs are accessible by the data owner at any time [215]. This implementation is relevant to learn from and can be seen as a potential instantiation of inverse transparency in software. Yet, we identify two issues: *First*, the approach is limited in scope to systems that do not need to know available data before their use. Therefore, it is not easily transferable to advanced PA using, e.g., big data analysis. *Second*, the work considers neither the developer nor the user perspective, which are important to judge the feasibility of the solution. *PrivacyStreams* [142], meanwhile, also aims at Java development, but specifically for Android. It is a library for transparency-enabled stream-based data access in mobile apps. Its goal is to facilitate developers' efforts toward providing their users with transparency over data usages in their apps. The authors recognize the importance of developers' judgment and, therefore, present two developer studies. Their lab study with ten Android developers [142, Sec. 7.1] focuses on programming efficiency and subjective feedback, showing acceptance of the approach. In a field study with five developers [142, Sec. 7.2], they assess the retrofitting of existing apps with their library, showing some required effort and issues but overall positive feedback. Yet, issues remain. *First*, the scope of this solution is also limited, with Android apps not being a relevant scenario for PA. *Second*, the work does not consider the user perspective, either. Finally, then, Bemmann et al. [27] also focus on mobile apps, but they present a study with users to determine the influence of transparency and control measures on users' willingness to share their data. Thereby, they recognize, as we do, the importance of assessing the user perspective. Their study is specifically focused on the use case of a mobile sensing application. That means they track data *collection*, not *usage*. As their main contribution, they present a large study with 227 participants [27, Sec. 5]. Their findings are relevant for us, as they suggest that transparency measures can be important but should be accompanied by control tools [27, Sec. 7]. Yet, we identify: *First*, they do not present a transparency solution sufficient for PA in the workplace, as they focus on mobile sensing. *Second*, they do not consider the developer perspective in their study.

In summary, we find that transparency-enhancing approaches can be useful, but existing research does not comprehensively address the issues we identify. To start with, most research covers the consumer context [e.g., 102], which means that results cannot be easily transferred to the workplace context and PA. Furthermore, previous works are either only conceptual [e.g., 12], only empirical [e.g., 27], limited in scope [e.g., 142], or lack comprehensive evaluation of their proposed solution [e.g., 215].

3.1.4. Conclusion

The *risk of data misusage* in PA can be addressed with *preventive*, *detective*, and *transparency-enhancing* approaches, which are complementary. Both preventive and detective systems can offer important baseline protection but are insufficient on their own. They require all allowed usages to be defined in advance, which is not always possible. Furthermore, they do not improve the *lack of transparency* in PA. Therefore, transparency-enhancing approaches are necessary. Existing research on these mainly considers the consumer context, which differs from the workplace. We lack a comprehensive design approach for inverse transparent PA in the workplace context (⇒ **G1: Inverse Transparency in People Analytics**). And the developed approach needs to be empirically studied both from the developer and user perspective to support its suitability (⇒ **G2: Empirical Studies of the Developer and User Perspective**).

3.2. Transparency for Sensitive Data

For most data, the retroactive accountability afforded by inverse transparency is, in our view, sufficient to deter misusage. This calculus changes for sensitive data such as health data. Analysis of these entails increased risks for the affected individual [see, e.g., 14], and consequently employees' willingness to share them is significantly reduced [220, Fig. 4].

As specific useful applications for sensitive data exist, though, we want to enable sharing of sensitive data without requiring employees to give their employer control over them. Instead, we propose that they should keep the generated data under their own governance, for example on their personal device. In the case of specific, beneficial applications of their data, data owners can then share them directly with the interested data consumer, which we refer to as a *data exchange*. Technically speaking, that means we do not want to depend on a trusted third party (TTP) to enable inverse transparency in this scenario. As we introduce in Section 1.4.2, this introduces the challenge of plausible deniability, allowing the data consumer to repudiate having received the data. Thereby, accountability is endangered. To prevent this, we aim for the decentralized data exchange between the data owner and data consumer to be provably and securely recorded, ensuring that it cannot be repudiated. This requires decentralized solutions for (1) *non-repudiable data exchange* (see Section 2.8) and, given our legal context, (2) *GDPR-compliant transparency logs*.

3.2.1. Weaker Notions of Trusted Third Party

Being able to utilize some form of TTP can significantly reduce the technical complexity. Therefore, a few previous works propose utilizing weaker notions of TTP, such as using technical components to fill the role of the TTP, as they could be considered neutral.

One common approach is to employ trusted hardware as a substitute TTP. For example, smart cards [e.g., 128, 145], trusted platform modules such as ARM TrustZone [e.g., 10, 139], or Intel Software Guard Extensions (SGX) [e.g., 123, 164, 187, 237] can be used. These approaches still require comparatively high levels of trust, though. In fact, the behavior of the trusted hardware cannot be easily verified by participating parties. The utilized hardware is only considered secure because its manufacturer is assumed to be trusted [see 52, Fig. 2]. Furthermore, the requirement for specialized hardware increases technical complexity and costs.

Alternatively, a smart contract can be employed as a TTP. For example, some works implement arbitrated exchange [e.g., 63] or optimistic fair exchange [e.g., 65] with smart contracts. The technology has also been employed to protect the integrity of the created logs [e.g., 21, 192]. Compared to trusted hardware, the actual code that is executed could—at least in theory—be verified by all parties. Yet, this means that users would be expected to perform a security audit or have the technical knowledge to be able to judge the trustworthiness of a smart contract. Various vulnerabilities and security issues with existing smart contracts [see, e.g., 46] show that this is a difficult problem.

Independently of the concrete technical component utilized, some level of trust is still required. Therefore, such approaches do not truly remove the TTP.

3.2.2. Fully Decentralized Solutions

There are solutions that remove dependence on any TTP, making them fully decentralized. The approaches differ technically for the two challenges of *non-repudiable data exchange* and *GDPR-compliant transparency logs*. Therefore, we discuss them separately below.

Non-repudiable Data Exchange

Many researchers consider non-repudiable fair exchange [see 130, 239] to be impossible without a TTP and therefore do not attempt it [e.g., 63, 134, 227]; [see also 84, 166]. Recently, though, a promising area of research towards decentralized non-repudiable data exchange has emerged: delivering data via blockchain.

In its most basic form, the approach is based on simply appending the data to a blockchain [see, e.g., 66, 237]. This approach can ensure non-repudiation of origin, as the sending of the data is tracked in the blockchain. Regarding non-repudiation of receipt, though, issues arise. The core assumption in this regard is often that, as the blockchain is publicly accessible, the receipt of data is simply "undeniable" [237, p. 61]. Yet, as we have shown previously, this assumption does not hold. Confidential data cannot be shared this way, as they would become publicly accessible [7, Sec. 2]. And even if confidentiality is not required, plausible deniability remains [7, Sec. 3]. Promising to solve the issues, staged protocols have been proposed [e.g., 45, 228]. Fundamentally, they work by appending only an unreadable part of the data to the blockchain. The other part can, for example, be sent directly. Therefore, confidentiality is preserved. Yet, these protocols still depend on the non-repudiation of data delivery via blockchain append, which allows for the same plausible deniability regarding data receipt [7, Sec. 4.1]. Meaning, non-repudiation cannot be guaranteed in either approach, rendering them insufficient. Accordingly, data delivery via blockchain does not represent a sufficient solution.

Beyond these, to our knowledge, only two non-repudiation protocols exist that do not depend on a TTP: the Markowitch and Roggeman (M&R) protocol [150] and the Mitsianis protocol [155]; [see 130, Sec. 3]. As we do not have access to the manuscript describing the Mitsianis protocol, we follow the description in Kremer et al. [130]. Both protocols function similarly, with the main difference being that the number of rounds in M&R is chosen dynamically, while it is fixed for Mitsianis [130, Sec. 3.3]. As this property of the M&R protocol is an important foundation of its security properties [see 16, p. 31], we consider the Mitsianis protocol insufficient. The M&R protocol, meanwhile, has been shown to be probabilistically secure [16]. Yet, due to its conceptual nature, it does not present concrete algorithms for the required identity verification and time-asymmetric encryption [see 150]. These are fundamental for the practicability and necessitate a solution that can be shown to fulfill the required technical properties.

GDPR-compliant Transparency Logs

After the data have been exchanged, meaning sent from the data owner to the data consumer, and evidence of this interaction was created, this evidence has to be stored securely. Thereby, we prevent retroactive deletion or manipulation of the logs. To that end, techniques from *secure logging* aim to guarantee the *authenticity* (correctness) and *completeness* of logged events [78, Sec. III]. As noted above, we aim for a decentralized approach, meaning one requiring no TTP. Considering secure logging generally, such decentralized solutions exist [e.g., 105, 146]; [see 78]. For our context, though, two additional constraints arise. *First*, we consider the sub-problem of *transparency logging*[3] [190], which entails that

[3] Also referred to as *secure usage logging* [6] or *data usage auditing* [116].

the data owner needs to be able to view the generated logs affecting their data at any time, even though they may not generate them. *Second*, we additionally require compliance with the GDPR, as personal data are stored. In the following, we analyze if relevant previous works abide by these constraints.

Many decentralized logging schemes are not applicable to the scenario of transparency logging, as they assume the more common auditing use case [e.g., 96, 105, 116, 146, 161, 234]. In this scenario, a limited number of *verifiers* [26] can access *all* logs for auditing purposes [see, e.g., 146, Sec. 4.1]. Other parties, specifically the data owner in our case, cannot access the logs by default. The auditing is also often assumed to be done only periodically [see, e.g., 96, Sec. 3.1], which is why continuous access to the logs is not considered. Therefore, two issues arise: *First*, the confidentiality of the logs is not given, as auditors get full access. This, in turn, means that these schemes are often not compliant with the GDPR's confidentiality requirements by default. *Second*, data owners do not receive direct transparency over usages of their data. Therefore, the schemes are by design only applicable for *detective* systems, but not for our *transparency-enhancing* system (cf. Section 3.1).

When considering the schemes that are applicable, we find that some are based on custom data structures that provide the required security properties [e.g., 174, 221]. For example, *Insynd* [174] is very closely related to our solution. This logging system is motivated by the transparency logging use case as well. It is based on the secure data structure *Balloon* [189] of which it inherits the security properties. Depending on the application scenario, *Insynd* can be a sensible solution and offers both high performance [174, Fig. 4] and theoretical unlinkability [174, Thm. 3]. Yet, two issues remain. *First*, Peeters and Pulls assume the existence of an identity verification algorithm but do not present it. *Second*, to ensure the consistency of events, they require a set of TTPs they refer to as *monitors*, who continuously monitor the log data stored on the server [189, p. 633]. While the main algorithm does not depend on a TTP, the forward security of stored log entries—which is fundamental for log integrity—is only guaranteed if one is present. Therefore, it is not applicable in the decentralized scenario. An alternative solution is presented by Tomescu et al., who developed an *append-only authenticated dictionary* [221]. Similarly to *Balloon*, it aims to allow an individual to verify that an element they know is in the set and that the data structure was not retroactively modified. It is designed to support a malicious server [221, Fig. 1]. This means that the data structure protects against an adversary that has taken control of the central logging server. Yet, three issues remain. *First*, to function, the scheme requires a trusted setup phase [221, p. 1313], which we consider a (weaker) notion of a TTP. *Second*, the data structure leaks the existence of users' public identifiers [221, p. 1309] and, by design, cannot support deletion requests. Therefore, it is incompatible with the GDPR. *Third*, in its current form, the data structure is computationally impractical, as it can grow to sizes of hundreds of GiBs [221, p. 1300].

Custom data structures require complex security proofs and can have practical limitations. As an alternative, blockchain has been shown to be a feasible technology for the context of secure logging [see 86, 203]. Accordingly, various decentralized logging schemes based on blockchain have been proposed [e.g., 107, 191, 202, 207, 222]. As the underlying blockchain technology is unchanged for these approaches, they inherit useful security properties, including immutability and availability. To simplify, we forego blockchain forks as an issue and assume (eventual) consistency in the network. Therefore, the only open question relevant for our context is if their data storage strategy is compliant with the GDPR, specifically its requirements towards confidentiality and erasability [see 2, Sec. 3.2]. In the following, we show the limitations of notable related works exemplarily. Thereby, we derive the gap in literature. A comprehensive discussion of approaches to GDPR-compliant use of blockchain can be found in our paper [see 2, Sec. 6.2].

One common approach to enable confidentiality and erasability is applied by *EngraveChain* [207]: encrypting all data stored in the blockchain. Their encryption layer is based on asymmetric encryption, using a list of known public keys of participants to encrypt data for them [207, Sec. 3.1]. This approach can improve confidentiality compared to storing the plaintext in a blockchain. Yet, it exhibits two fundamental issues: *First*, if the full content of the block is encrypted, querying history quickly becomes computationally infeasible. This reduces the *de facto* transparency that is attainable, potentially making the approach unusable for large datasets. *Second*, encryption itself only guarantees pseudonymity of data, which means that data protection requirements still apply [144, Sec. 2.2]. Therefore, the approach is not GDPR-compliant. To remedy this issue, Putz et al. [191] employ a relatively common approach: *hashing out*, which means storing only the hash of the logs in the blockchain [see also 107, 202, 222]. The plaintext logs are stored in a regular database [e.g., 191] or protected with other security measures such as a local permissioned blockchain [e.g., 202]. There are two fundamental issues, though: *First*, since the data themselves are not secured with the blockchain, the solution depends on a TTP to manage them. It could delete or modify arbitrary entries, with only their hash remaining. Therefore, the approach can only preserve knowledge about the *existence* of entries, not their *content*. While such tampering can be detected and theoretically prosecuted, the created accountability is limited, as options for *plausible deniability* remain. For example, in case of a corrupted hard disk or malicious encryption of company systems, data loss can be realistic. Ensuring the *contents* are preserved as well would require additional security systems, which then have to solve the same challenges that the blockchain was meant to address. Accordingly, the approach cannot guarantee information security. *Second*, critically, the underlying assumption is that the hash stored in the blockchain is not considered personal data. Then, confidentiality would be guaranteed, and data protection laws would not apply. Yet, even cryptographic hashes are only considered *pseudonymous*, as they are vulnerable to brute force attacks [144, Sec. 3.2]. Therefore, as noted above, the stored data still fall under GDPR provisions, making the approach not GDPR-compliant.

The inherent conflict with the GDPR lies in blockchain's immutability [165]. To overcome it, some authors propose to just modify the underlying blockchain technically to enable mutability. For example, Farshid et al. [74] propose a *forgetting blockchain* that automatically deletes blocks after some time has passed, Ateniese et al. [19] suggest the use of *chameleon hash* functions to effectively delete blocks, and Deuber et al. design an algorithm for *mutability by consensus*. All these approaches have individual limitations, but the most important issue is that they weaken blockchain's immutability. This precludes the inherent forward integrity of the logs, which is our reason to consider blockchain in the first place. Therefore, they are insufficient for our purposes. [following 2, Sec. 6.2]

In summary, existing decentralized logging schemes are not applicable to our scenario. They either assume the auditing use case, which is not transferable [e.g., 96] or use custom data structures with practical limitations [e.g., 221]. Blockchain has been shown to be a feasible alternative, but existing approaches are not GDPR-compliant [e.g., 207] or weaken the security guarantees provided by blockchain to achieve compliance [e.g., 74].

3.2.3. Conclusion

When dealing with sensitive data, we do not want to depend on a TTP to enable inverse transparency, minimizing the required trust. Instead, we propose a decentralized approach: Sensitive data should only be sent directly from the data owner to the data consumer, with their exchange being securely recorded. As elaborated above, we accordingly need to ensure (1) non-repudiable data exchange and (2) GDPR-compliant transparency logs. For (1), the M&R protocol is applicable, yet we lack concrete algorithms for the required identity verification and time-asymmetric encryption (⇒ **G3: Algorithms for Non-repudiable Data Exchange**). For (2), we find that blockchain can provide the desired security guarantees. Yet, existing research does not solve the requirement of GDPR compliance while also maintaining the forward integrity afforded by blockchain (⇒ **G4: GDPR-Compliant Decentralized Transparency Log**).

3.3. Fostering Trust Through User Interface Design

One important aspect of providing trusted transparency (addressing problem **II**) is the underlying technical infrastructure that we focus on above. Yet, that does not suffice. The design of the user interface of the utilized transparency dashboard also directly influences the trustworthiness and by extension the individual's intention to use it [3, p. 182]. User trust is influenced by environmental factors such as the type of system, which is why concrete implementations of trustworthy software should take them into account. Yet, independently of the usage context, specific user interface design facets can influence trust and are therefore important to consider [see, e.g., 48]. Previous works on user trust in software can be distinguished as either focusing on specific use cases or considering the trustworthiness of software generally.

3.3.1. Trustworthiness in Specific Use Cases

Most research on user trust focuses on specific use cases and does not provide general guidance. An often-researched usage context are e-commerce websites [e.g., 51, 69, 70, 100, 201]. For example, Faisal et al. study user preferences for typical web design factors, such as typography and navigation, to determine their influence on user satisfaction and trust [70, Fig. 1]. Meanwhile, Saw and Inthiran consider more e-commerce-specific features, such as if parcel tracking services are offered, and correlate them with typical personality traits [201, Tab. 5]. Thereby, they show which design features can increase trust based on the user's assumed personality [201, p. 375].

Another typical use case for trust research is automated driving [e.g., 22, 64, 80]. As individuals delegate responsibility to the vehicle, trust in the automation is important to shape their attitudes as consumers [22, p. 1911]. Frison et al. [80], for example, consider the influence of the driver-vehicle interface specifically on the user's trust. As an indicator of (mis)trust, they measure, among other factors, the braking behavior of users [80, Fig. 2]. Another example is the research of Azevedo-Sa et al. [22], who design a user trust estimator for an automated driving system. In a study, participants had to rate how detected and missed hazards influenced their trust level in the system [22, Sec. 5.1].

These and similar works are important to understand trustworthiness for the specific use cases they consider. Their focus on a single application is reasonable, as environmental factors are an important aspect of user trust [see, e.g., 104, p. 416]; [3, Sec. 3]. Yet, the identified measures are, by their nature, context-specific and may therefore not be generally applicable.

3.3.2. Trustworthiness of Software Generally

Some reviews and studies consider trustworthiness of software generally [e.g., 48, 77, 79, 104, 138, 167, 210, 216]. A notable example is the work by Sutcliffe, who builds a cognitive model of trust based on a review of psychological literature [216, Fig. 2]. Based on the model, they analyze the use case of e-science collaboration to determine the concrete factors that influence trust in this context [216, Sec. 6]. Their work is relevant as a theoretical foundation, as it offers a robust framework for the process of decision-making based on the trust relationship. Yet, Sutcliffe does not consider concrete trustworthiness factors that can be targeted to *improve* user trust. Instead, they only aim at evaluating trust in an existing system.

In a larger study, Söllner et al. present a model of concrete factors of software that elicit and improve trust. Their model is based on intuition and a selection of relevant literature. To evaluate the influence of individual factors on overall trust, they performed a large-scale study with undergraduate students. Participants used a mobile application and answered a questionnaire [211, p. 8]. The results confirm that the identified factors have an influence on overall trust in the app [211, Fig. 3]. Their work is very comprehensive and an important step forward. Yet, it exhibits two issues. *First*, the trustworthiness factors are not elicited in a systematic approach. The chosen study method, meanwhile, is not suitable to determine if relevant factors were omitted. Therefore, the completeness of the model is not sufficiently ensured. *Second*, they do not focus on the user interface and therefore do not present measures to create a trustworthy user interface. Their study, furthermore, does not manipulate the identified factors in the utilized app to measure the effectiveness of such interventions. Instead, they focus on broad themes that affect technical artifacts in general. Accordingly, they do not provide guidance on how to design a trustworthy software interface.

Finally, French et al. [79] performed a systematic literature review on trust in automation. Thereby, they address the lack of a systematic approach. Their work provides an extensive overview of trust in automation, models to explain it, trustworthiness factors, and measurement instruments. Yet, they do not comprehensively address our research problem. *First*, even though they provide an overview of trustworthiness factors [see 79, Sec. 6.2], their focus on autonomous systems specifically makes their work less applicable to our scenario. For example, they discuss the levels of autonomy as one relevant dimension [79, Sec. 6.2.5]. *Second*, they also do not present concrete measures to improve the trustworthiness. Therefore, their work is also not sufficient to provide guidance for user interface design.

3.3.3. Conclusion

Existing research on user trust is often limited to specific use cases. And most works that do consider trust in software more generally do not follow a systematic approach. Importantly, there is no concrete guidance available on how to design a trustworthy user interface specifically. Accordingly, we identify a research gap (⇒ **G5: Trustworthy User Interface Design**)

3.4. Motivating Employees to Share Their Data

Finally, we take a more strategic view and consider what can motivate employees to share their data. Grounded in the privacy calculus theory (see Section 2.4.1), we determine two potential strategies: increasing *awareness* to counter perceived risks [see also 89], and providing *appeal*, which targets perceived benefits [4]. With the latter, we can address the *missing appeal* (problem **III**) of PA.

3.4.1. Awareness Strategy

Privacy awareness refers to an individual's ability to judge the risks associated with their disclosure [59, p. 278]. Different terms are used interchangeably for this concept, among them the individual's *level of knowledge* or their *data literacy* [50, Sec. 3.1]. Most existing studies on individuals' data disclosure focus on this aspect. They try to increase the user's awareness and data literacy on what happens with their data after the disclosure [e.g., 59, 180, 193, 194].

In order to ensure informed consent, the privacy awareness of individuals is required. Furthermore, it is a prerequisite for an individual's privacy calculus [50, p. 4022]. There are open questions as to how an individual's awareness influences their disclosure decision, though. Intuitively, we may assume that more awareness leads to more concerns, which, in turn, inhibit data disclosure [see, e.g., 59, p. 280]. For example, Risius et al. find that making users aware of privacy threats can induce privacy-protecting behaviors [194, Sec. 3]. However, Gambino et al. note that uncertainty over how data are processed makes individuals feel unsafe [83, p. 2841]. This would mean that some level of transparency over how the data are processed could reduce this feeling and increase voluntary sharing. Along those lines, Deuker hypothesize that awareness could have a positive effect if combined with control measures for the individual. This combination could increase sharing volume and accuracy [59, p. 281]. In a study on workplace data sharing, Patil and Lai find promising results in this direction: Their system's transparency over which information is collected did not lead to users making more privacy-conservative choices and instead reassured them to share more [170, p. 109]. Finally, to improve employees' privacy awareness, inverse transparency can be provided, which means giving them oversight of the usage of their data. Gierlich-Joas et al. hypothesize that this may reduce privacy concerns [89, Fig. 2], which are an important factor in the disclosure decision (see Section 2.4.3). Indirectly, the privacy concerns affect the perceived risks, which in turn influences the privacy calculus.

We present contributions that aim to provide inverse transparency (see Section 1.5). This could indirectly positively influence employees' disclosure decision. Yet, addressing perceived risks is not sufficient, as the perceived benefits are an important factor as well (see Section 2.4.1). Therefore, we also need to consider appeal strategies.

3.4.2. Appeal Strategy

Research on increasing the appeal of software tools is mainly focused on the consumer context [e.g., 20, 47, 99, 159, 200, 229]. In comparison, investigations of the workplace context are rare [4]. Existing studies in the workplace context mostly consider appeal only on the side, with no systematic overview of possible appeal strategies available. For example, Mettler and Winter [152] investigate if the addition of social features increases employees' information sharing, which has been shown for the consumer context. They find that individuals are more calculating in the workplace compared to their private life [152, p. 111]. Complementary, Stock-Homburg and Hannig [212] investigate if employees follow their intentions when disclosing information to service robots. They consider the privacy paradox, which suggests that individuals' expressed *intentions* differ from their enacted *behaviors* in the context of privacy (see Section 2.4.4). Again, though, they do not investigate appeal in more detail and only consider the abstract *benefit* of the investigated robots as a positive motivational factor [212, p. 2]. Finally, Mettler and Wulf [153] discuss a more recent phenomenon: wearables used for employee analytics. They adopt the perspective of technological affordances [see 182] to explain positive motivational factors for wearable use. For example, they identify the affordance of *establishing workplace security* by, e.g., the wearable detecting bad posture [153, p. 250]. As these motivational factors are wearable-specific, though, they are not generalizable to PA more broadly.

3.4.3. Conclusion

Employees can be motivated to share their data by applying *awareness* or *appeal* strategies. This thesis contains contributions that aim to improve awareness, but previous works identify the additional necessity of appeal strategies. Thereby, we can alleviate the *missing appeal* of PA. For the consumer context, numerous works suggest solutions to increase the appeal of software tools. For the workplace context, though, only few—context-specific—works exist. Most importantly, no systematic guidance on how to increase the appeal of PA for employees is available. Accordingly, we identify a research gap (⇒ **G6: Employee Appeal of PA**)

Part II.

Conceptual and Technical Solutions

4. Rethinking People Analytics With Inverse Transparency by Design

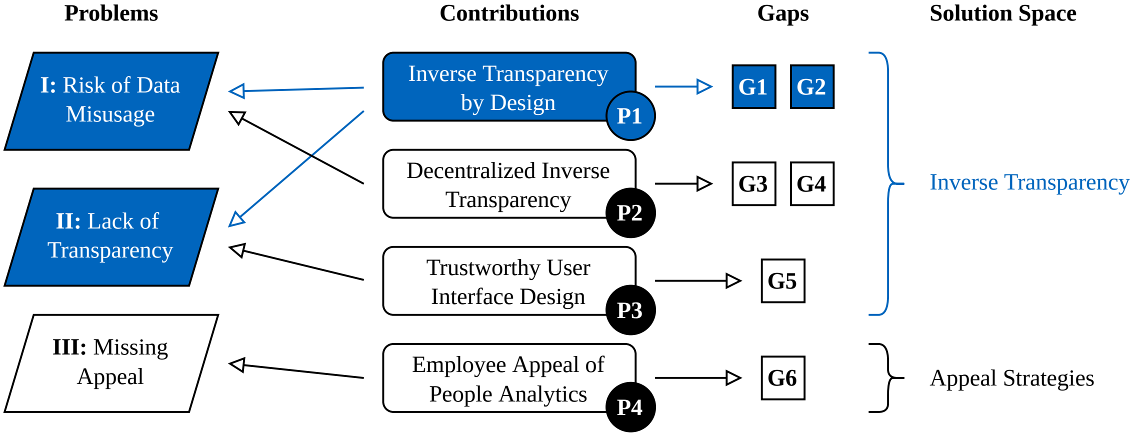


Figure 4.1.: Relationship of our first contribution to the big picture.

Summary. The following summary is partially adapted from our paper [see 1].

- *Problem:* PA are increasingly used in the workplace, but they implicate an elevated *risk of data misusage* (problem **I**) for employees, exacerbated by a *lack of transparency* (problem **II**). This inhibits useful applications of PA.
- *Solution:* We think that PA should provide *inverse transparency* (see Section 2.5), giving data owners more oversight and enabling accountability of data consumers. This requires a comprehensive solution approach and empirical studies of its suitability.
- *Gaps:* Existing research on transparency-enhancing approaches mainly considers the consumer context, which differs from the workplace context of PA. Therefore, it does not offer a solution for inverse transparent PA (**G1**). Furthermore, while empirical studies exist for other contexts and concepts, they are not transferable to our context or do not consider all perspectives (**G2**).
- *Contribution:* We present the concept of *inverse transparency by design* and theoretically deliberate its implications for software design, as well as requirements towards usage log integrity and trustworthiness. Then, we conduct exploratory studies of the developer and user perspective. We find promising results regarding the suitability of our concept. Developers consider it valuable and technically feasible. They

could integrate inverse transparency into PA without inhibiting core functionality. Users, meanwhile, perceive inverse transparency as beneficial and feel empowered by it. They unanimously agree that it would be an improvement for the workplace.

- *Limitations:* Our empirical studies are preliminary and therefore limited in their significance. To best study a fundamental change in the work and interaction of employees, we conducted laboratory studies with university students. Thereby, we could support internal validity and establish causality. Yet, this choice means that the external validity of our results may be limited.

Author Contributions. VZ and AP developed the original context and problem statement. Based on this, VZ conceived of the solution idea. VZ designed the concept of the two studies and discussed it with AP. VZ then conducted the studies and evaluated them. Finally, VZ wrote the manuscript in close discussion with AP.

Copyright Note. Valentin Zieglmeier and Alexander Pretschner. “Rethinking People Analytics With Inverse Transparency by Design.” *Proceedings of the ACM on Human-Computer Interaction* 7.CSCW2, Article 292 (2023).

On the following pages, the published paper is reprinted in accordance with the ACM author rights to reuse their own work in a dissertation. The version of record can be found with the following DOI: `10.1145/3610083`.

Rethinking People Analytics With Inverse Transparency by Design

VALENTIN ZIEGLMEIER, Technical University of Munich, Germany
ALEXANDER PRETSCHNER, Technical University of Munich, Germany

Employees work in increasingly digital environments that enable advanced analytics. Yet, they lack oversight over the systems that process their data. That means that potential analysis errors or hidden biases are hard to uncover. Recent data protection legislation tries to tackle these issues, but it is inadequate. It does not prevent data misusage while at the same time stifling sensible use cases for data.

We think the conflict between data protection and increasingly data-driven systems should be solved differently. When access to an employees' data is given, all usages should be made transparent to them, according to the concept of *inverse transparency*. This allows individuals to benefit from sensible data usage while addressing the potentially harmful consequences of data misusage. To accomplish this, we propose a new design approach for workforce analytics software we refer to as *inverse transparency by design*.

To understand the developer and user perspectives on the proposal, we conduct two exploratory studies with students. First, we let small teams of developers implement analytics tools with inverse transparency by design to uncover how they judge the approach and how it materializes in their developed tools. We find that architectural changes are made without inhibiting core functionality. The developers consider our approach valuable and technically feasible. Second, we conduct a user study over three months to let participants experience the provided inverse transparency and reflect on their experience. The study models a software development workplace where most work processes are already digital. Participants perceive the transparency as beneficial and feel empowered by it. They unanimously agree that it would be an improvement for the workplace. We conclude that inverse transparency by design is a promising approach to realize accepted and responsible people analytics.

CCS Concepts: • **Human-centered computing**; • **Security and privacy** → **Social aspects of security and privacy**; • **Software and its engineering** → *Designing software*;

Additional Key Words and Phrases: Data sovereignty, Privacy by design, HR analytics, Qualitative study

ACM Reference Format:
Valentin Zieglmeier and Alexander Pretschner. 2023. Rethinking People Analytics With Inverse Transparency by Design. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 292 (October 2023), 29 pages. https://doi.org/10.1145/3610083

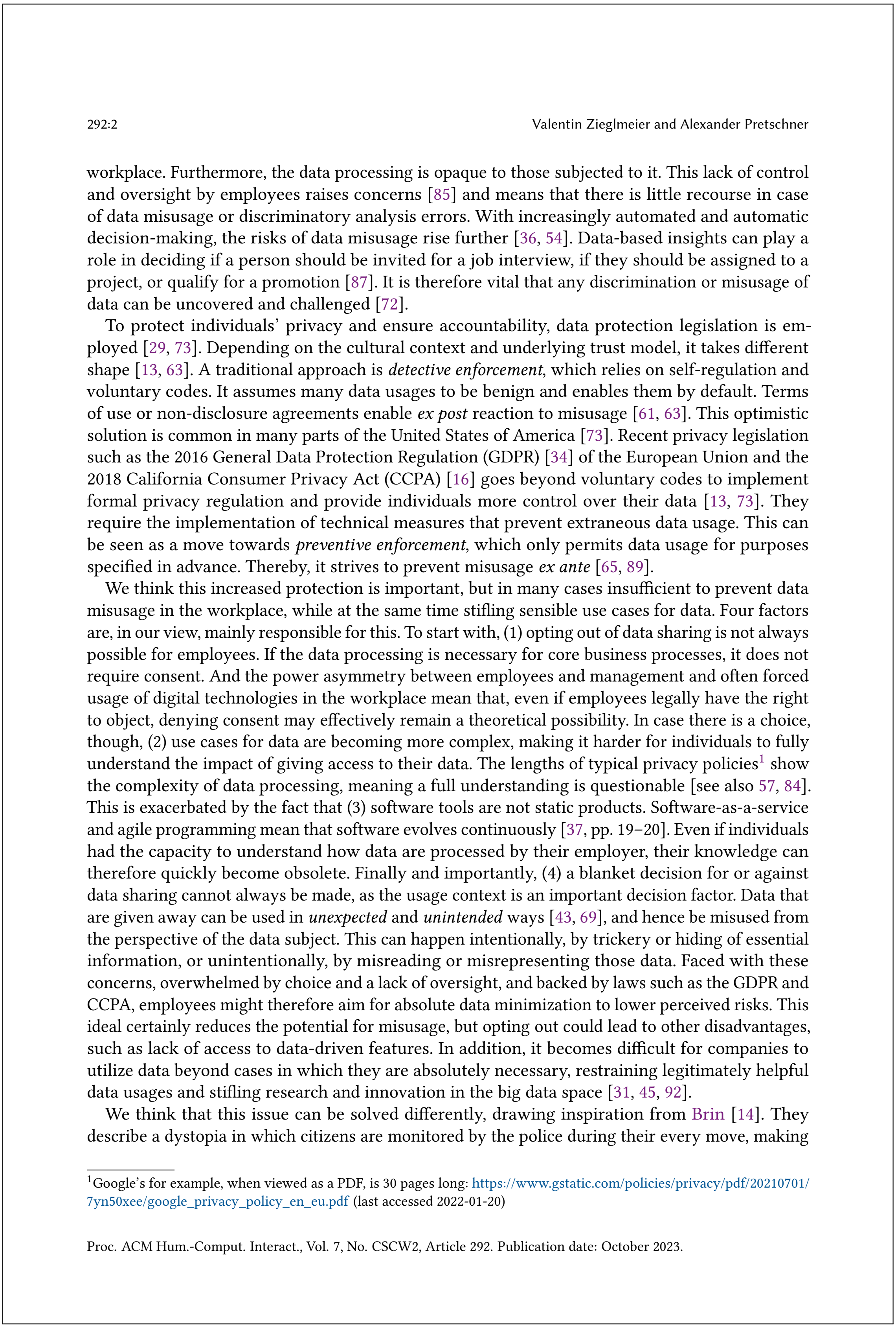
292

1 INTRODUCTION

Workplaces are becoming increasingly digital. Everything from employee communication to the status of tasks and work items is stored and handled digitally. This enables advanced people analytics that process employee data to speed up processes or help with decision-making. But, contrary to the consumer context, employees often have no choice which tools to use in their

Authors' addresses: Valentin Zieglmeier, valentin.zieglmeier@tum.de; Alexander Pretschner, alexander.pretschner@tum.de, Technical University of Munich, TUM School of Computation, Information and Technology, Chair of Software and Systems Engineering, Boltzmannstr. 3, 85748 Garching, Germany.

2573-0142/2023/10-ART292 $15.00
https://doi.org/10.1145/3610083



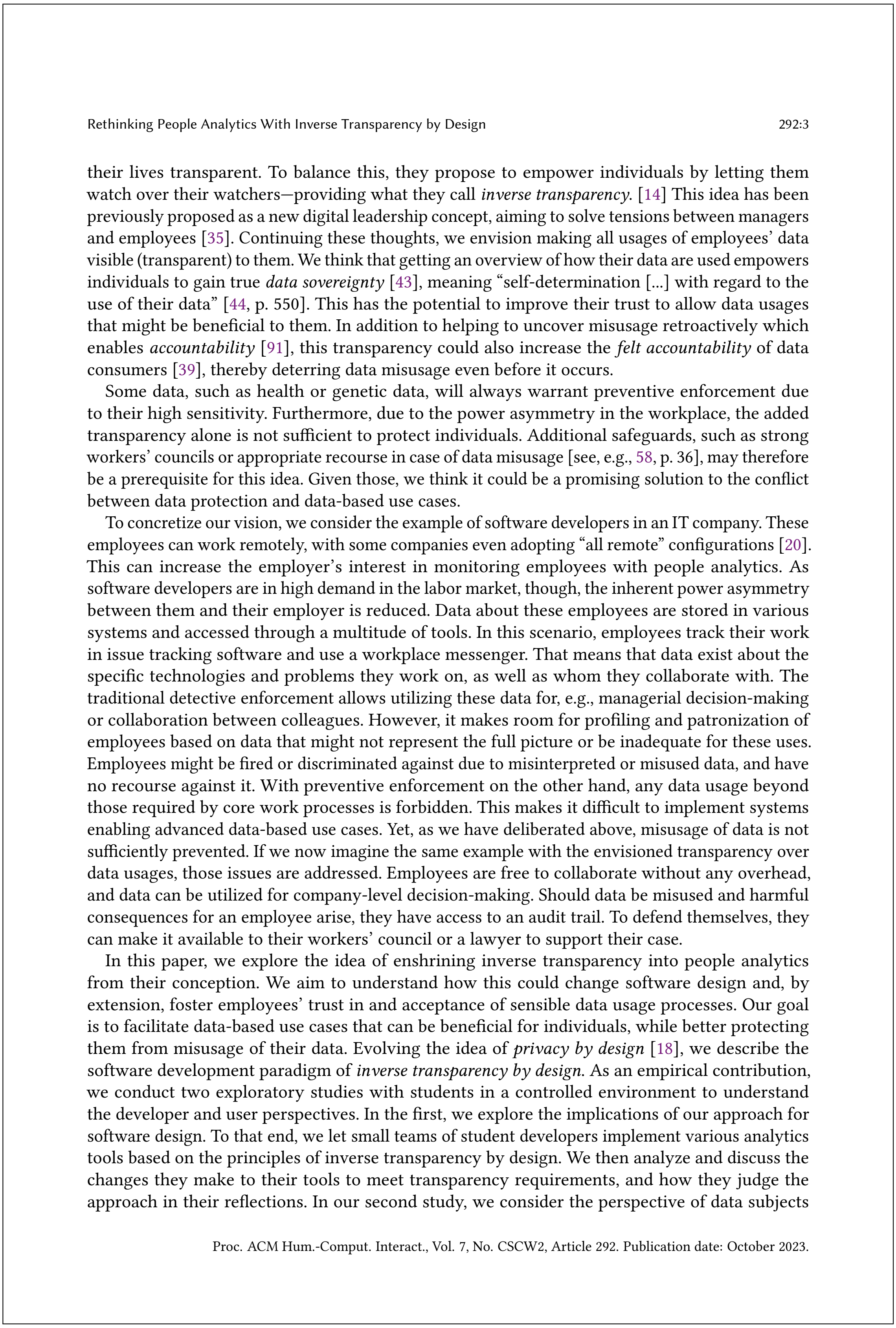
workplace. Furthermore, the data processing is opaque to those subjected to it. This lack of control and oversight by employees raises concerns [85] and means that there is little recourse in case of data misusage or discriminatory analysis errors. With increasingly automated and automatic decision-making, the risks of data misusage rise further [36, 54]. Data-based insights can play a role in deciding if a person should be invited for a job interview, if they should be assigned to a project, or qualify for a promotion [87]. It is therefore vital that any discrimination or misusage of data can be uncovered and challenged [72].

To protect individuals' privacy and ensure accountability, data protection legislation is employed [29, 73]. Depending on the cultural context and underlying trust model, it takes different shape [13, 63]. A traditional approach is *detective enforcement*, which relies on self-regulation and voluntary codes. It assumes many data usages to be benign and enables them by default. Terms of use or non-disclosure agreements enable *ex post* reaction to misusage [61, 63]. This optimistic solution is common in many parts of the United States of America [73]. Recent privacy legislation such as the 2016 General Data Protection Regulation (GDPR) [34] of the European Union and the 2018 California Consumer Privacy Act (CCPA) [16] goes beyond voluntary codes to implement formal privacy regulation and provide individuals more control over their data [13, 73]. They require the implementation of technical measures that prevent extraneous data usage. This can be seen as a move towards *preventive enforcement*, which only permits data usage for purposes specified in advance. Thereby, it strives to prevent misusage *ex ante* [65, 89].

We think this increased protection is important, but in many cases insufficient to prevent data misusage in the workplace, while at the same time stifling sensible use cases for data. Four factors are, in our view, mainly responsible for this. To start with, (1) opting out of data sharing is not always possible for employees. If the data processing is necessary for core business processes, it does not require consent. And the power asymmetry between employees and management and often forced usage of digital technologies in the workplace mean that, even if employees legally have the right to object, denying consent may effectively remain a theoretical possibility. In case there is a choice, though, (2) use cases for data are becoming more complex, making it harder for individuals to fully understand the impact of giving access to their data. The lengths of typical privacy policies[1] show the complexity of data processing, meaning a full understanding is questionable [see also 57, 84]. This is exacerbated by the fact that (3) software tools are not static products. Software-as-a-service and agile programming mean that software evolves continuously [37, pp. 19–20]. Even if individuals had the capacity to understand how data are processed by their employer, their knowledge can therefore quickly become obsolete. Finally and importantly, (4) a blanket decision for or against data sharing cannot always be made, as the usage context is an important decision factor. Data that are given away can be used in *unexpected* and *unintended* ways [43, 69], and hence be misused from the perspective of the data subject. This can happen intentionally, by trickery or hiding of essential information, or unintentionally, by misreading or misrepresenting those data. Faced with these concerns, overwhelmed by choice and a lack of oversight, and backed by laws such as the GDPR and CCPA, employees might therefore aim for absolute data minimization to lower perceived risks. This ideal certainly reduces the potential for misusage, but opting out could lead to other disadvantages, such as lack of access to data-driven features. In addition, it becomes difficult for companies to utilize data beyond cases in which they are absolutely necessary, restraining legitimately helpful data usages and stifling research and innovation in the big data space [31, 45, 92].

We think that this issue can be solved differently, drawing inspiration from Brin [14]. They describe a dystopia in which citizens are monitored by the police during their every move, making

[1] Google's for example, when viewed as a PDF, is 30 pages long: https://www.gstatic.com/policies/privacy/pdf/20210701/7yn50xee/google_privacy_policy_en_eu.pdf (last accessed 2022-01-20)



their lives transparent. To balance this, they propose to empower individuals by letting them watch over their watchers—providing what they call *inverse transparency.* [14] This idea has been previously proposed as a new digital leadership concept, aiming to solve tensions between managers and employees [35]. Continuing these thoughts, we envision making all usages of employees' data visible (transparent) to them. We think that getting an overview of how their data are used empowers individuals to gain true *data sovereignty* [43], meaning "self-determination [...] with regard to the use of their data" [44, p. 550]. This has the potential to improve their trust to allow data usages that might be beneficial to them. In addition to helping to uncover misusage retroactively which enables *accountability* [91], this transparency could also increase the *felt accountability* of data consumers [39], thereby deterring data misusage even before it occurs.

Some data, such as health or genetic data, will always warrant preventive enforcement due to their high sensitivity. Furthermore, due to the power asymmetry in the workplace, the added transparency alone is not sufficient to protect individuals. Additional safeguards, such as strong workers' councils or appropriate recourse in case of data misusage [see, e.g., 58, p. 36], may therefore be a prerequisite for this idea. Given those, we think it could be a promising solution to the conflict between data protection and data-based use cases.

To concretize our vision, we consider the example of software developers in an IT company. These employees can work remotely, with some companies even adopting "all remote" configurations [20]. This can increase the employer's interest in monitoring employees with people analytics. As software developers are in high demand in the labor market, though, the inherent power asymmetry between them and their employer is reduced. Data about these employees are stored in various systems and accessed through a multitude of tools. In this scenario, employees track their work in issue tracking software and use a workplace messenger. That means that data exist about the specific technologies and problems they work on, as well as whom they collaborate with. The traditional detective enforcement allows utilizing these data for, e.g., managerial decision-making or collaboration between colleagues. However, it makes room for profiling and patronization of employees based on data that might not represent the full picture or be inadequate for these uses. Employees might be fired or discriminated against due to misinterpreted or misused data, and have no recourse against it. With preventive enforcement on the other hand, any data usage beyond those required by core work processes is forbidden. This makes it difficult to implement systems enabling advanced data-based use cases. Yet, as we have deliberated above, misusage of data is not sufficiently prevented. If we now imagine the same example with the envisioned transparency over data usages, those issues are addressed. Employees are free to collaborate without any overhead, and data can be utilized for company-level decision-making. Should data be misused and harmful consequences for an employee arise, they have access to an audit trail. To defend themselves, they can make it available to their workers' council or a lawyer to support their case.

In this paper, we explore the idea of enshrining inverse transparency into people analytics from their conception. We aim to understand how this could change software design and, by extension, foster employees' trust in and acceptance of sensible data usage processes. Our goal is to facilitate data-based use cases that can be beneficial for individuals, while better protecting them from misusage of their data. Evolving the idea of *privacy by design* [18], we describe the software development paradigm of *inverse transparency by design.* As an empirical contribution, we conduct two exploratory studies with students in a controlled environment to understand the developer and user perspectives. In the first, we explore the implications of our approach for software design. To that end, we let small teams of student developers implement various analytics tools based on the principles of inverse transparency by design. We then analyze and discuss the changes they make to their tools to meet transparency requirements, and how they judge the approach in their reflections. In our second study, we consider the perspective of data subjects on our concept. Therefore, we conduct a controlled laboratory study with students that mirrors the real-world use case of a software development department. Participants worked for three months in a workplace-like setting utilizing transparency-enabled people analytics. We examine and deliberate the user experience and personal perspectives of participants. Note that, for both studies, we worked with university students in a controlled environment. This was an intentional choice to best study a fundamental change in the work and interactions of employees. Thereby, we remove potential confounding factors [see, e.g., 12, 60] to support internal validity [28] and establish causality [47]. The artificial nature of the studies may limit external validity, though. We discuss the impact and our reasoning in more detail in Section 6.

In all, we contribute a comprehensive conceptualization and preliminary evaluation of inverse transparency by design, a new approach for workplace software development. When studying the developer perspective, we find that building with inverse transparency by design does not inhibit core functionality, with developers considering the concept practical. In our study of the user perspective, participants experiencing inverse transparency in practice find it beneficial and feel empowered by it. Given the choice, they would unanimously opt for it in their workplace. We conclude that moving towards incorporating inverse transparency by design is a promising direction for people analytics.

2 RELATED WORK

We begin by describing other ideas besides inverse transparency that aim to find a middle ground between necessary data protection and sensible uses for personal data. Then, we give an overview of related works that specifically propose to provide transparency over data usages to ensure accountability. Finally, we narrow down further to the technical realization. We discuss other works that aim to change personal data processing such that individual data usages are tracked.

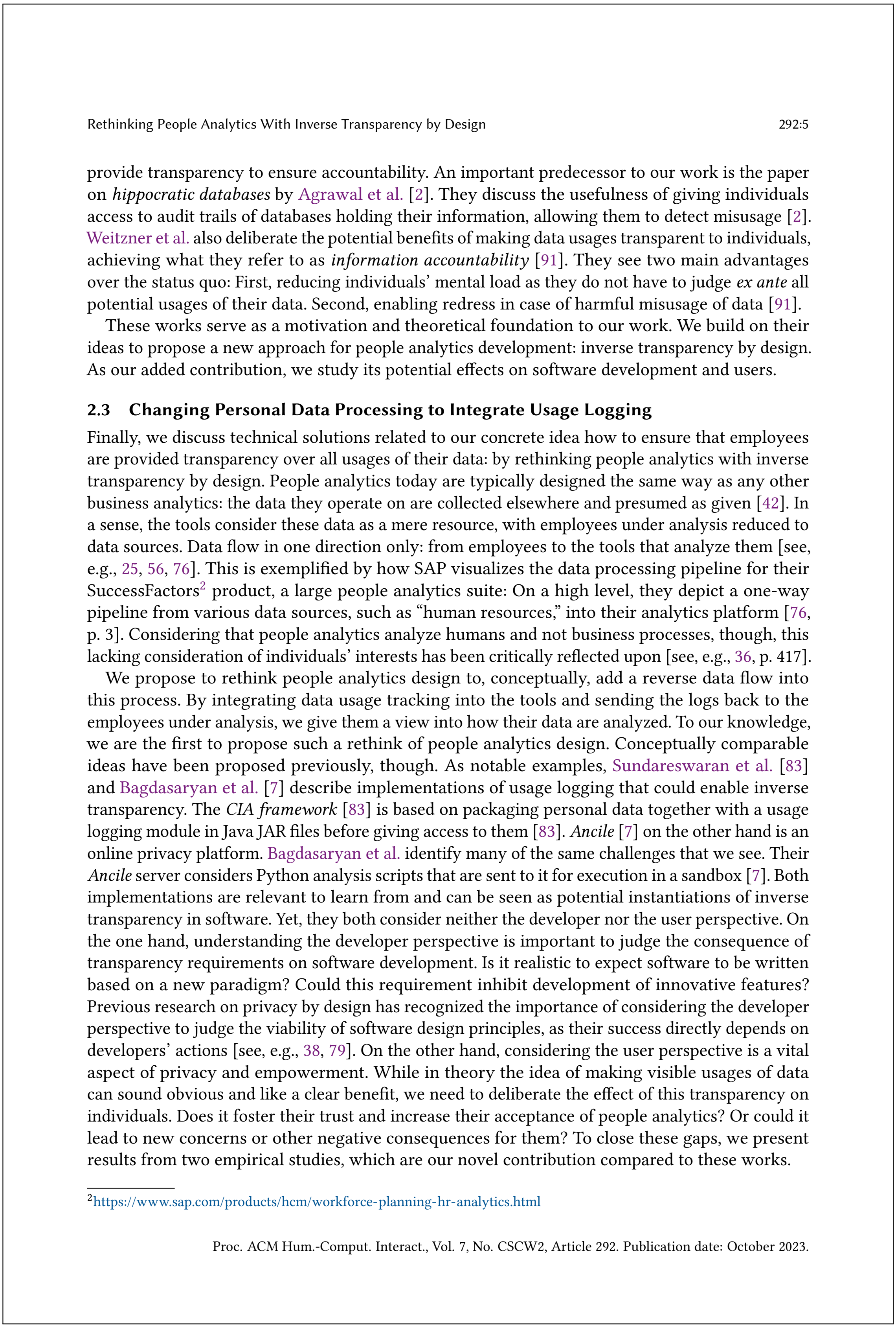
2.1 Balancing Data Protection and Sensible Data Usage

The conflict between data protection and sensible data usage has been considered in various works. Cate observed that "many uses of personal information pose no risk of harm to individuals, while creating significant benefits for data subjects and society more broadly" [17, p. 37]. As a solution, they propose *use-based privacy*, which entails defining collective norms that specify acceptable uses [11, 17]. For example, these could explicitly exclude discriminatory uses of data where they are obvious. Then, the norms are enforced automatically by systems that process data [see, e.g., 11]. A similar strategy underlies *distributed usage control*, which instead of collective norms aims to enforce user-defined usage policies [62] with comparable technical challenges [see, e.g., 90].

Independently of who defines the usage policies, though, it is in our view impossible to know all acceptable usage patterns for data in advance. Therefore, we consider it important to enable more flexibility and instead provide individuals with inverse transparency, allowing them direct oversight over how their data are used. Potential misusage of data can then be handled retroactively, ensuring accountability. Still, the idea to establish collective norms defining appropriate usage of data, enshrined for example in laws or company agreements, is compelling. They could serve as a minimum safeguard for individuals, with inverse transparency helping to protect them for data usage that goes beyond the basic use cases.

2.2 Providing Transparency Over Data Usages to Ensure Accountability

Our work is not the first to introduce the idea of *inverse transparency*, or more broadly aiming to ensure accountability by giving individuals oversight over usages of their data. The general concept of inverse transparency was originally conceived of and presented by Brin [14] (see Section 1). In the software development context specifically, similar concepts have been developed, also aiming to



provide transparency to ensure accountability. An important predecessor to our work is the paper on *hippocratic databases* by Agrawal et al. [2]. They discuss the usefulness of giving individuals access to audit trails of databases holding their information, allowing them to detect misusage [2]. Weitzner et al. also deliberate the potential benefits of making data usages transparent to individuals, achieving what they refer to as *information accountability* [91]. They see two main advantages over the status quo: First, reducing individuals' mental load as they do not have to judge *ex ante* all potential usages of their data. Second, enabling redress in case of harmful misusage of data [91].

These works serve as a motivation and theoretical foundation to our work. We build on their ideas to propose a new approach for people analytics development: inverse transparency by design. As our added contribution, we study its potential effects on software development and users.

2.3 Changing Personal Data Processing to Integrate Usage Logging

Finally, we discuss technical solutions related to our concrete idea how to ensure that employees are provided transparency over all usages of their data: by rethinking people analytics with inverse transparency by design. People analytics today are typically designed the same way as any other business analytics: the data they operate on are collected elsewhere and presumed as given [42]. In a sense, the tools consider these data as a mere resource, with employees under analysis reduced to data sources. Data flow in one direction only: from employees to the tools that analyze them [see, e.g., 25, 56, 76]. This is exemplified by how SAP visualizes the data processing pipeline for their SuccessFactors[2] product, a large people analytics suite: On a high level, they depict a one-way pipeline from various data sources, such as "human resources," into their analytics platform [76, p. 3]. Considering that people analytics analyze humans and not business processes, though, this lacking consideration of individuals' interests has been critically reflected upon [see, e.g., 36, p. 417].

We propose to rethink people analytics design to, conceptually, add a reverse data flow into this process. By integrating data usage tracking into the tools and sending the logs back to the employees under analysis, we give them a view into how their data are analyzed. To our knowledge, we are the first to propose such a rethink of people analytics design. Conceptually comparable ideas have been proposed previously, though. As notable examples, Sundareswaran et al. [83] and Bagdasaryan et al. [7] describe implementations of usage logging that could enable inverse transparency. The *CIA framework* [83] is based on packaging personal data together with a usage logging module in Java JAR files before giving access to them [83]. *Ancile* [7] on the other hand is an online privacy platform. Bagdasaryan et al. identify many of the same challenges that we see. Their *Ancile* server considers Python analysis scripts that are sent to it for execution in a sandbox [7]. Both implementations are relevant to learn from and can be seen as potential instantiations of inverse transparency in software. Yet, they both consider neither the developer nor the user perspective. On the one hand, understanding the developer perspective is important to judge the consequence of transparency requirements on software development. Is it realistic to expect software to be written based on a new paradigm? Could this requirement inhibit development of innovative features? Previous research on privacy by design has recognized the importance of considering the developer perspective to judge the viability of software design principles, as their success directly depends on developers' actions [see, e.g., 38, 79]. On the other hand, considering the user perspective is a vital aspect of privacy and empowerment. While in theory the idea of making visible usages of data can sound obvious and like a clear benefit, we need to deliberate the effect of this transparency on individuals. Does it foster their trust and increase their acceptance of people analytics? Or could it lead to new concerns or other negative consequences for them? To close these gaps, we present results from two empirical studies, which are our novel contribution compared to these works.

[2] https://www.sap.com/products/hcm/workforce-planning-hr-analytics.html

3 CONCEPT

Currently, data usage processes in the workplace happen without oversight of the employees concerned by them. To tackle this, we think that people analytics should be built with inverse transparency by design, a next step after privacy by design. This means they should be built in such a way that data usages can be traced back and attributed. As a first step, we discuss our idea as a theoretical concept, outlining the basic requirements for inverse transparency. Second, we deliberate the potential implications for software design. Finally, we discuss the problems of usage log integrity and system trustworthiness as necessary prerequisites for our concept to function.

From here on, instead of talking specifically about employees or managers, we will refer to the more generic concepts of *data owners* and *data consumers*. This reflects that these roles might be reversed or even inhabited by the same person (accessing their own data). We follow the definition by Pretschner et al., who state that for each datum, there is a *data owner* [64]. They “[possess] the rights to the data” [64, p. 40].[3] They also define the role of the *data consumer* [64]. We personify the data consumer in our concept. In cases of algorithmic data usages, the processing system could be considered the data consumer.

3.1 Requirements for Inverse Transparency

The basis of inverse transparency is to give data owners an overview of all data usages. That requires (1) monitoring every usage of data, (2) verifying the authenticity of these events and storing them, and (3) making this information transparent to data owners. According to the separation of concerns, we can imagine each requirement being fulfilled by individual tools, but the functionality can also be integrated into the operating system or directly into the software that provides data access.

To enable the three steps of our vision, let us therefore consider three (conceptual) components:

(1) *Monitor*: Track data usages
(2) *Safekeeper*: Store monitored usages
(3) *Display*: Make stored usages transparent

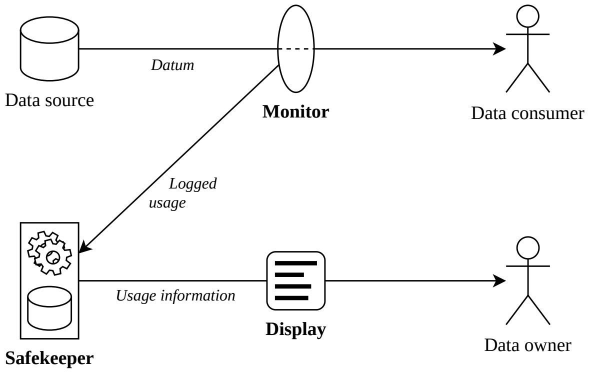


Fig. 1. Providing inverse transparency on a conceptual level. The data consumer accesses a datum from the data source. Their usage is logged and stored. The data owner can now retrieve the logged usage information.

[3] For example, in case the GDPR applies, this would correspond to the “data subject,” meaning the individual identifiable through the data [34, Art. 4].

This process is visualized in Figure 1. On this level of abstraction, we can already postulate: Usage tracking requires oversight over all data processing. Therefore, integrating it into the analysis tool itself is ideal. That matches our concept that these tools be built with inverse transparency by design. But it seems as if the tasks of storing monitored usages and making them transparent could be extracted and shared between multiple tools. In that case, we could reduce development effort of tool developers and potentially increase security, as only one database would need to be protected. For data owners, having a single tool to watch over how their data are used has the potential to reduce mental load and may therefore be preferable, too.

3.2 Implications for Software Design

In order to make all data usages transparent, we need to not only track all occurring usages, but also prevent circumvention of the framework logging. Therefore, it is important to ensure that data never leave controlled environments. Accessing them may only be allowed through tools that provide inverse transparency, thereby ensuring that usages are logged [91]. This is why we envision inverse transparency by design: No software tool should be distributed without enabling this transparency. In our concept, this means that tools need to, on a conceptual level, include a *Monitor* that tracks data processing. We can then imagine companies running their own, private log store providing a standardized API for adding and retrieving log entries. Tools they license or deploy are then required to integrate into the company inverse transparency infrastructure, sending the logged data usages to their private *Safekeeper* API.

At first glance, this seems reasonable. Yet, when we look closer, we find that data are usually not "sent" to data consumers, but continuously move between (sub) systems, are aggregated, copied, converted, or moved. A "data usage" may for example just be a software tool starting up—simply to show a data-driven start page, data accesses can be necessary. Therefore, we have to consider what a data usage is and how software developers can ensure that it is logged. For most cases, our concept of developing with inverse transparency by design deals exactly with this problem: Instead of trying to retrofit usage logging to existing software, we expect developers to already recognize the interests of data owners in the design phase. In some cases, though, building with inverse transparency in mind may require a paradigm shift. Specifically, consider a scenario that we refer to as *ambient usage* of data. We illustrate this with a typical home page of an analytics tool. On load,

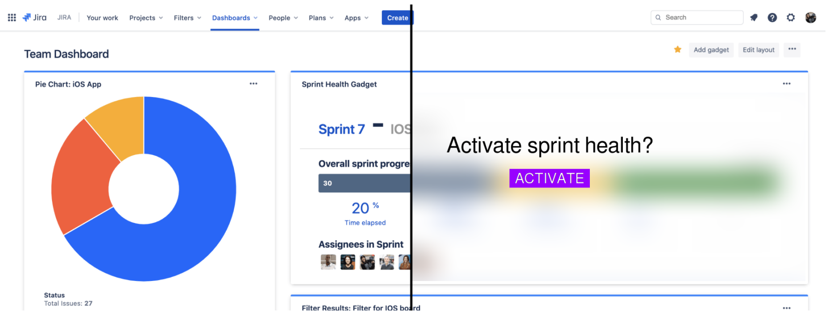
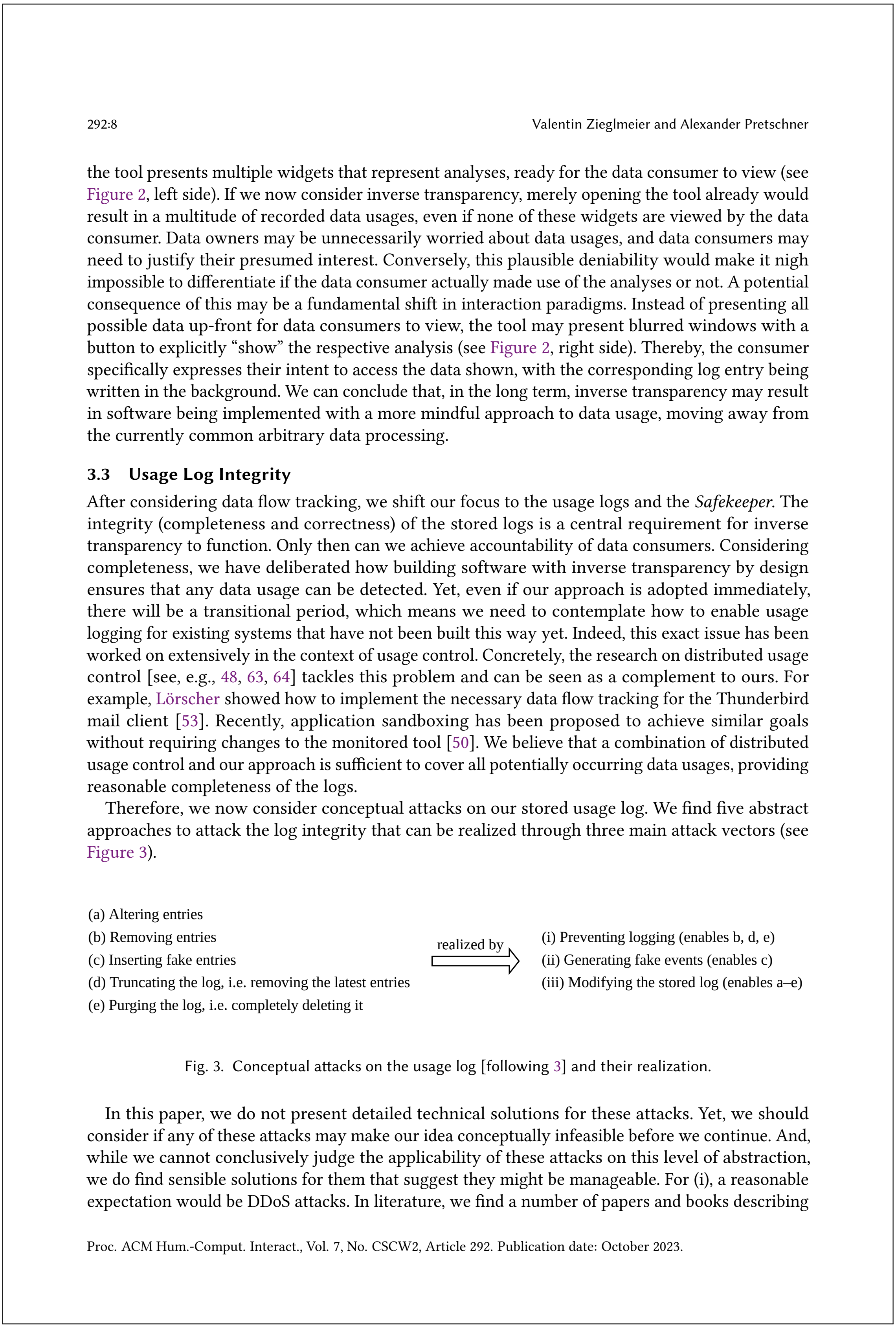


Fig. 2. Example of how interaction paradigms may change as a consequence of inverse transparency, illustrated at the example of Jira Software (image source: Atlassian; modifications: the authors). On the left side, the status quo in Jira: analyses are run and insights presented automatically when the tool is opened, a pattern we refer to as *ambient usage* of data. The right side illustrates how inverse transparency may transform this paradigm: the insights are hidden and need to be explicitly activated, signaling intent by the data consumer.



the tool presents multiple widgets that represent analyses, ready for the data consumer to view (see Figure 2, left side). If we now consider inverse transparency, merely opening the tool already would result in a multitude of recorded data usages, even if none of these widgets are viewed by the data consumer. Data owners may be unnecessarily worried about data usages, and data consumers may need to justify their presumed interest. Conversely, this plausible deniability would make it nigh impossible to differentiate if the data consumer actually made use of the analyses or not. A potential consequence of this may be a fundamental shift in interaction paradigms. Instead of presenting all possible data up-front for data consumers to view, the tool may present blurred windows with a button to explicitly "show" the respective analysis (see Figure 2, right side). Thereby, the consumer specifically expresses their intent to access the data shown, with the corresponding log entry being written in the background. We can conclude that, in the long term, inverse transparency may result in software being implemented with a more mindful approach to data usage, moving away from the currently common arbitrary data processing.

3.3 Usage Log Integrity

After considering data flow tracking, we shift our focus to the usage logs and the *Safekeeper*. The integrity (completeness and correctness) of the stored logs is a central requirement for inverse transparency to function. Only then can we achieve accountability of data consumers. Considering completeness, we have deliberated how building software with inverse transparency by design ensures that any data usage can be detected. Yet, even if our approach is adopted immediately, there will be a transitional period, which means we need to contemplate how to enable usage logging for existing systems that have not been built this way yet. Indeed, this exact issue has been worked on extensively in the context of usage control. Concretely, the research on distributed usage control [see, e.g., 48, 63, 64] tackles this problem and can be seen as a complement to ours. For example, Lörscher showed how to implement the necessary data flow tracking for the Thunderbird mail client [53]. Recently, application sandboxing has been proposed to achieve similar goals without requiring changes to the monitored tool [50]. We believe that a combination of distributed usage control and our approach is sufficient to cover all potentially occurring data usages, providing reasonable completeness of the logs.

Therefore, we now consider conceptual attacks on our stored usage log. We find five abstract approaches to attack the log integrity that can be realized through three main attack vectors (see Figure 3).

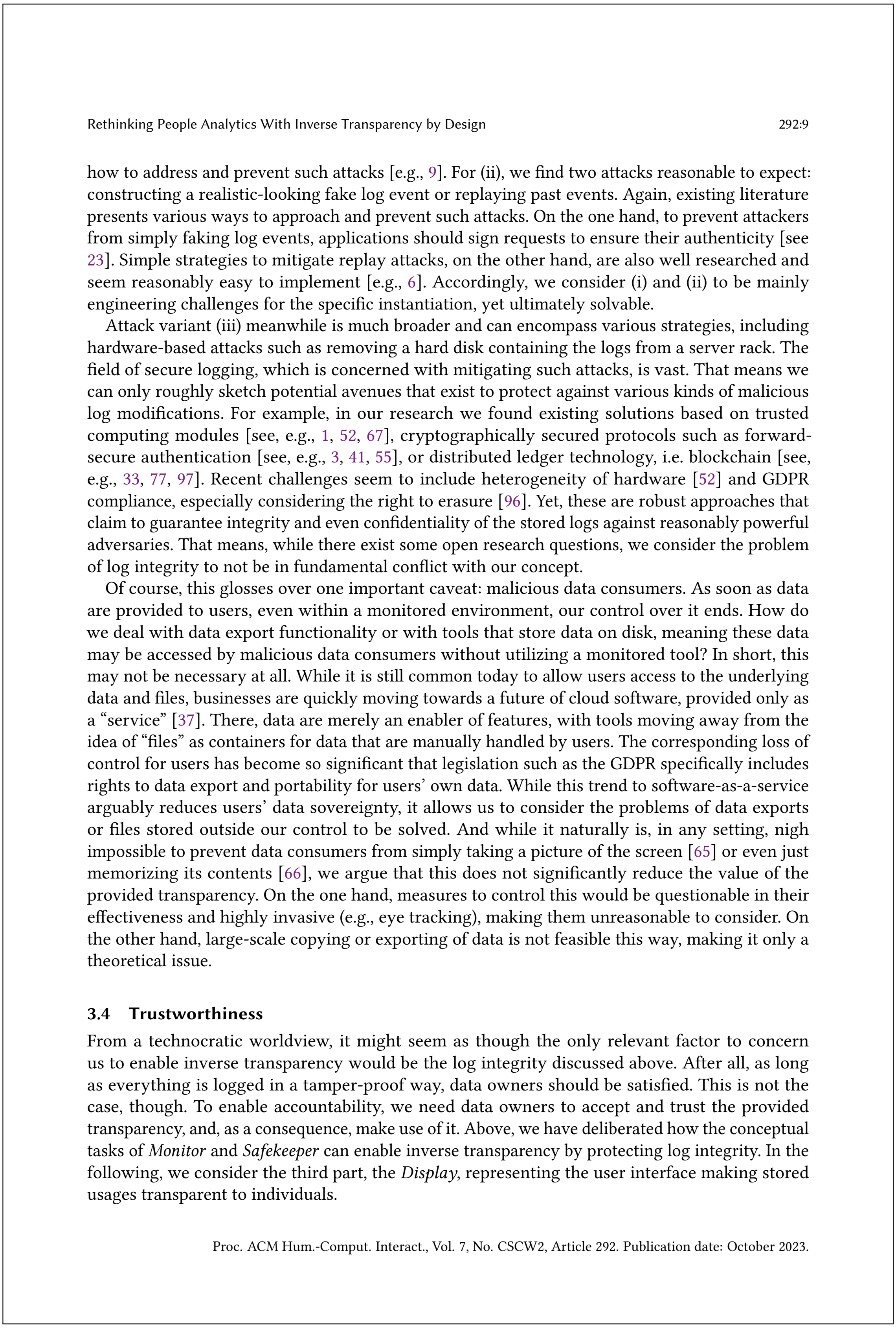
(a) Altering entries
(b) Removing entries
(c) Inserting fake entries
(d) Truncating the log, i.e. removing the latest entries
(e) Purging the log, i.e. completely deleting it

realized by ⇒

(i) Preventing logging (enables b, d, e)
(ii) Generating fake events (enables c)
(iii) Modifying the stored log (enables a–e)

Fig. 3. Conceptual attacks on the usage log [following 3] and their realization.

In this paper, we do not present detailed technical solutions for these attacks. Yet, we should consider if any of these attacks may make our idea conceptually infeasible before we continue. And, while we cannot conclusively judge the applicability of these attacks on this level of abstraction, we do find sensible solutions for them that suggest they might be manageable. For (i), a reasonable expectation would be DDoS attacks. In literature, we find a number of papers and books describing



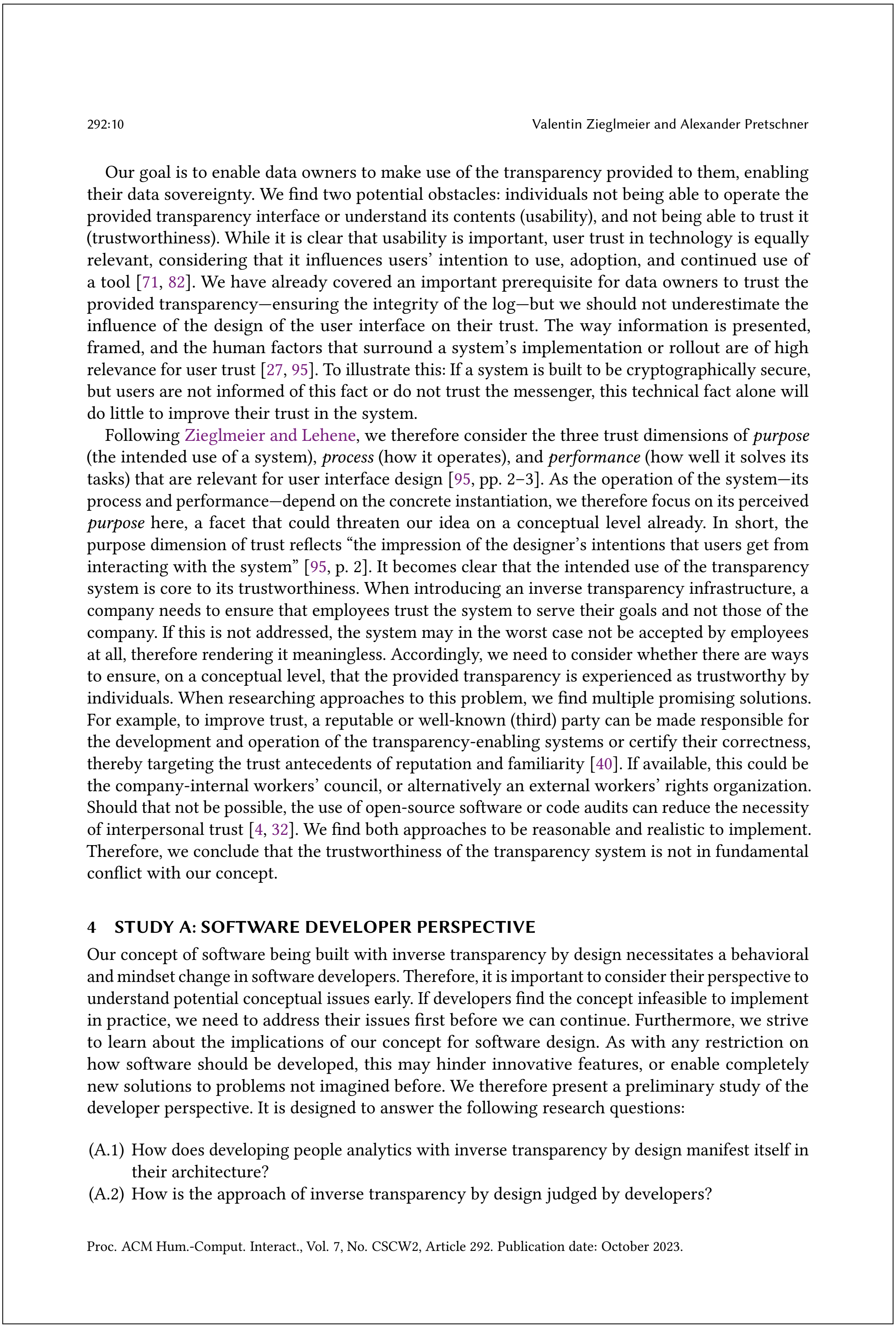
how to address and prevent such attacks [e.g., 9]. For (ii), we find two attacks reasonable to expect: constructing a realistic-looking fake log event or replaying past events. Again, existing literature presents various ways to approach and prevent such attacks. On the one hand, to prevent attackers from simply faking log events, applications should sign requests to ensure their authenticity [see 23]. Simple strategies to mitigate replay attacks, on the other hand, are also well researched and seem reasonably easy to implement [e.g., 6]. Accordingly, we consider (i) and (ii) to be mainly engineering challenges for the specific instantiation, yet ultimately solvable.

Attack variant (iii) meanwhile is much broader and can encompass various strategies, including hardware-based attacks such as removing a hard disk containing the logs from a server rack. The field of secure logging, which is concerned with mitigating such attacks, is vast. That means we can only roughly sketch potential avenues that exist to protect against various kinds of malicious log modifications. For example, in our research we found existing solutions based on trusted computing modules [see, e.g., 1, 52, 67], cryptographically secured protocols such as forward-secure authentication [see, e.g., 3, 41, 55], or distributed ledger technology, i.e. blockchain [see, e.g., 33, 77, 97]. Recent challenges seem to include heterogeneity of hardware [52] and GDPR compliance, especially considering the right to erasure [96]. Yet, these are robust approaches that claim to guarantee integrity and even confidentiality of the stored logs against reasonably powerful adversaries. That means, while there exist some open research questions, we consider the problem of log integrity to not be in fundamental conflict with our concept.

Of course, this glosses over one important caveat: malicious data consumers. As soon as data are provided to users, even within a monitored environment, our control over it ends. How do we deal with data export functionality or with tools that store data on disk, meaning these data may be accessed by malicious data consumers without utilizing a monitored tool? In short, this may not be necessary at all. While it is still common today to allow users access to the underlying data and files, businesses are quickly moving towards a future of cloud software, provided only as a "service" [37]. There, data are merely an enabler of features, with tools moving away from the idea of "files" as containers for data that are manually handled by users. The corresponding loss of control for users has become so significant that legislation such as the GDPR specifically includes rights to data export and portability for users' own data. While this trend to software-as-a-service arguably reduces users' data sovereignty, it allows us to consider the problems of data exports or files stored outside our control to be solved. And while it naturally is, in any setting, nigh impossible to prevent data consumers from simply taking a picture of the screen [65] or even just memorizing its contents [66], we argue that this does not significantly reduce the value of the provided transparency. On the one hand, measures to control this would be questionable in their effectiveness and highly invasive (e.g., eye tracking), making them unreasonable to consider. On the other hand, large-scale copying or exporting of data is not feasible this way, making it only a theoretical issue.

3.4 Trustworthiness

From a technocratic worldview, it might seem as though the only relevant factor to concern us to enable inverse transparency would be the log integrity discussed above. After all, as long as everything is logged in a tamper-proof way, data owners should be satisfied. This is not the case, though. To enable accountability, we need data owners to accept and trust the provided transparency, and, as a consequence, make use of it. Above, we have deliberated how the conceptual tasks of *Monitor* and *Safekeeper* can enable inverse transparency by protecting log integrity. In the following, we consider the third part, the *Display*, representing the user interface making stored usages transparent to individuals.



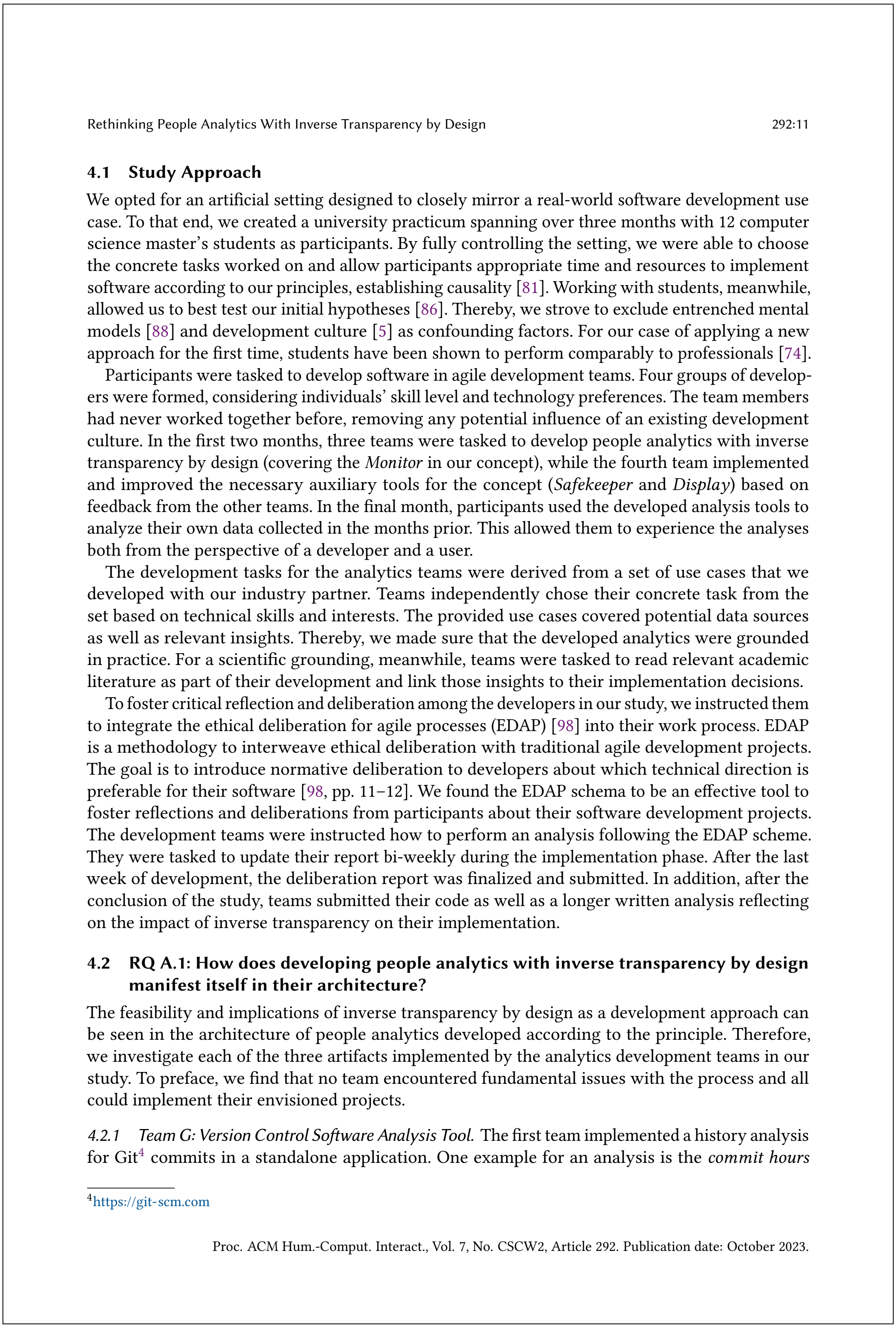
Our goal is to enable data owners to make use of the transparency provided to them, enabling their data sovereignty. We find two potential obstacles: individuals not being able to operate the provided transparency interface or understand its contents (usability), and not being able to trust it (trustworthiness). While it is clear that usability is important, user trust in technology is equally relevant, considering that it influences users' intention to use, adoption, and continued use of a tool [71, 82]. We have already covered an important prerequisite for data owners to trust the provided transparency—ensuring the integrity of the log—but we should not underestimate the influence of the design of the user interface on their trust. The way information is presented, framed, and the human factors that surround a system's implementation or rollout are of high relevance for user trust [27, 95]. To illustrate this: If a system is built to be cryptographically secure, but users are not informed of this fact or do not trust the messenger, this technical fact alone will do little to improve their trust in the system.

Following Zieglmeier and Lehene, we therefore consider the three trust dimensions of *purpose* (the intended use of a system), *process* (how it operates), and *performance* (how well it solves its tasks) that are relevant for user interface design [95, pp. 2–3]. As the operation of the system—its process and performance—depend on the concrete instantiation, we therefore focus on its perceived *purpose* here, a facet that could threaten our idea on a conceptual level already. In short, the purpose dimension of trust reflects "the impression of the designer's intentions that users get from interacting with the system" [95, p. 2]. It becomes clear that the intended use of the transparency system is core to its trustworthiness. When introducing an inverse transparency infrastructure, a company needs to ensure that employees trust the system to serve their goals and not those of the company. If this is not addressed, the system may in the worst case not be accepted by employees at all, therefore rendering it meaningless. Accordingly, we need to consider whether there are ways to ensure, on a conceptual level, that the provided transparency is experienced as trustworthy by individuals. When researching approaches to this problem, we find multiple promising solutions. For example, to improve trust, a reputable or well-known (third) party can be made responsible for the development and operation of the transparency-enabling systems or certify their correctness, thereby targeting the trust antecedents of reputation and familiarity [40]. If available, this could be the company-internal workers' council, or alternatively an external workers' rights organization. Should that not be possible, the use of open-source software or code audits can reduce the necessity of interpersonal trust [4, 32]. We find both approaches to be reasonable and realistic to implement. Therefore, we conclude that the trustworthiness of the transparency system is not in fundamental conflict with our concept.

4 STUDY A: SOFTWARE DEVELOPER PERSPECTIVE

Our concept of software being built with inverse transparency by design necessitates a behavioral and mindset change in software developers. Therefore, it is important to consider their perspective to understand potential conceptual issues early. If developers find the concept infeasible to implement in practice, we need to address their issues first before we can continue. Furthermore, we strive to learn about the implications of our concept for software design. As with any restriction on how software should be developed, this may hinder innovative features, or enable completely new solutions to problems not imagined before. We therefore present a preliminary study of the developer perspective. It is designed to answer the following research questions:

(A.1) How does developing people analytics with inverse transparency by design manifest itself in their architecture?
(A.2) How is the approach of inverse transparency by design judged by developers?



4.1 Study Approach

We opted for an artificial setting designed to closely mirror a real-world software development use case. To that end, we created a university practicum spanning over three months with 12 computer science master's students as participants. By fully controlling the setting, we were able to choose the concrete tasks worked on and allow participants appropriate time and resources to implement software according to our principles, establishing causality [81]. Working with students, meanwhile, allowed us to best test our initial hypotheses [86]. Thereby, we strove to exclude entrenched mental models [88] and development culture [5] as confounding factors. For our case of applying a new approach for the first time, students have been shown to perform comparably to professionals [74].

Participants were tasked to develop software in agile development teams. Four groups of developers were formed, considering individuals' skill level and technology preferences. The team members had never worked together before, removing any potential influence of an existing development culture. In the first two months, three teams were tasked to develop people analytics with inverse transparency by design (covering the *Monitor* in our concept), while the fourth team implemented and improved the necessary auxiliary tools for the concept (*Safekeeper* and *Display*) based on feedback from the other teams. In the final month, participants used the developed analysis tools to analyze their own data collected in the months prior. This allowed them to experience the analyses both from the perspective of a developer and a user.

The development tasks for the analytics teams were derived from a set of use cases that we developed with our industry partner. Teams independently chose their concrete task from the set based on technical skills and interests. The provided use cases covered potential data sources as well as relevant insights. Thereby, we made sure that the developed analytics were grounded in practice. For a scientific grounding, meanwhile, teams were tasked to read relevant academic literature as part of their development and link those insights to their implementation decisions.

To foster critical reflection and deliberation among the developers in our study, we instructed them to integrate the ethical deliberation for agile processes (EDAP) [98] into their work process. EDAP is a methodology to interweave ethical deliberation with traditional agile development projects. The goal is to introduce normative deliberation to developers about which technical direction is preferable for their software [98, pp. 11–12]. We found the EDAP schema to be an effective tool to foster reflections and deliberations from participants about their software development projects. The development teams were instructed how to perform an analysis following the EDAP scheme. They were tasked to update their report bi-weekly during the implementation phase. After the last week of development, the deliberation report was finalized and submitted. In addition, after the conclusion of the study, teams submitted their code as well as a longer written analysis reflecting on the impact of inverse transparency on their implementation.

4.2 RQ A.1: How does developing people analytics with inverse transparency by design manifest itself in their architecture?

The feasibility and implications of inverse transparency by design as a development approach can be seen in the architecture of people analytics developed according to the principle. Therefore, we investigate each of the three artifacts implemented by the analytics development teams in our study. To preface, we find that no team encountered fundamental issues with the process and all could implement their envisioned projects.

4.2.1 Team G: Version Control Software Analysis Tool. The first team implemented a history analysis for Git[4] commits in a standalone application. One example for an analysis is the *commit hours*

[4] https://git-scm.com

overview. It summarizes for each developer the number of commits per hour of the day. The analyses are implemented as tabs that the data consumer can switch between. To implement inverse transparency, they include an additional screen in their tool that may not be necessary otherwise: a selection screen for the requested analysis. It manifests a simplistic solution for the issue of *ambient usage* in the context of inverse transparency that we deliberated above (see Section 3.2). Only those analyses that the data consumer explicitly selects are loaded, with a usage log immediately created for every selected one (see Figure 4). After the analysis report has been created, the tool logs no additional accesses even if it is opened again, which is interesting albeit inconsequential. We consider this the most basic form of implementing inverse transparency: the interaction paradigm and provided features do not have to change, instead the tool adds a step to the process before starting the data processing and just records any potential data usage immediately.

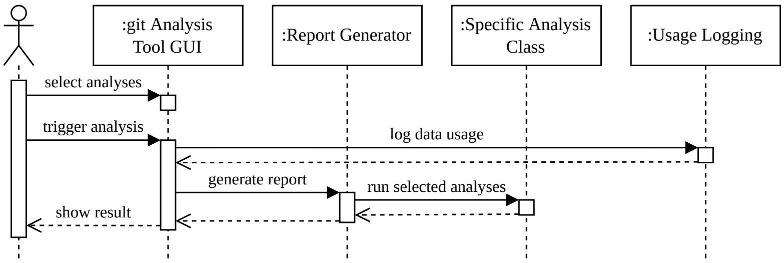


Fig. 4. Sequence diagram of how team G realized inverse transparency. Data consumers have to explicitly select analyses before any data are processed. This data usage is logged once, before the report is generated.

Team G worked together with team S on their implementation of inverse transparency. They implemented a shared library that was then integrated by both into their code. Hereby, we could already see potential synergies when software is implemented with inverse transparency by design: Basic functionality does not have to be reimplemented for every tool and may instead be shared, minimizing the development overhead.

4.2.2 Team J: Issue Tracking Software Analysis Plugins. Team J implemented four analysis gadgets for Jira Software,[5] distributed as one plugin. One example is the *supporter analysis*. It ranks team members by who performed most code reviews. Data consumers can then choose any of the gadgets independently to be added to their main application dashboard, where they have to be explicitly activated to run. Considering its architecture, the artifact implements inverse transparency in a notable fashion—a single module provides both data retrieval and usage logging functionality, allowing the various plugins to share this code. Thereby, developers do not need to consider inverse transparency for every newly created analysis. In the front end code, this manifests as a single function invocation that logs the data usage via the shared back end (see Figure 5). After it completes, the tool continues its operation. On failure, it shows an error message, ensuring that the logging successfully completes before data are presented to the data consumer.

4.2.3 Team S: Chat Analysis Tool. Team S implemented a standalone analysis tool for Slack[6] workspaces. For example, one analysis their tool can perform is the *network analysis*. It creates a social network graph based on who is “mentioned” by whom in their messages. Various analyses such as this can be chosen on the main screen and then independently triggered with a button. For

[5] https://www.atlassian.com/software/jira
[6] https://slack.com/

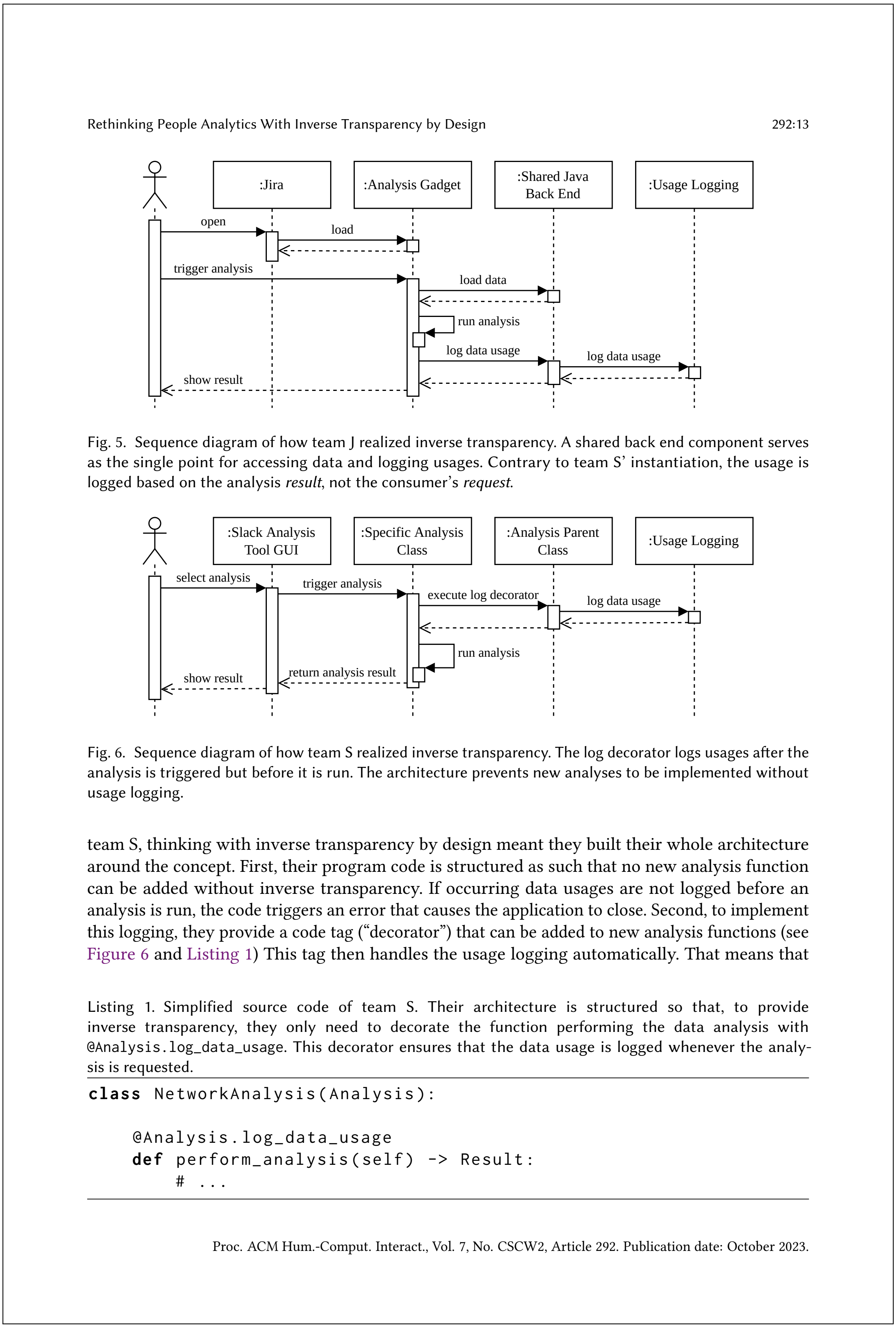


Fig. 5. Sequence diagram of how team J realized inverse transparency. A shared back end component serves as the single point for accessing data and logging usages. Contrary to team S' instantiation, the usage is logged based on the analysis *result*, not the consumer's *request*.

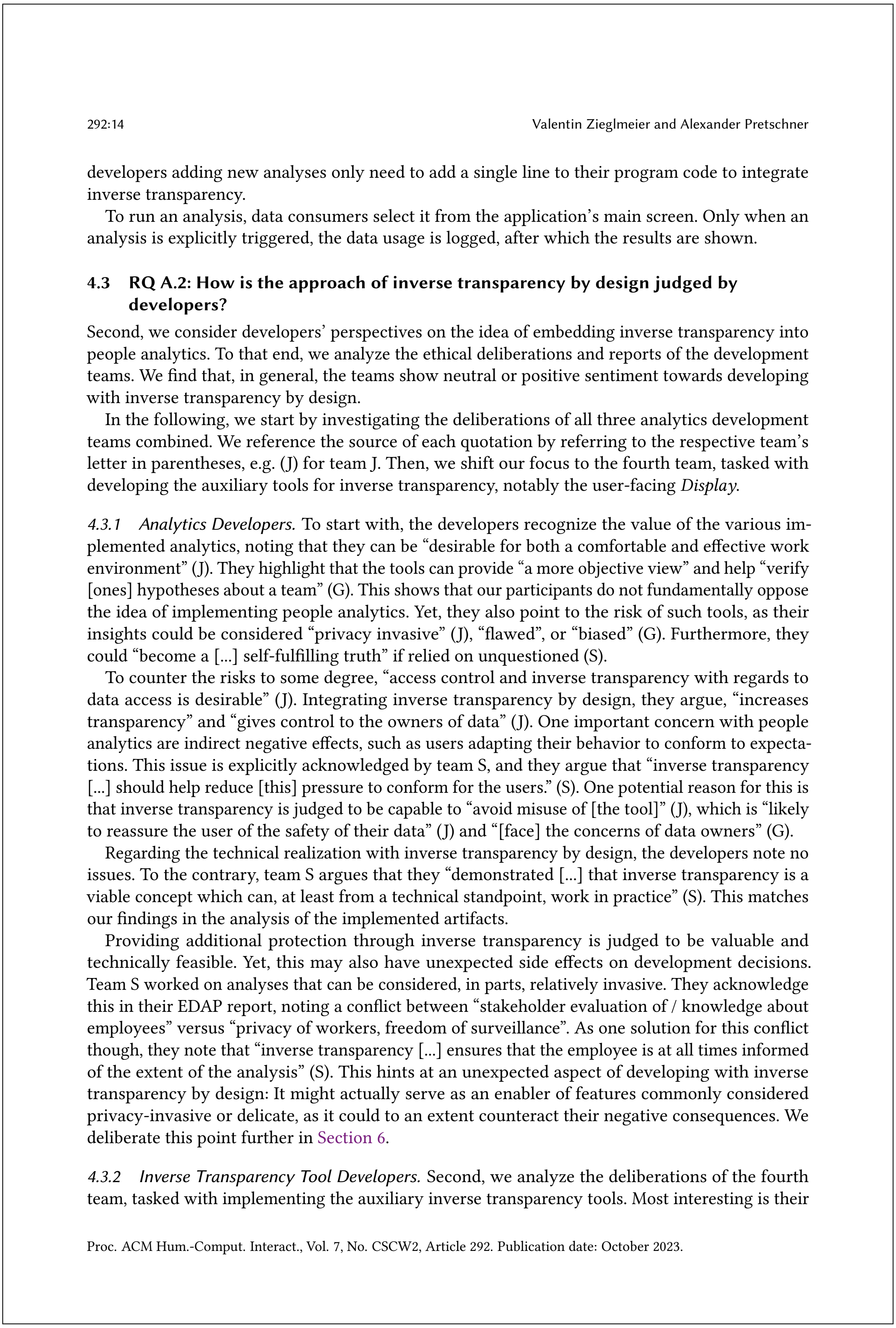
Fig. 6. Sequence diagram of how team S realized inverse transparency. The log decorator logs usages after the analysis is triggered but before it is run. The architecture prevents new analyses to be implemented without usage logging.

team S, thinking with inverse transparency by design meant they built their whole architecture around the concept. First, their program code is structured as such that no new analysis function can be added without inverse transparency. If occurring data usages are not logged before an analysis is run, the code triggers an error that causes the application to close. Second, to implement this logging, they provide a code tag ("decorator") that can be added to new analysis functions (see Figure 6 and Listing 1) This tag then handles the usage logging automatically. That means that

Listing 1. Simplified source code of team S. Their architecture is structured so that, to provide inverse transparency, they only need to decorate the function performing the data analysis with `@Analysis.log_data_usage`. This decorator ensures that the data usage is logged whenever the analysis is requested.

```
class NetworkAnalysis(Analysis):

    @Analysis.log_data_usage
    def perform_analysis(self) -> Result:
        # ...
```



developers adding new analyses only need to add a single line to their program code to integrate inverse transparency.

To run an analysis, data consumers select it from the application's main screen. Only when an analysis is explicitly triggered, the data usage is logged, after which the results are shown.

4.3 RQ A.2: How is the approach of inverse transparency by design judged by developers?

Second, we consider developers' perspectives on the idea of embedding inverse transparency into people analytics. To that end, we analyze the ethical deliberations and reports of the development teams. We find that, in general, the teams show neutral or positive sentiment towards developing with inverse transparency by design.

In the following, we start by investigating the deliberations of all three analytics development teams combined. We reference the source of each quotation by referring to the respective team's letter in parentheses, e.g. (J) for team J. Then, we shift our focus to the fourth team, tasked with developing the auxiliary tools for inverse transparency, notably the user-facing *Display*.

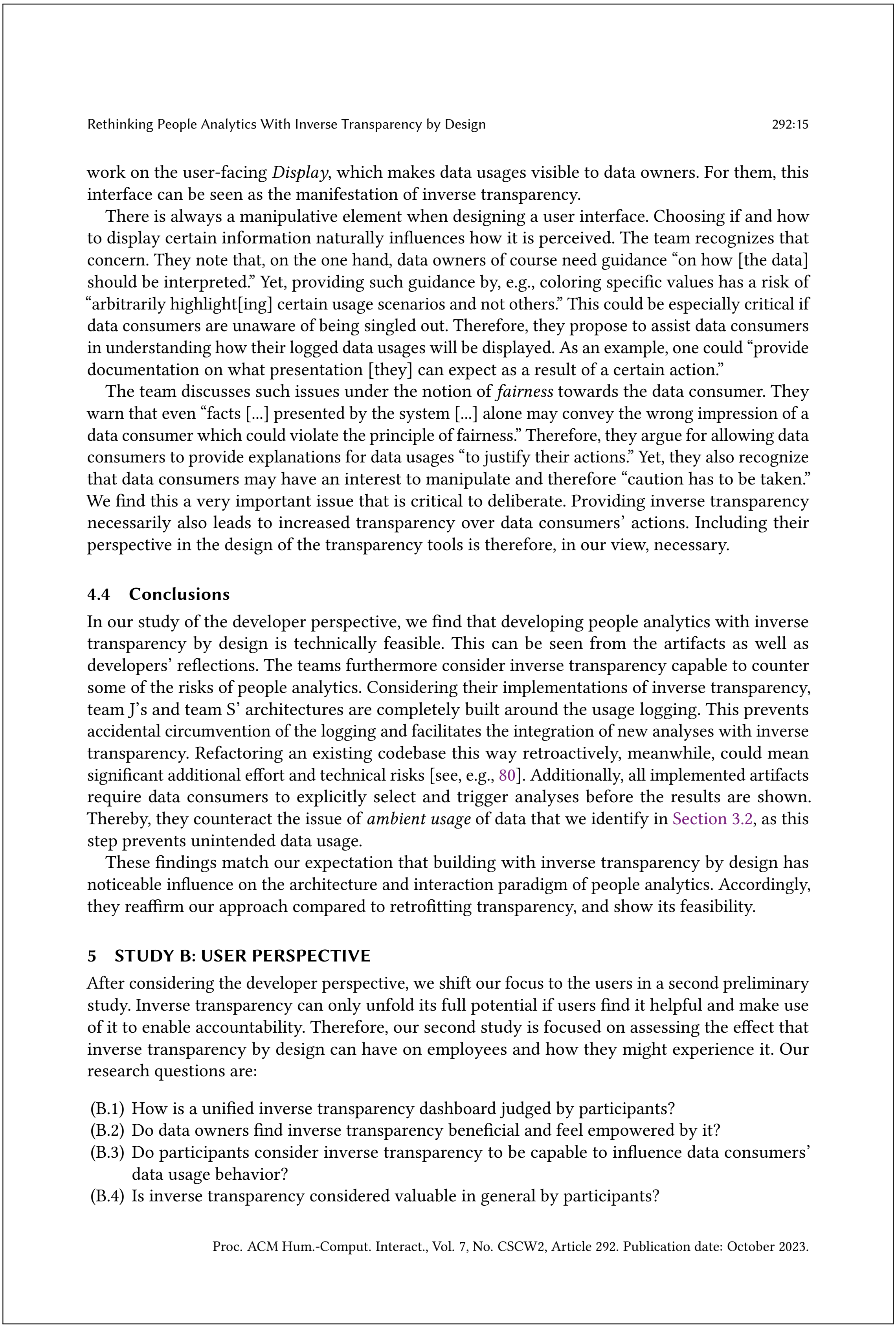
4.3.1 Analytics Developers. To start with, the developers recognize the value of the various implemented analytics, noting that they can be “desirable for both a comfortable and effective work environment” (J). They highlight that the tools can provide “a more objective view” and help “verify [ones] hypotheses about a team” (G). This shows that our participants do not fundamentally oppose the idea of implementing people analytics. Yet, they also point to the risk of such tools, as their insights could be considered “privacy invasive” (J), “flawed”, or “biased” (G). Furthermore, they could “become a [...] self-fulfilling truth” if relied on unquestioned (S).

To counter the risks to some degree, “access control and inverse transparency with regards to data access is desirable” (J). Integrating inverse transparency by design, they argue, “increases transparency” and “gives control to the owners of data” (J). One important concern with people analytics are indirect negative effects, such as users adapting their behavior to conform to expectations. This issue is explicitly acknowledged by team S, and they argue that “inverse transparency [...] should help reduce [this] pressure to conform for the users.” (S). One potential reason for this is that inverse transparency is judged to be capable to “avoid misuse of [the tool]” (J), which is “likely to reassure the user of the safety of their data” (J) and “[face] the concerns of data owners” (G).

Regarding the technical realization with inverse transparency by design, the developers note no issues. To the contrary, team S argues that they “demonstrated [...] that inverse transparency is a viable concept which can, at least from a technical standpoint, work in practice” (S). This matches our findings in the analysis of the implemented artifacts.

Providing additional protection through inverse transparency is judged to be valuable and technically feasible. Yet, this may also have unexpected side effects on development decisions. Team S worked on analyses that can be considered, in parts, relatively invasive. They acknowledge this in their EDAP report, noting a conflict between “stakeholder evaluation of / knowledge about employees” versus “privacy of workers, freedom of surveillance”. As one solution for this conflict though, they note that “inverse transparency [...] ensures that the employee is at all times informed of the extent of the analysis” (S). This hints at an unexpected aspect of developing with inverse transparency by design: It might actually serve as an enabler of features commonly considered privacy-invasive or delicate, as it could to an extent counteract their negative consequences. We deliberate this point further in Section 6.

4.3.2 Inverse Transparency Tool Developers. Second, we analyze the deliberations of the fourth team, tasked with implementing the auxiliary inverse transparency tools. Most interesting is their



work on the user-facing *Display*, which makes data usages visible to data owners. For them, this interface can be seen as the manifestation of inverse transparency.

There is always a manipulative element when designing a user interface. Choosing if and how to display certain information naturally influences how it is perceived. The team recognizes that concern. They note that, on the one hand, data owners of course need guidance "on how [the data] should be interpreted." Yet, providing such guidance by, e.g., coloring specific values has a risk of "arbitrarily highlight[ing] certain usage scenarios and not others." This could be especially critical if data consumers are unaware of being singled out. Therefore, they propose to assist data consumers in understanding how their logged data usages will be displayed. As an example, one could "provide documentation on what presentation [they] can expect as a result of a certain action."

The team discusses such issues under the notion of *fairness* towards the data consumer. They warn that even "facts [...] presented by the system [...] alone may convey the wrong impression of a data consumer which could violate the principle of fairness." Therefore, they argue for allowing data consumers to provide explanations for data usages "to justify their actions." Yet, they also recognize that data consumers may have an interest to manipulate and therefore "caution has to be taken." We find this a very important issue that is critical to deliberate. Providing inverse transparency necessarily also leads to increased transparency over data consumers' actions. Including their perspective in the design of the transparency tools is therefore, in our view, necessary.

4.4 Conclusions

In our study of the developer perspective, we find that developing people analytics with inverse transparency by design is technically feasible. This can be seen from the artifacts as well as developers' reflections. The teams furthermore consider inverse transparency capable to counter some of the risks of people analytics. Considering their implementations of inverse transparency, team J's and team S' architectures are completely built around the usage logging. This prevents accidental circumvention of the logging and facilitates the integration of new analyses with inverse transparency. Refactoring an existing codebase this way retroactively, meanwhile, could mean significant additional effort and technical risks [see, e.g., 80]. Additionally, all implemented artifacts require data consumers to explicitly select and trigger analyses before the results are shown. Thereby, they counteract the issue of *ambient usage* of data that we identify in Section 3.2, as this step prevents unintended data usage.

These findings match our expectation that building with inverse transparency by design has noticeable influence on the architecture and interaction paradigm of people analytics. Accordingly, they reaffirm our approach compared to retrofitting transparency, and show its feasibility.

5 STUDY B: USER PERSPECTIVE

After considering the developer perspective, we shift our focus to the users in a second preliminary study. Inverse transparency can only unfold its full potential if users find it helpful and make use of it to enable accountability. Therefore, our second study is focused on assessing the effect that inverse transparency by design can have on employees and how they might experience it. Our research questions are:

(B.1) How is a unified inverse transparency dashboard judged by participants?
(B.2) Do data owners find inverse transparency beneficial and feel empowered by it?
(B.3) Do participants consider inverse transparency to be capable to influence data consumers' data usage behavior?
(B.4) Is inverse transparency considered valuable in general by participants?

5.1 Study Approach

We conducted a laboratory study with students, which was specifically designed for our research objectives. Inverse transparency represents a fundamental change in the working environment and interactions of employees. Introducing it to an existing workplace could have triggered confounding change management issues, most notably individual's resistance to change [12], which can limit the success of change initiatives [51]. By working with students and fully controlling the study environment, we could ensure that inverse transparency was actually experienced and used over an extended period of time [47]. Furthermore, having removed external influences of an existing workplace, such as schedule pressure [60], means that we can cleanly attribute any observed effects to inverse transparency instead of potential confounders [28]. Accordingly, we created a university practicum running over three months for our study. To increase the representativeness of the study environment, we emulated the real-world use case of a remote software development workplace as closely as possible. That means having small teams of developers build software in agile teams, each lead by a team lead. Work items and their status were tracked in the issue tracking software Jira Software, messages were exchanged over the business messenger Slack, and code was versioned with the distributed version control software Git. Teams only interacted through digital collaboration tools, representing an all-remote configuration.

In total, 15 master's or final year bachelor's students in computer science were tasked to conduct work in the setting described above. Working with computer science students specifically allowed us to realistically model the software development use case. They were split up in four groups of four students each, one with three (see Figure 7). Each group decided on a team lead to guide the development process and serve as the data consumer. We explicitly gave team leads the task to conduct data-driven management, utilizing analysis tools built with inverse transparency by design to analyze their team members' data and write reports on their performance. Due to the all-remote configuration, team leads had to rely on these analytics to get a full picture of their teams. The other team members worked as developers, representing the data owners in our concept. They were instructed to specifically utilize those data-driven collaboration tools that their team leads' analysis tools were compatible with. That means they worked with Jira, Slack and Git, as described above. The data collected in these tools were then made available to their respective team leads to analyze via people analytics built with inverse transparency by design.

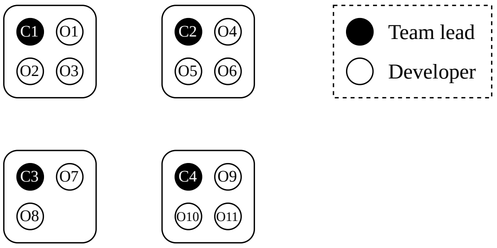


Fig. 7. Participant organization in our study ($n = 15$). Three teams of four and one of three were formed. Each team had one team lead, serving as a data consumer (C1–C4), with the rest of the team working as developers, representing data owners (O1–O11).

We employ a mixed-method evaluation design. During the process, regular written self-reflections were submitted. In addition, participants answered two questionnaires after the study, one of which covers their experience with inverse transparency over the course of the study, the other their opinion of inverse transparency in general (the questions are listed in Appendix A.2). Our findings summarize the results from all of these evaluations.

5.2 Workflow and Utilized Tools

The teams in study B worked with people analytics that were built by the developers in study A with inverse transparency by design. The employed workflow, visualized in Figure 8, was an instantiation of our concept (see Figure 1 on page 6), with developers representing data owners and team leads representing data consumers. Developers used the tools Jira, Slack, and Git for their work and collaboration. The data collected in these tools were then accessed by the team leads via the people analytics created in study B by team J, team S, and team G. As these analytics were built with inverse transparency by design, any data usage was logged and stored in a database (the *Safekeeper*) to be made available to the data owners. To access usage information relating to their data, developers were given access to a transparency dashboard (the *Display*) that provides a direct view into the tracked usage logs. This dashboard was developed by the fourth team in study A as part of the auxiliary transparency tools, in addition to the *Safekeeper* that stored usage data, and a single sign-on server. Collectively, we refer to these tools as the *inverse transparency toolchain* [93].

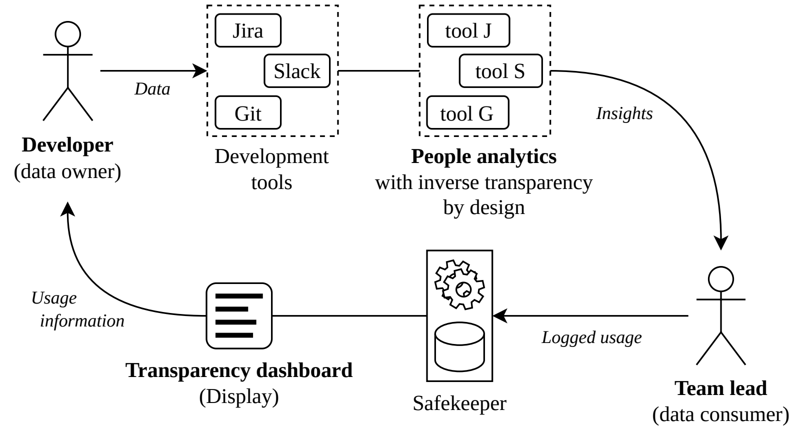
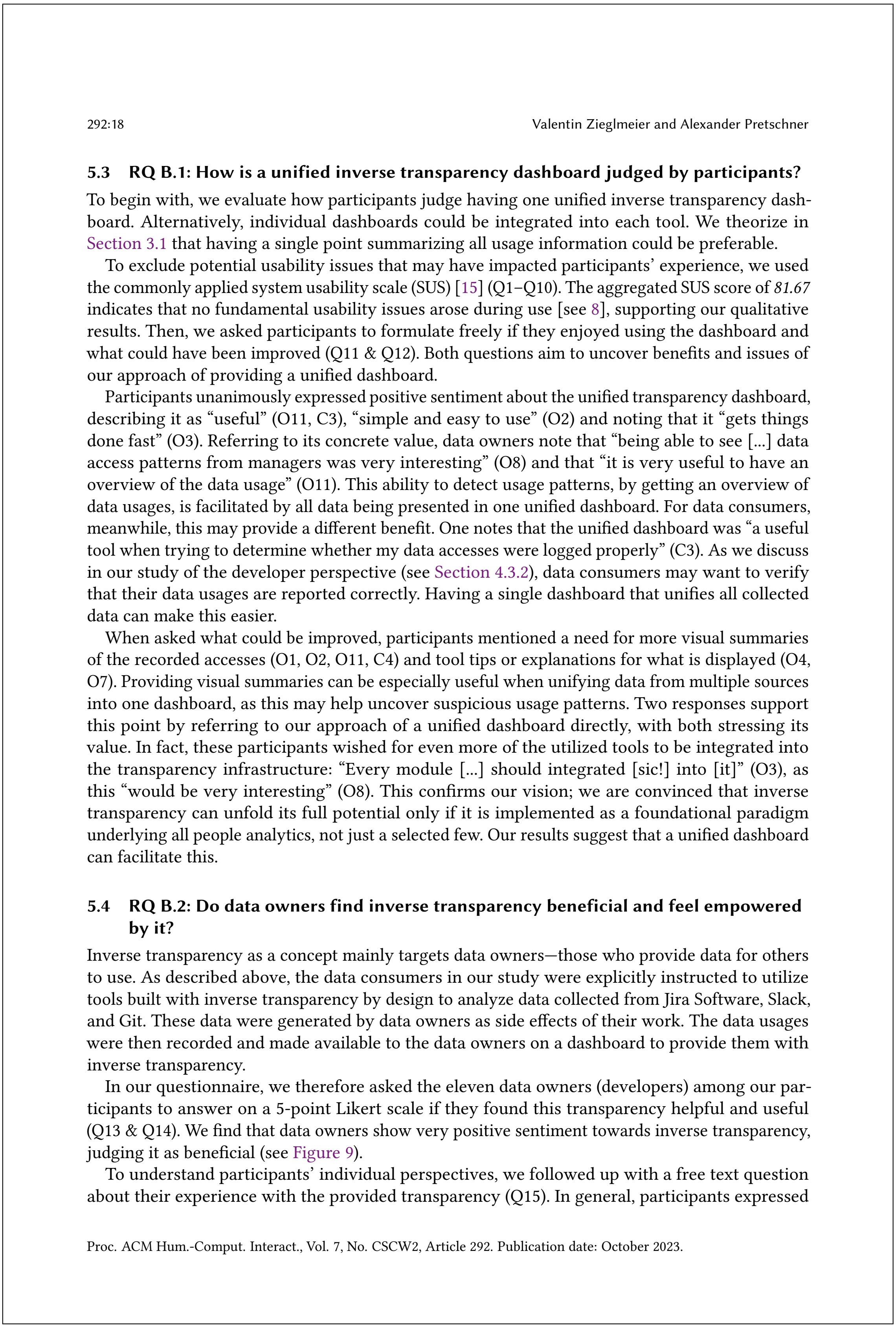


Fig. 8. The workflow and utilized tools in study B. Developers worked with development tools that collect data on their work and collaboration. These data were then analyzed by the team leads with the people analytics created in study A. The tools logged their data usage, storing this data in the *Safekeeper*. The usage information was then made available to developers via the transparency dashboard.

The purpose of the transparency dashboard is to make usage information available to data owners. It is a standalone tool that unifies all logged usage data relating to an individual, independently of the concrete analytics used. All study participants got accounts for the inverse transparency toolchain that they could use to log into the transparency dashboard. To illustrate our following elaborations on the dashboard, find a screenshot of it in Appendix A.1. After logging in, users find two views into the data: an overview area and a detailed table with individual entries. The overview area is meant to give a quick view how one's data were used in the past days. To find out exactly what was accessed, by whom, and when, the table and its filtering functionality can be used.



5.3 RQ B.1: How is a unified inverse transparency dashboard judged by participants?

To begin with, we evaluate how participants judge having one unified inverse transparency dashboard. Alternatively, individual dashboards could be integrated into each tool. We theorize in Section 3.1 that having a single point summarizing all usage information could be preferable.

To exclude potential usability issues that may have impacted participants' experience, we used the commonly applied system usability scale (SUS) [15] (Q1–Q10). The aggregated SUS score of *81.67* indicates that no fundamental usability issues arose during use [see 8], supporting our qualitative results. Then, we asked participants to formulate freely if they enjoyed using the dashboard and what could have been improved (Q11 & Q12). Both questions aim to uncover benefits and issues of our approach of providing a unified dashboard.

Participants unanimously expressed positive sentiment about the unified transparency dashboard, describing it as "useful" (O11, C3), "simple and easy to use" (O2) and noting that it "gets things done fast" (O3). Referring to its concrete value, data owners note that "being able to see [...] data access patterns from managers was very interesting" (O8) and that "it is very useful to have an overview of the data usage" (O11). This ability to detect usage patterns, by getting an overview of data usages, is facilitated by all data being presented in one unified dashboard. For data consumers, meanwhile, this may provide a different benefit. One notes that the unified dashboard was "a useful tool when trying to determine whether my data accesses were logged properly" (C3). As we discuss in our study of the developer perspective (see Section 4.3.2), data consumers may want to verify that their data usages are reported correctly. Having a single dashboard that unifies all collected data can make this easier.

When asked what could be improved, participants mentioned a need for more visual summaries of the recorded accesses (O1, O2, O11, C4) and tool tips or explanations for what is displayed (O4, O7). Providing visual summaries can be especially useful when unifying data from multiple sources into one dashboard, as this may help uncover suspicious usage patterns. Two responses support this point by referring to our approach of a unified dashboard directly, with both stressing its value. In fact, these participants wished for even more of the utilized tools to be integrated into the transparency infrastructure: "Every module [...] should integrated [sic!] into [it]" (O3), as this "would be very interesting" (O8). This confirms our vision; we are convinced that inverse transparency can unfold its full potential only if it is implemented as a foundational paradigm underlying all people analytics, not just a selected few. Our results suggest that a unified dashboard can facilitate this.

5.4 RQ B.2: Do data owners find inverse transparency beneficial and feel empowered by it?

Inverse transparency as a concept mainly targets data owners—those who provide data for others to use. As described above, the data consumers in our study were explicitly instructed to utilize tools built with inverse transparency by design to analyze data collected from Jira Software, Slack, and Git. These data were generated by data owners as side effects of their work. The data usages were then recorded and made available to the data owners on a dashboard to provide them with inverse transparency.

In our questionnaire, we therefore asked the eleven data owners (developers) among our participants to answer on a 5-point Likert scale if they found this transparency helpful and useful (Q13 & Q14). We find that data owners show very positive sentiment towards inverse transparency, judging it as beneficial (see Figure 9).

To understand participants' individual perspectives, we followed up with a free text question about their experience with the provided transparency (Q15). In general, participants expressed

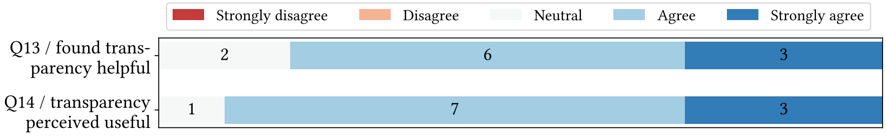


Fig. 9. Results of the questions for data owners (n = 11) about how helpful and useful they perceived inverse transparency to be. They were asked to express their agreement on a five-point Likert scale (X-axis), from 1 for "strongly disagree" to 5 for "strongly agree."

positive sentiment, explaining that the transparency "makes me feel safe" (O1) and that "it is a great approach" (O4) or "a good tool" (O10). "I personally like to use it" (O11), wrote one, with another expressing: "I experienced it well" (O5). Notably though, one participant had a very different experience. They write: "This transparancy [sic!] makes me even more aware of the monitoring, so im [sic!] not able to forget it." (O9). Thereby, they touch upon an important point. Many data usages may be benign, but being confronted with them may still have an influence. For most of our participants, the additional transparency induced a feeling of safety, but some may find that it actually creates a sense of worry and a feeling of being monitored. We reflect this thought in Section 6.

5.5 RQ B.3: Do participants consider inverse transparency to be capable to influence data consumers' data usage behavior?

Receiving transparency over how data are used is judged as helpful and useful by data owners. Yet, the question arises if this transparency influences data consumers, as it might elicit a feeling of "being watched." Potential chilling effects could be desired (by preventing data misusage), or problematic (if legitimate usage of data is hindered). Therefore, we asked participants about the potential influence of inverse transparency on data consumers' data usage behavior.

First, we asked the data owners (n = 11) if they consider inverse transparency a meaningful deterrent for potential data misusage (Q16). We find just one participant disagreeing with the statement, with all others agreeing (4/11) or agreeing strongly (6/11). In a free text answer, the critical respondent reveals that they consider the "transparency [to be] high but not high enough to be bad", judging the risk of misusage in our concrete setting to be low in general. They add: "I believe if usage of data is within good intention. It does not effect user behavior." This suggests that they might have misunderstood the question to be about our concrete scenario, not the theoretical possibility of misusage.

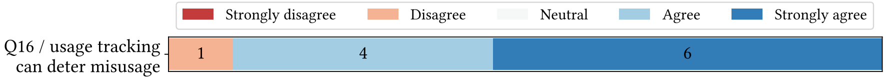


Fig. 10. Results of the question to data owners if they think that inverse transparency can deter misusage. Data owners (n = 11) were asked to express their agreement on a five-point Likert scale (X-axis), from 1 for "strongly disagree" to 5 for "strongly agree."

Second, we asked the data consumers (n = 4) to elaborate if the monitoring of their data usage behavior influenced their actions (Q17). Two consumers were sure, answering "I believe so" (C3) and “Definitely” (C2). Both explained that they think the monitoring reduced the number of their accesses. “Without having to worry about potentially having to explain the frequency of my accesses to the data owners, I would likely have conducted the analyses more often” (C3). The third consumer responded that it only influenced them “barely, but I can’t be 100% sure” (C1), with the final participant being sure that the answer would be no: “I would have acted the same” (C4). All of our participants could not know, but it is notable that opinions diverge so strongly. With two participants clearly worrying about the impression of their data usage on data owners, while another did not even consider it, other confounding factors may have been responsible. For example, the team culture with regards to data-driven management may change the acceptance of the team lead utilizing analyses. We conclude that inverse transparency is capable, but not guaranteed, to influence data consumers in their data usage behavior.

5.6 RQ B.4: Is inverse transparency considered valuable in general by participants?

Finally, we want to understand if participants would prefer inverse transparency over the status quo, which for them is the GDPR. While data owners considered the provided transparency helpful and useful in our specific scenario (see RQ B.2), they may still judge the value of inverse transparency differently if it was introduced to their workplace. Therefore, we asked them for their agreement to four statements:

- Inverse transparency improves upon the protection of the GDPR. (Q18)
- I would prefer inverse transparency over just having the right to consent to or reject data usages outright. (Q19)
- If my company offered me the choice, I would like to have access to data usage tracking. (Q20)
- I would feel safer knowing how my data are accessed in detail. (Q21)

In addition, we allowed participants to elaborate if they wanted to explain their responses (Q22).

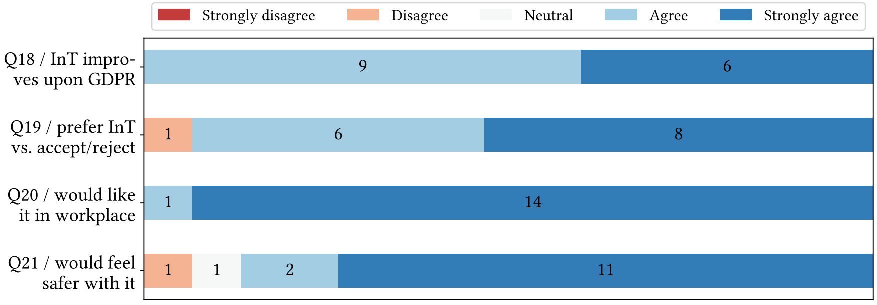


Fig. 11. Results of the questions about inverse transparency generally (abbreviated as “InT” above). All participants (n = 15) were asked to express their agreement on a five-point Likert scale (X-axis), from 1 for “strongly disagree” to 5 for “strongly agree.”

Responses are overall very positive, with only one respondent disagreeing for questions Q19 and Q21, respectively. For question Q21, the critical respondent elaborated that they “would not trust that the access data provided to [them] is actually correct” (O9). Possibly, if an integrity guarantee was given, their sense of safety could be improved. Especially striking is the result for question Q20, with all participants agreeing and almost all (14/15) even expressing strong agreement. This

suggests that inverse transparency is considered especially valuable by participants when imagining a workplace setting. This sentiment is best summarized by the following quotes: “I think that if implemented correctly, Inverse Transparency could be a game changer for privacy” (C3), with another respondent adding: “being able to track how my data is used is beneficial to myself, to transparancy [sic!] within the team and ultimately to the relationship between data owners and data consumers” (O6).

5.7 Conclusions

Our study of the user perspective shows that inverse transparency is judged positively and as an improvement over the status quo. All participants agreed that it would be a valuable addition to the workplace. This suggests inverse transparency is capable to empower employees. We furthermore find promising results regarding our goal of creating accountability and deterring misusage of data. Data owners believed that the inverse transparency could influence data consumers’ usage behavior, an impression that some data consumers confirmed for their case. Finally, considering the purpose dimension of trust (see Section 3.4), participants judge inverse transparency as beneficial for themselves and their teams. The introduction of inverse transparency is considered valuable for employees, which is reflected in that all participants would like it in the workplace. This shows that the purpose of inverse transparency is perceived as benevolent and supportive of employees.

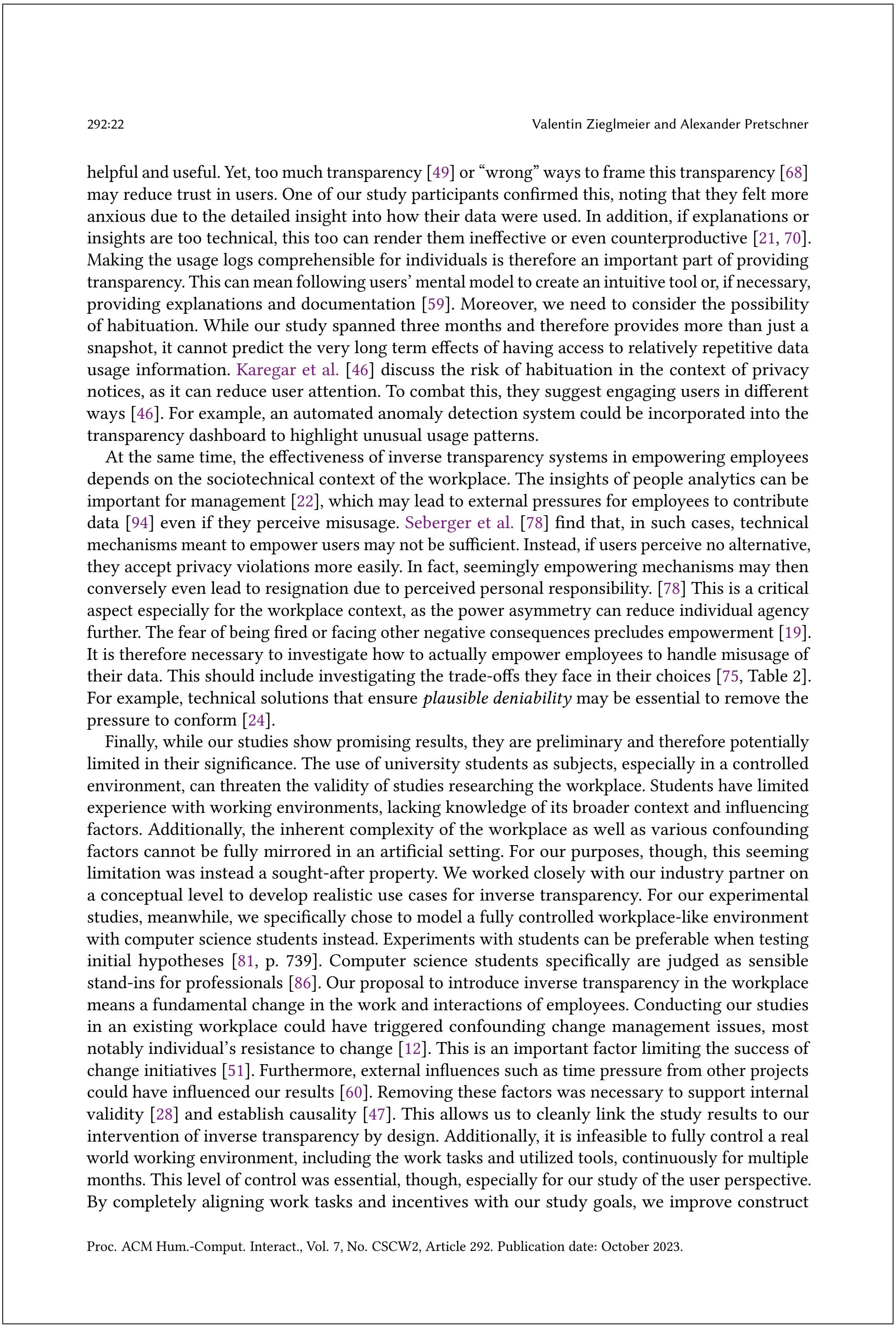
We conclude that inverse transparency is capable to meet our goals of empowering employees and creating accountability for data usages. It is considered an improvement of the status quo, with the perceived benevolent purpose and felt benefits supporting trust in the concept.

6 DISCUSSION

At first glance, expecting a rethink of people analytics design may seem ambitious at best. Yet, we have seen with the introduction of the GDPR how quickly software firms can adapt and update their tools. We think the key driver for this speed were their customers—the companies buying these tools for their use. To ensure their GDPR compliance, they expect their suppliers to build their software accordingly and integrate the necessary tools for, e.g., anonymization or deletion of user data. Now, companies are faced with potent privacy legislation on the one hand, and privacy concerns of their employees [see 85] on the other hand. Ignoring these concerns can lead to anything from reactance [26] to dissatisfaction with the employer or psychological distress [10]. This means that it is in companies’ own interest to address their employees’ concerns. Furthermore, where workers’ councils are active, they may have a say in if analysis tools are bought, using their power to prevent invasive uses of employee data. Inverse transparency could offer a solution for these scenarios. Not all data analyses are problematic, and some might even be beneficial for data owners. Tools built with inverse transparency by design can unlock this potential by empowering data owners with data sovereignty and holding data consumers accountable in case of data misusage.

The shift to develop with inverse transparency by design may have a more pronounced influence on software design than it first appears. We have deliberated the potential changes that can be envisioned when considering *ambient usage* of data. Developers may need to change fundamental interaction paradigms as they include data owners as stakeholders in their design. On the other hand, as we could see in our preliminary study of the developer perspective, inverse transparency may serve as an enabler for potentially privacy-invasive but useful features. Knowing that data owners will be able to supervise usage of their data may reduce the need to be overly cautious upfront, enabling the ethical development of innovative features.

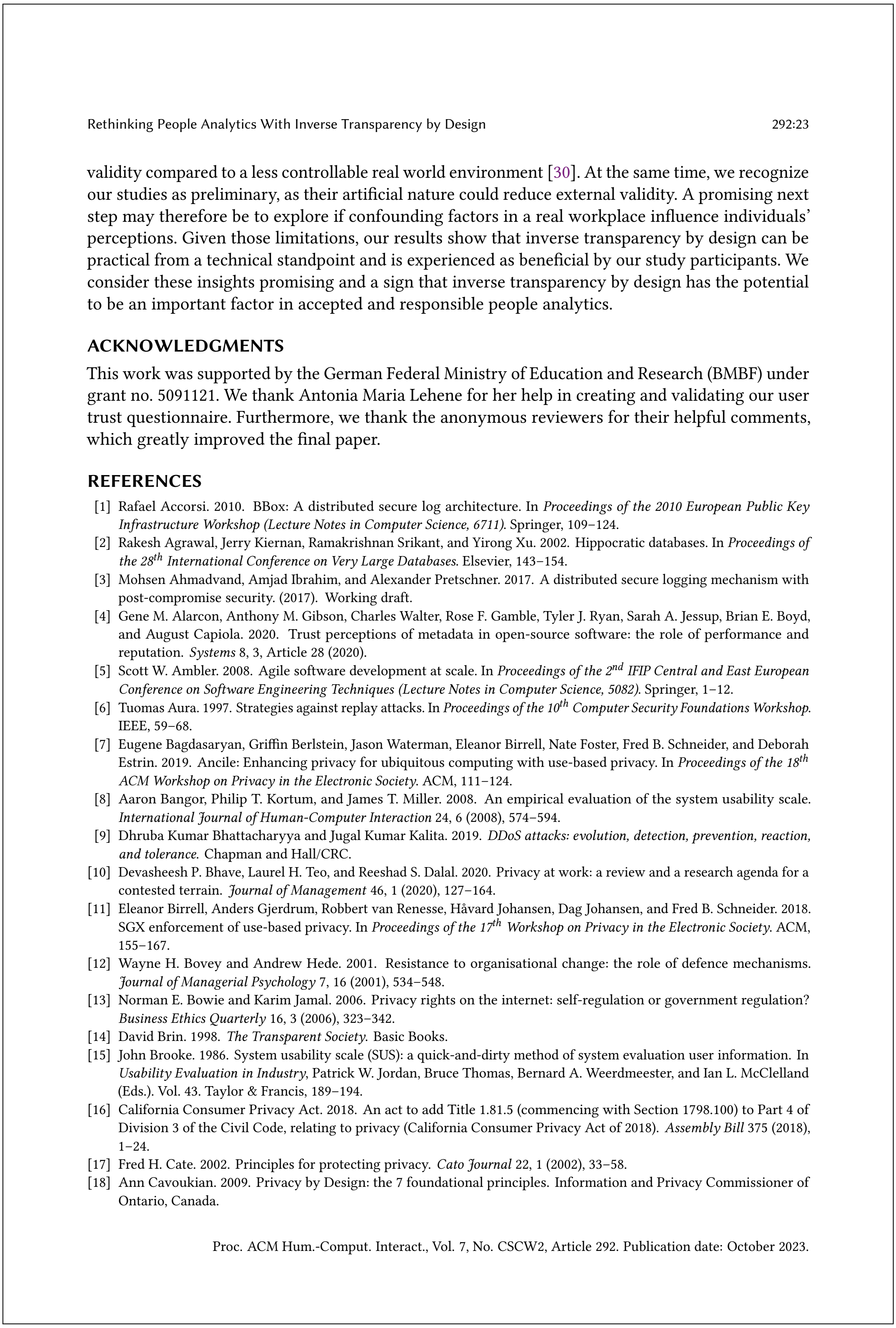
Considering the users, providing inverse transparency is a valuable step but not sufficient on its own. On the one hand, participants in our preliminary study of the user perspective showed a clear interest in inverse transparency. They felt empowered and considered the provided transparency



helpful and useful. Yet, too much transparency [49] or "wrong" ways to frame this transparency [68] may reduce trust in users. One of our study participants confirmed this, noting that they felt more anxious due to the detailed insight into how their data were used. In addition, if explanations or insights are too technical, this too can render them ineffective or even counterproductive [21, 70]. Making the usage logs comprehensible for individuals is therefore an important part of providing transparency. This can mean following users' mental model to create an intuitive tool or, if necessary, providing explanations and documentation [59]. Moreover, we need to consider the possibility of habituation. While our study spanned three months and therefore provides more than just a snapshot, it cannot predict the very long term effects of having access to relatively repetitive data usage information. Karegar et al. [46] discuss the risk of habituation in the context of privacy notices, as it can reduce user attention. To combat this, they suggest engaging users in different ways [46]. For example, an automated anomaly detection system could be incorporated into the transparency dashboard to highlight unusual usage patterns.

At the same time, the effectiveness of inverse transparency systems in empowering employees depends on the sociotechnical context of the workplace. The insights of people analytics can be important for management [22], which may lead to external pressures for employees to contribute data [94] even if they perceive misusage. Seberger et al. [78] find that, in such cases, technical mechanisms meant to empower users may not be sufficient. Instead, if users perceive no alternative, they accept privacy violations more easily. In fact, seemingly empowering mechanisms may then conversely even lead to resignation due to perceived personal responsibility. [78] This is a critical aspect especially for the workplace context, as the power asymmetry can reduce individual agency further. The fear of being fired or facing other negative consequences precludes empowerment [19]. It is therefore necessary to investigate how to actually empower employees to handle misusage of their data. This should include investigating the trade-offs they face in their choices [75, Table 2]. For example, technical solutions that ensure *plausible deniability* may be essential to remove the pressure to conform [24].

Finally, while our studies show promising results, they are preliminary and therefore potentially limited in their significance. The use of university students as subjects, especially in a controlled environment, can threaten the validity of studies researching the workplace. Students have limited experience with working environments, lacking knowledge of its broader context and influencing factors. Additionally, the inherent complexity of the workplace as well as various confounding factors cannot be fully mirrored in an artificial setting. For our purposes, though, this seeming limitation was instead a sought-after property. We worked closely with our industry partner on a conceptual level to develop realistic use cases for inverse transparency. For our experimental studies, meanwhile, we specifically chose to model a fully controlled workplace-like environment with computer science students instead. Experiments with students can be preferable when testing initial hypotheses [81, p. 739]. Computer science students specifically are judged as sensible stand-ins for professionals [86]. Our proposal to introduce inverse transparency in the workplace means a fundamental change in the work and interactions of employees. Conducting our studies in an existing workplace could have triggered confounding change management issues, most notably individual's resistance to change [12]. This is an important factor limiting the success of change initiatives [51]. Furthermore, external influences such as time pressure from other projects could have influenced our results [60]. Removing these factors was necessary to support internal validity [28] and establish causality [47]. This allows us to cleanly link the study results to our intervention of inverse transparency by design. Additionally, it is infeasible to fully control a real world working environment, including the work tasks and utilized tools, continuously for multiple months. This level of control was essential, though, especially for our study of the user perspective. By completely aligning work tasks and incentives with our study goals, we improve construct



validity compared to a less controllable real world environment [30]. At the same time, we recognize our studies as preliminary, as their artificial nature could reduce external validity. A promising next step may therefore be to explore if confounding factors in a real workplace influence individuals' perceptions. Given those limitations, our results show that inverse transparency by design can be practical from a technical standpoint and is experienced as beneficial by our study participants. We consider these insights promising and a sign that inverse transparency by design has the potential to be an important factor in accepted and responsible people analytics.

ACKNOWLEDGMENTS

This work was supported by the German Federal Ministry of Education and Research (BMBF) under grant no. 5091121. We thank Antonia Maria Lehene for her help in creating and validating our user trust questionnaire. Furthermore, we thank the anonymous reviewers for their helpful comments, which greatly improved the final paper.

REFERENCES

[1] Rafael Accorsi. 2010. BBox: A distributed secure log architecture. In *Proceedings of the 2010 European Public Key Infrastructure Workshop (Lecture Notes in Computer Science, 6711)*. Springer, 109–124.

[2] Rakesh Agrawal, Jerry Kiernan, Ramakrishnan Srikant, and Yirong Xu. 2002. Hippocratic databases. In *Proceedings of the 28th International Conference on Very Large Databases*. Elsevier, 143–154.

[3] Mohsen Ahmadvand, Amjad Ibrahim, and Alexander Pretschner. 2017. A distributed secure logging mechanism with post-compromise security. (2017). Working draft.

[4] Gene M. Alarcon, Anthony M. Gibson, Charles Walter, Rose F. Gamble, Tyler J. Ryan, Sarah A. Jessup, Brian E. Boyd, and August Capiola. 2020. Trust perceptions of metadata in open-source software: the role of performance and reputation. *Systems* 8, 3, Article 28 (2020).

[5] Scott W. Ambler. 2008. Agile software development at scale. In *Proceedings of the 2nd IFIP Central and East European Conference on Software Engineering Techniques (Lecture Notes in Computer Science, 5082)*. Springer, 1–12.

[6] Tuomas Aura. 1997. Strategies against replay attacks. In *Proceedings of the 10th Computer Security Foundations Workshop*. IEEE, 59–68.

[7] Eugene Bagdasaryan, Griffin Berlstein, Jason Waterman, Eleanor Birrell, Nate Foster, Fred B. Schneider, and Deborah Estrin. 2019. Ancile: Enhancing privacy for ubiquitous computing with use-based privacy. In *Proceedings of the 18th ACM Workshop on Privacy in the Electronic Society*. ACM, 111–124.

[8] Aaron Bangor, Philip T. Kortum, and James T. Miller. 2008. An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction* 24, 6 (2008), 574–594.

[9] Dhruba Kumar Bhattacharyya and Jugal Kumar Kalita. 2019. *DDoS attacks: evolution, detection, prevention, reaction, and tolerance*. Chapman and Hall/CRC.

[10] 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 (2020), 127–164.

[11] Eleanor Birrell, Anders Gjerdrum, Robbert van Renesse, Håvard Johansen, Dag Johansen, and Fred B. Schneider. 2018. SGX enforcement of use-based privacy. In *Proceedings of the 17th Workshop on Privacy in the Electronic Society*. ACM, 155–167.

[12] Wayne H. Bovey and Andrew Hede. 2001. Resistance to organisational change: the role of defence mechanisms. *Journal of Managerial Psychology* 7, 16 (2001), 534–548.

[13] Norman E. Bowie and Karim Jamal. 2006. Privacy rights on the internet: self-regulation or government regulation? *Business Ethics Quarterly* 16, 3 (2006), 323–342.

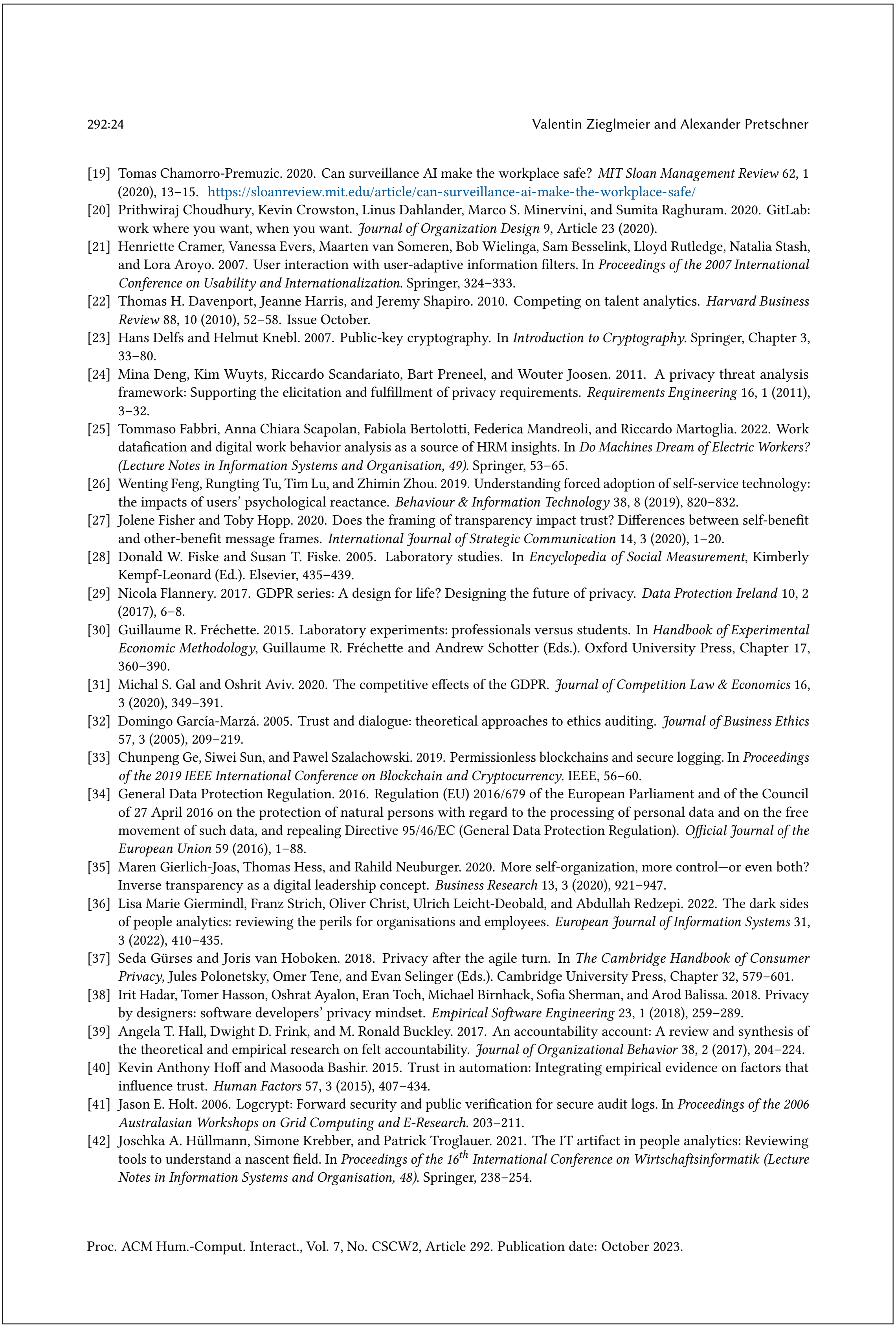
[14] David Brin. 1998. *The Transparent Society*. Basic Books.

[15] John Brooke. 1986. System usability scale (SUS): a quick-and-dirty method of system evaluation user information. In *Usability Evaluation in Industry*, Patrick W. Jordan, Bruce Thomas, Bernard A. Weerdmeester, and Ian L. McClelland (Eds.). Vol. 43. Taylor & Francis, 189–194.

[16] California Consumer Privacy Act. 2018. An act to add Title 1.81.5 (commencing with Section 1798.100) to Part 4 of Division 3 of the Civil Code, relating to privacy (California Consumer Privacy Act of 2018). *Assembly Bill* 375 (2018), 1–24.

[17] Fred H. Cate. 2002. Principles for protecting privacy. *Cato Journal* 22, 1 (2002), 33–58.

[18] Ann Cavoukian. 2009. Privacy by Design: the 7 foundational principles. Information and Privacy Commissioner of Ontario, Canada.



[19] Tomas Chamorro-Premuzic. 2020. Can surveillance AI make the workplace safe? *MIT Sloan Management Review* 62, 1 (2020), 13–15. https://sloanreview.mit.edu/article/can-surveillance-ai-make-the-workplace-safe/

[20] Prithwiraj Choudhury, Kevin Crowston, Linus Dahlander, Marco S. Minervini, and Sumita Raghuram. 2020. GitLab: work where you want, when you want. *Journal of Organization Design* 9, Article 23 (2020).

[21] Henriette Cramer, Vanessa Evers, Maarten van Someren, Bob Wielinga, Sam Besselink, Lloyd Rutledge, Natalia Stash, and Lora Aroyo. 2007. User interaction with user-adaptive information filters. In *Proceedings of the 2007 International Conference on Usability and Internationalization*. Springer, 324–333.

[22] Thomas H. Davenport, Jeanne Harris, and Jeremy Shapiro. 2010. Competing on talent analytics. *Harvard Business Review* 88, 10 (2010), 52–58. Issue October.

[23] Hans Delfs and Helmut Knebl. 2007. Public-key cryptography. In *Introduction to Cryptography*. Springer, Chapter 3, 33–80.

[24] Mina Deng, Kim Wuyts, Riccardo Scandariato, Bart Preneel, and Wouter Joosen. 2011. A privacy threat analysis framework: Supporting the elicitation and fulfillment of privacy requirements. *Requirements Engineering* 16, 1 (2011), 3–32.

[25] Tommaso Fabbri, Anna Chiara Scapolan, Fabiola Bertolotti, Federica Mandreoli, and Riccardo Martoglia. 2022. Work datafication and digital work behavior analysis as a source of HRM insights. In *Do Machines Dream of Electric Workers? (Lecture Notes in Information Systems and Organisation, 49)*. Springer, 53–65.

[26] Wenting Feng, Rungting Tu, Tim Lu, and Zhimin Zhou. 2019. Understanding forced adoption of self-service technology: the impacts of users' psychological reactance. *Behaviour & Information Technology* 38, 8 (2019), 820–832.

[27] Jolene Fisher and Toby Hopp. 2020. Does the framing of transparency impact trust? Differences between self-benefit and other-benefit message frames. *International Journal of Strategic Communication* 14, 3 (2020), 1–20.

[28] Donald W. Fiske and Susan T. Fiske. 2005. Laboratory studies. In *Encyclopedia of Social Measurement*, Kimberly Kempf-Leonard (Ed.). Elsevier, 435–439.

[29] Nicola Flannery. 2017. GDPR series: A design for life? Designing the future of privacy. *Data Protection Ireland* 10, 2 (2017), 6–8.

[30] Guillaume R. Fréchette. 2015. Laboratory experiments: professionals versus students. In *Handbook of Experimental Economic Methodology*, Guillaume R. Fréchette and Andrew Schotter (Eds.). Oxford University Press, Chapter 17, 360–390.

[31] Michal S. Gal and Oshrit Aviv. 2020. The competitive effects of the GDPR. *Journal of Competition Law & Economics* 16, 3 (2020), 349–391.

[32] Domingo García-Marzá. 2005. Trust and dialogue: theoretical approaches to ethics auditing. *Journal of Business Ethics* 57, 3 (2005), 209–219.

[33] Chunpeng Ge, Siwei Sun, and Pawel Szalachowski. 2019. Permissionless blockchains and secure logging. In *Proceedings of the 2019 IEEE International Conference on Blockchain and Cryptocurrency*. IEEE, 56–60.

[34] General Data Protection Regulation. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union* 59 (2016), 1–88.

[35] Maren Gierlich-Joas, Thomas Hess, and Rahild Neuburger. 2020. More self-organization, more control—or even both? Inverse transparency as a digital leadership concept. *Business Research* 13, 3 (2020), 921–947.

[36] Lisa Marie Giermindl, Franz Strich, Oliver Christ, Ulrich Leicht-Deobald, and Abdullah Redzepi. 2022. The dark sides of people analytics: reviewing the perils for organisations and employees. *European Journal of Information Systems* 31, 3 (2022), 410–435.

[37] Seda Gürses and Joris van Hoboken. 2018. Privacy after the agile turn. In *The Cambridge Handbook of Consumer Privacy*, Jules Polonetsky, Omer Tene, and Evan Selinger (Eds.). Cambridge University Press, Chapter 32, 579–601.

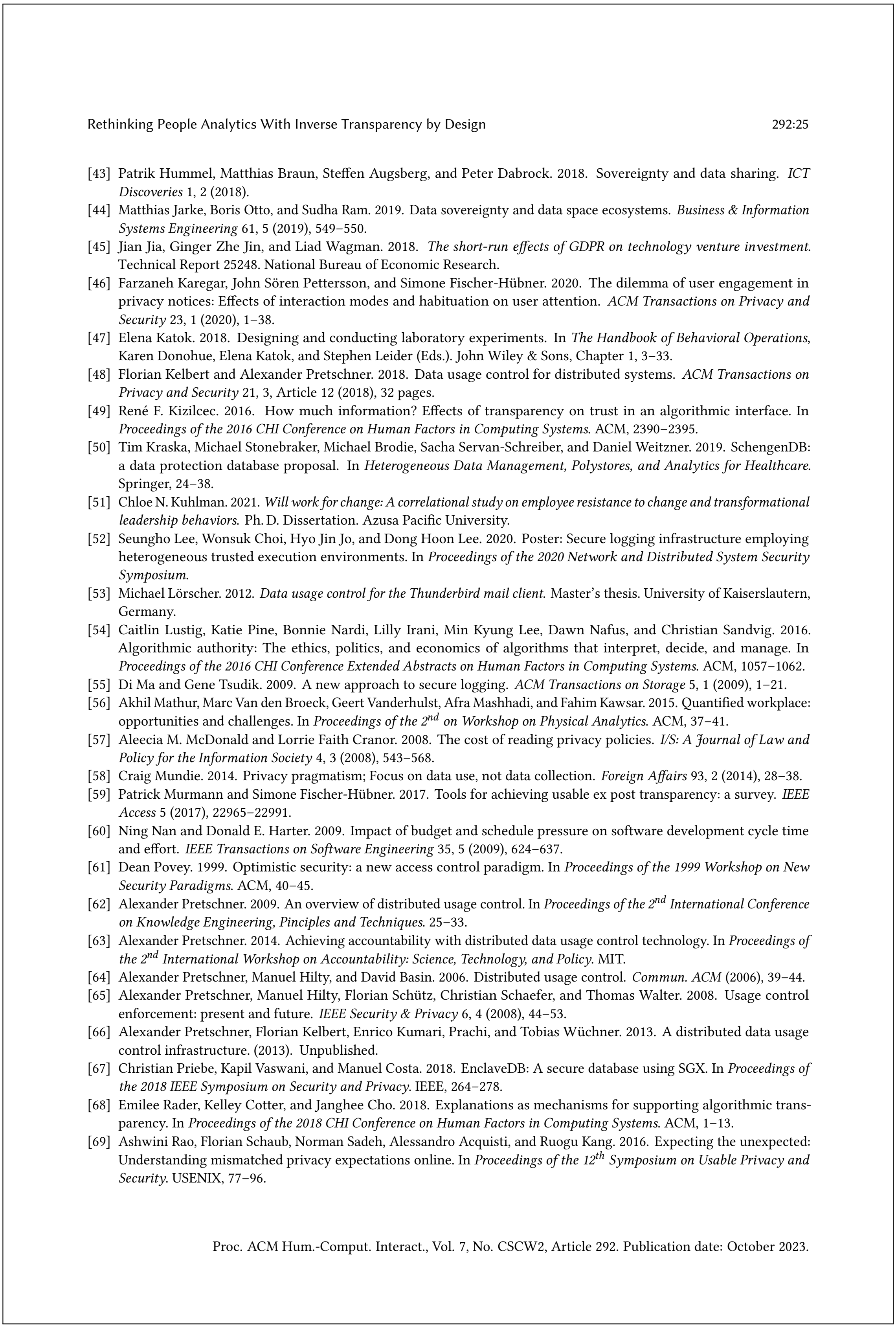
[38] Irit Hadar, Tomer Hasson, Oshrat Ayalon, Eran Toch, Michael Birnhack, Sofia Sherman, and Arod Balissa. 2018. Privacy by designers: software developers' privacy mindset. *Empirical Software Engineering* 23, 1 (2018), 259–289.

[39] Angela T. Hall, Dwight D. Frink, and M. Ronald Buckley. 2017. An accountability account: A review and synthesis of the theoretical and empirical research on felt accountability. *Journal of Organizational Behavior* 38, 2 (2017), 204–224.

[40] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors* 57, 3 (2015), 407–434.

[41] Jason E. Holt. 2006. Logcrypt: Forward security and public verification for secure audit logs. In *Proceedings of the 2006 Australasian Workshops on Grid Computing and E-Research*. 203–211.

[42] Joschka A. Hüllmann, Simone Krebber, and Patrick Troglauer. 2021. The IT artifact in people analytics: Reviewing tools to understand a nascent field. In *Proceedings of the 16th International Conference on Wirtschaftsinformatik (Lecture Notes in Information Systems and Organisation, 48)*. Springer, 238–254.



[43] Patrik Hummel, Matthias Braun, Steffen Augsberg, and Peter Dabrock. 2018. Sovereignty and data sharing. *ICT Discoveries* 1, 2 (2018).

[44] Matthias Jarke, Boris Otto, and Sudha Ram. 2019. Data sovereignty and data space ecosystems. *Business & Information Systems Engineering* 61, 5 (2019), 549–550.

[45] Jian Jia, Ginger Zhe Jin, and Liad Wagman. 2018. *The short-run effects of GDPR on technology venture investment.* Technical Report 25248. National Bureau of Economic Research.

[46] Farzaneh Karegar, John Sören Pettersson, and Simone Fischer-Hübner. 2020. The dilemma of user engagement in privacy notices: Effects of interaction modes and habituation on user attention. *ACM Transactions on Privacy and Security* 23, 1 (2020), 1–38.

[47] Elena Katok. 2018. Designing and conducting laboratory experiments. In *The Handbook of Behavioral Operations*, Karen Donohue, Elena Katok, and Stephen Leider (Eds.). John Wiley & Sons, Chapter 1, 3–33.

[48] Florian Kelbert and Alexander Pretschner. 2018. Data usage control for distributed systems. *ACM Transactions on Privacy and Security* 21, 3, Article 12 (2018), 32 pages.

[49] René F. Kizilcec. 2016. How much information? Effects of transparency on trust in an algorithmic interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* ACM, 2390–2395.

[50] Tim Kraska, Michael Stonebraker, Michael Brodie, Sacha Servan-Schreiber, and Daniel Weitzner. 2019. SchengenDB: a data protection database proposal. In *Heterogeneous Data Management, Polystores, and Analytics for Healthcare.* Springer, 24–38.

[51] Chloe N. Kuhlman. 2021. *Will work for change: A correlational study on employee resistance to change and transformational leadership behaviors.* Ph. D. Dissertation. Azusa Pacific University.

[52] Seungho Lee, Wonsuk Choi, Hyo Jin Jo, and Dong Hoon Lee. 2020. Poster: Secure logging infrastructure employing heterogeneous trusted execution environments. In *Proceedings of the 2020 Network and Distributed System Security Symposium.*

[53] Michael Lörscher. 2012. *Data usage control for the Thunderbird mail client.* Master's thesis. University of Kaiserslautern, Germany.

[54] Caitlin Lustig, Katie Pine, Bonnie Nardi, Lilly Irani, Min Kyung Lee, Dawn Nafus, and Christian Sandvig. 2016. Algorithmic authority: The ethics, politics, and economics of algorithms that interpret, decide, and manage. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems.* ACM, 1057–1062.

[55] Di Ma and Gene Tsudik. 2009. A new approach to secure logging. *ACM Transactions on Storage* 5, 1 (2009), 1–21.

[56] Akhil Mathur, Marc Van den Broeck, Geert Vanderhulst, Afra Mashhadi, and Fahim Kawsar. 2015. Quantified workplace: opportunities and challenges. In *Proceedings of the 2nd on Workshop on Physical Analytics.* ACM, 37–41.

[57] Aleecia M. McDonald and Lorrie Faith Cranor. 2008. The cost of reading privacy policies. *I/S: A Journal of Law and Policy for the Information Society* 4, 3 (2008), 543–568.

[58] Craig Mundie. 2014. Privacy pragmatism; Focus on data use, not data collection. *Foreign Affairs* 93, 2 (2014), 28–38.

[59] Patrick Murmann and Simone Fischer-Hübner. 2017. Tools for achieving usable ex post transparency: a survey. *IEEE Access* 5 (2017), 22965–22991.

[60] Ning Nan and Donald E. Harter. 2009. Impact of budget and schedule pressure on software development cycle time and effort. *IEEE Transactions on Software Engineering* 35, 5 (2009), 624–637.

[61] Dean Povey. 1999. Optimistic security: a new access control paradigm. In *Proceedings of the 1999 Workshop on New Security Paradigms.* ACM, 40–45.

[62] Alexander Pretschner. 2009. An overview of distributed usage control. In *Proceedings of the 2nd International Conference on Knowledge Engineering, Pinciples and Techniques.* 25–33.

[63] Alexander Pretschner. 2014. Achieving accountability with distributed data usage control technology. In *Proceedings of the 2nd International Workshop on Accountability: Science, Technology, and Policy.* MIT.

[64] Alexander Pretschner, Manuel Hilty, and David Basin. 2006. Distributed usage control. *Commun. ACM* (2006), 39–44.

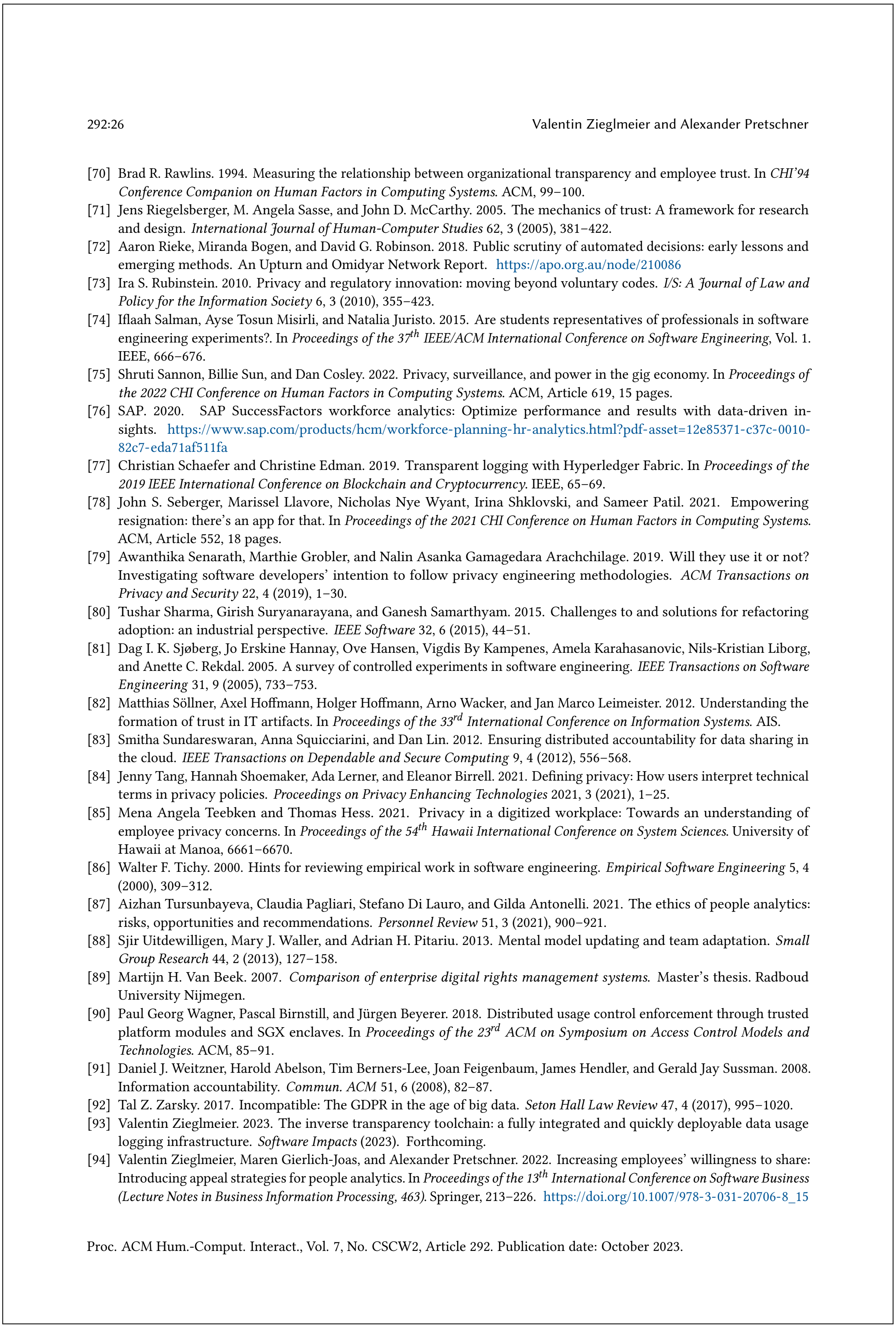
[65] Alexander Pretschner, Manuel Hilty, Florian Schütz, Christian Schaefer, and Thomas Walter. 2008. Usage control enforcement: present and future. *IEEE Security & Privacy* 6, 4 (2008), 44–53.

[66] Alexander Pretschner, Florian Kelbert, Enrico Kumari, Prachi, and Tobias Wüchner. 2013. A distributed data usage control infrastructure. (2013). Unpublished.

[67] Christian Priebe, Kapil Vaswani, and Manuel Costa. 2018. EnclaveDB: A secure database using SGX. In *Proceedings of the 2018 IEEE Symposium on Security and Privacy.* IEEE, 264–278.

[68] Emilee Rader, Kelley Cotter, and Janghee Cho. 2018. Explanations as mechanisms for supporting algorithmic transparency. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* ACM, 1–13.

[69] Ashwini Rao, Florian Schaub, Norman Sadeh, Alessandro Acquisti, and Ruogu Kang. 2016. Expecting the unexpected: Understanding mismatched privacy expectations online. In *Proceedings of the 12th Symposium on Usable Privacy and Security.* USENIX, 77–96.



[70] Brad R. Rawlins. 1994. Measuring the relationship between organizational transparency and employee trust. In *CHI'94 Conference Companion on Human Factors in Computing Systems*. ACM, 99–100.

[71] Jens Riegelsberger, M. Angela Sasse, and John D. McCarthy. 2005. The mechanics of trust: A framework for research and design. *International Journal of Human-Computer Studies* 62, 3 (2005), 381–422.

[72] Aaron Rieke, Miranda Bogen, and David G. Robinson. 2018. Public scrutiny of automated decisions: early lessons and emerging methods. An Upturn and Omidyar Network Report. https://apo.org.au/node/210086

[73] Ira S. Rubinstein. 2010. Privacy and regulatory innovation: moving beyond voluntary codes. *I/S: A Journal of Law and Policy for the Information Society* 6, 3 (2010), 355–423.

[74] Iflaah Salman, Ayse Tosun Misirli, and Natalia Juristo. 2015. Are students representatives of professionals in software engineering experiments?. In *Proceedings of the 37th IEEE/ACM International Conference on Software Engineering*, Vol. 1. IEEE, 666–676.

[75] Shruti Sannon, Billie Sun, and Dan Cosley. 2022. Privacy, surveillance, and power in the gig economy. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. ACM, Article 619, 15 pages.

[76] SAP. 2020. SAP SuccessFactors workforce analytics: Optimize performance and results with data-driven insights. https://www.sap.com/products/hcm/workforce-planning-hr-analytics.html?pdf-asset=12e85371-c37c-0010-82c7-eda71af511fa

[77] Christian Schaefer and Christine Edman. 2019. Transparent logging with Hyperledger Fabric. In *Proceedings of the 2019 IEEE International Conference on Blockchain and Cryptocurrency*. IEEE, 65–69.

[78] John S. Seberger, Marissel Llavore, Nicholas Nye Wyant, Irina Shklovski, and Sameer Patil. 2021. Empowering resignation: there's an app for that. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Article 552, 18 pages.

[79] Awanthika Senarath, Marthie Grobler, and Nalin Asanka Gamagedara Arachchilage. 2019. Will they use it or not? Investigating software developers' intention to follow privacy engineering methodologies. *ACM Transactions on Privacy and Security* 22, 4 (2019), 1–30.

[80] Tushar Sharma, Girish Suryanarayana, and Ganesh Samarthyam. 2015. Challenges to and solutions for refactoring adoption: an industrial perspective. *IEEE Software* 32, 6 (2015), 44–51.

[81] Dag I. K. Sjøberg, Jo Erskine Hannay, Ove Hansen, Vigdis By Kampenes, Amela Karahasanovic, Nils-Kristian Liborg, and Anette C. Rekdal. 2005. A survey of controlled experiments in software engineering. *IEEE Transactions on Software Engineering* 31, 9 (2005), 733–753.

[82] Matthias Söllner, Axel Hoffmann, Holger Hoffmann, Arno Wacker, and Jan Marco Leimeister. 2012. Understanding the formation of trust in IT artifacts. In *Proceedings of the 33rd International Conference on Information Systems*. AIS.

[83] Smitha Sundareswaran, Anna Squicciarini, and Dan Lin. 2012. Ensuring distributed accountability for data sharing in the cloud. *IEEE Transactions on Dependable and Secure Computing* 9, 4 (2012), 556–568.

[84] Jenny Tang, Hannah Shoemaker, Ada Lerner, and Eleanor Birrell. 2021. Defining privacy: How users interpret technical terms in privacy policies. *Proceedings on Privacy Enhancing Technologies* 2021, 3 (2021), 1–25.

[85] Mena Angela Teebken and Thomas Hess. 2021. Privacy in a digitized workplace: Towards an understanding of employee privacy concerns. In *Proceedings of the 54th Hawaii International Conference on System Sciences*. University of Hawaii at Manoa, 6661–6670.

[86] Walter F. Tichy. 2000. Hints for reviewing empirical work in software engineering. *Empirical Software Engineering* 5, 4 (2000), 309–312.

[87] Aizhan Tursunbayeva, Claudia Pagliari, Stefano Di Lauro, and Gilda Antonelli. 2021. The ethics of people analytics: risks, opportunities and recommendations. *Personnel Review* 51, 3 (2021), 900–921.

[88] Sjir Uitdewilligen, Mary J. Waller, and Adrian H. Pitariu. 2013. Mental model updating and team adaptation. *Small Group Research* 44, 2 (2013), 127–158.

[89] Martijn H. Van Beek. 2007. *Comparison of enterprise digital rights management systems*. Master's thesis. Radboud University Nijmegen.

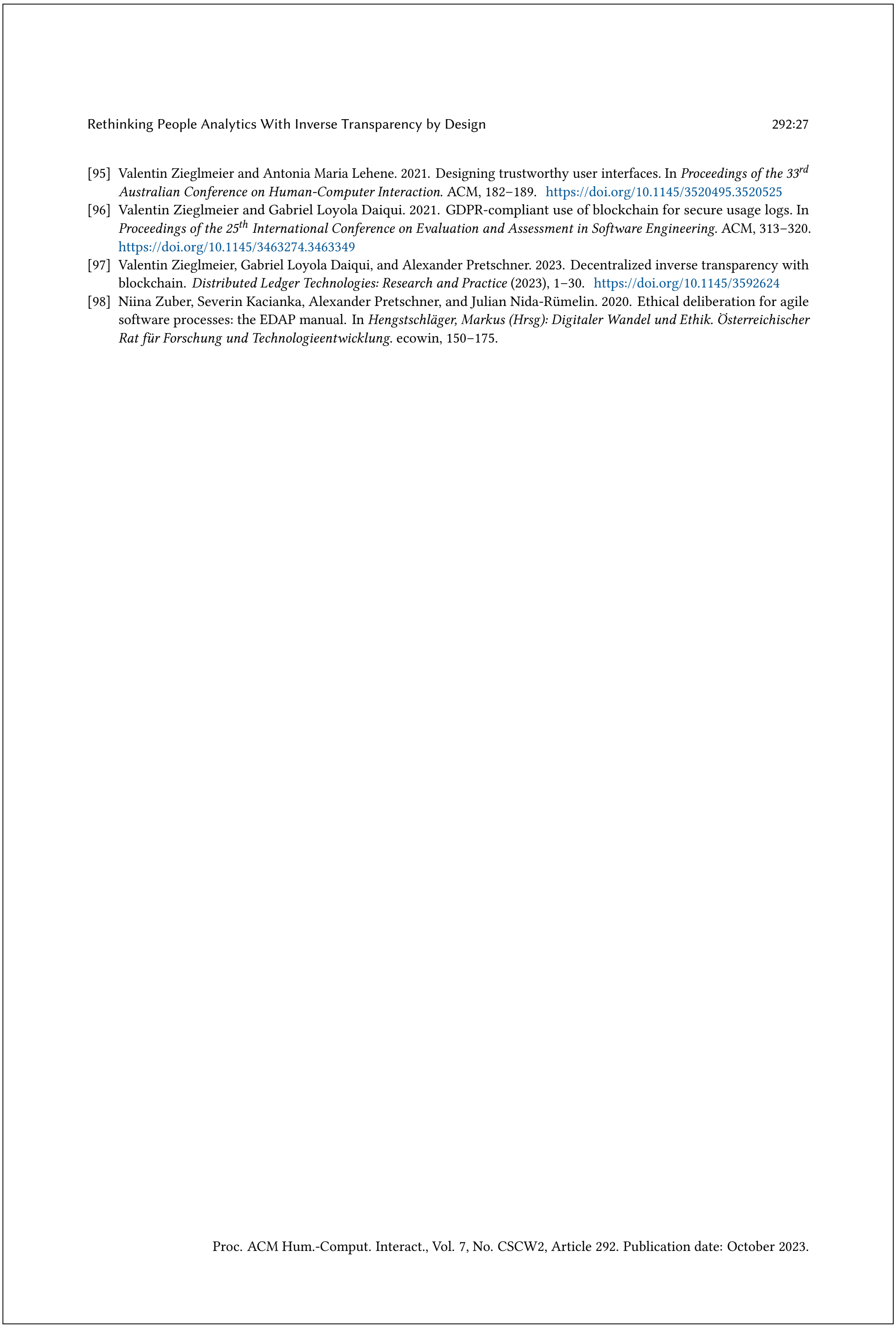
[90] Paul Georg Wagner, Pascal Birnstill, and Jürgen Beyerer. 2018. Distributed usage control enforcement through trusted platform modules and SGX enclaves. In *Proceedings of the 23rd ACM on Symposium on Access Control Models and Technologies*. ACM, 85–91.

[91] Daniel J. Weitzner, Harold Abelson, Tim Berners-Lee, Joan Feigenbaum, James Hendler, and Gerald Jay Sussman. 2008. Information accountability. *Commun. ACM* 51, 6 (2008), 82–87.

[92] Tal Z. Zarsky. 2017. Incompatible: The GDPR in the age of big data. *Seton Hall Law Review* 47, 4 (2017), 995–1020.

[93] Valentin Zieglmeier. 2023. The inverse transparency toolchain: a fully integrated and quickly deployable data usage logging infrastructure. *Software Impacts* (2023). Forthcoming.

[94] Valentin Zieglmeier, Maren Gierlich-Joas, and Alexander Pretschner. 2022. Increasing employees' willingness to share: Introducing appeal strategies for people analytics. In *Proceedings of the 13th International Conference on Software Business (Lecture Notes in Business Information Processing, 463)*. Springer, 213–226. https://doi.org/10.1007/978-3-031-20706-8_15



[95] Valentin Zieglmeier and Antonia Maria Lehene. 2021. Designing trustworthy user interfaces. In *Proceedings of the 33rd Australian Conference on Human-Computer Interaction*. ACM, 182–189. https://doi.org/10.1145/3520495.3520525

[96] Valentin Zieglmeier and Gabriel Loyola Daiqui. 2021. GDPR-compliant use of blockchain for secure usage logs. In *Proceedings of the 25th International Conference on Evaluation and Assessment in Software Engineering*. ACM, 313–320. https://doi.org/10.1145/3463274.3463349

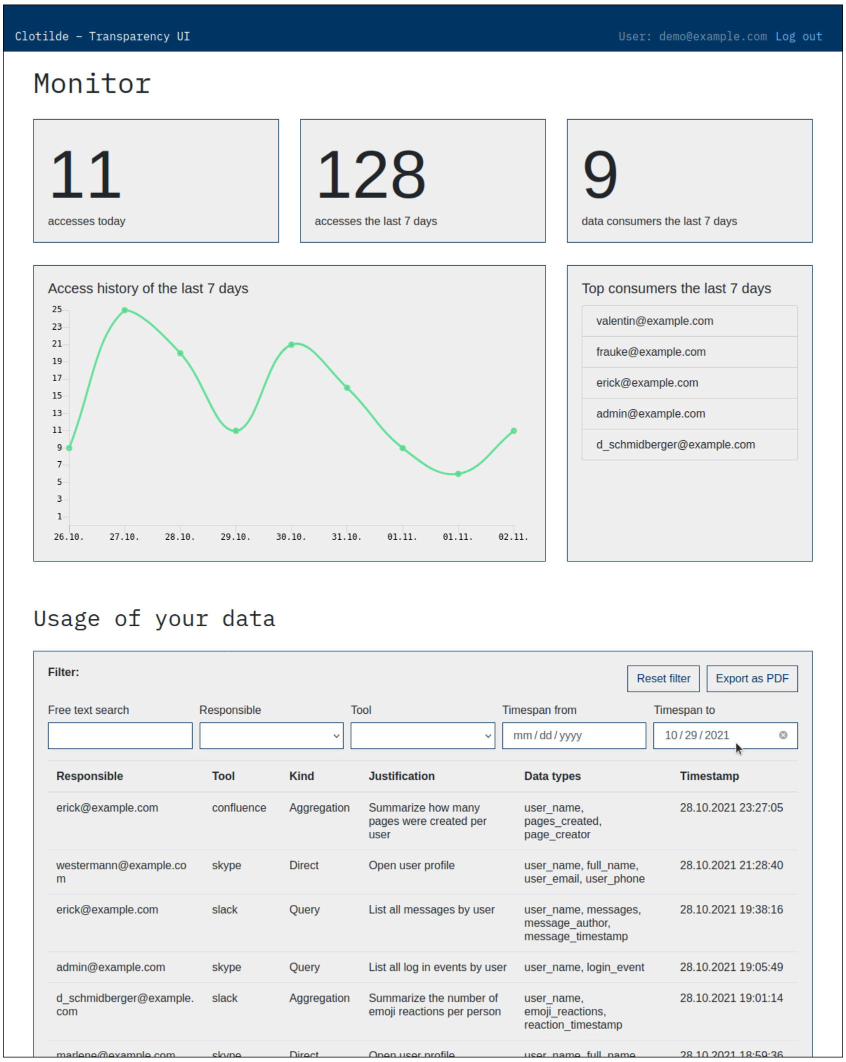
[97] Valentin Zieglmeier, Gabriel Loyola Daiqui, and Alexander Pretschner. 2023. Decentralized inverse transparency with blockchain. *Distributed Ledger Technologies: Research and Practice* (2023), 1–30. https://doi.org/10.1145/3592624

[98] Niina Zuber, Severin Kacianka, Alexander Pretschner, and Julian Nida-Rümelin. 2020. Ethical deliberation for agile software processes: the EDAP manual. In *Hengstschläger, Markus (Hrsg): Digitaler Wandel und Ethik. Österreichischer Rat für Forschung und Technologieentwicklung*. ecowin, 150–175.

A ADDENDUM FOR STUDY B

In the following, find a screenshot of the transparency dashboard (A.1) and a list of the questionnaire questions (A.2).

A.1 The transparency dashboard



Monitor

11
accesses today

128
accesses the last 7 days

9
data consumers the last 7 days

Top consumers the last 7 days

- valentin@example.com
- frauke@example.com
- erick@example.com
- admin@example.com
- d_schmidberger@example.com

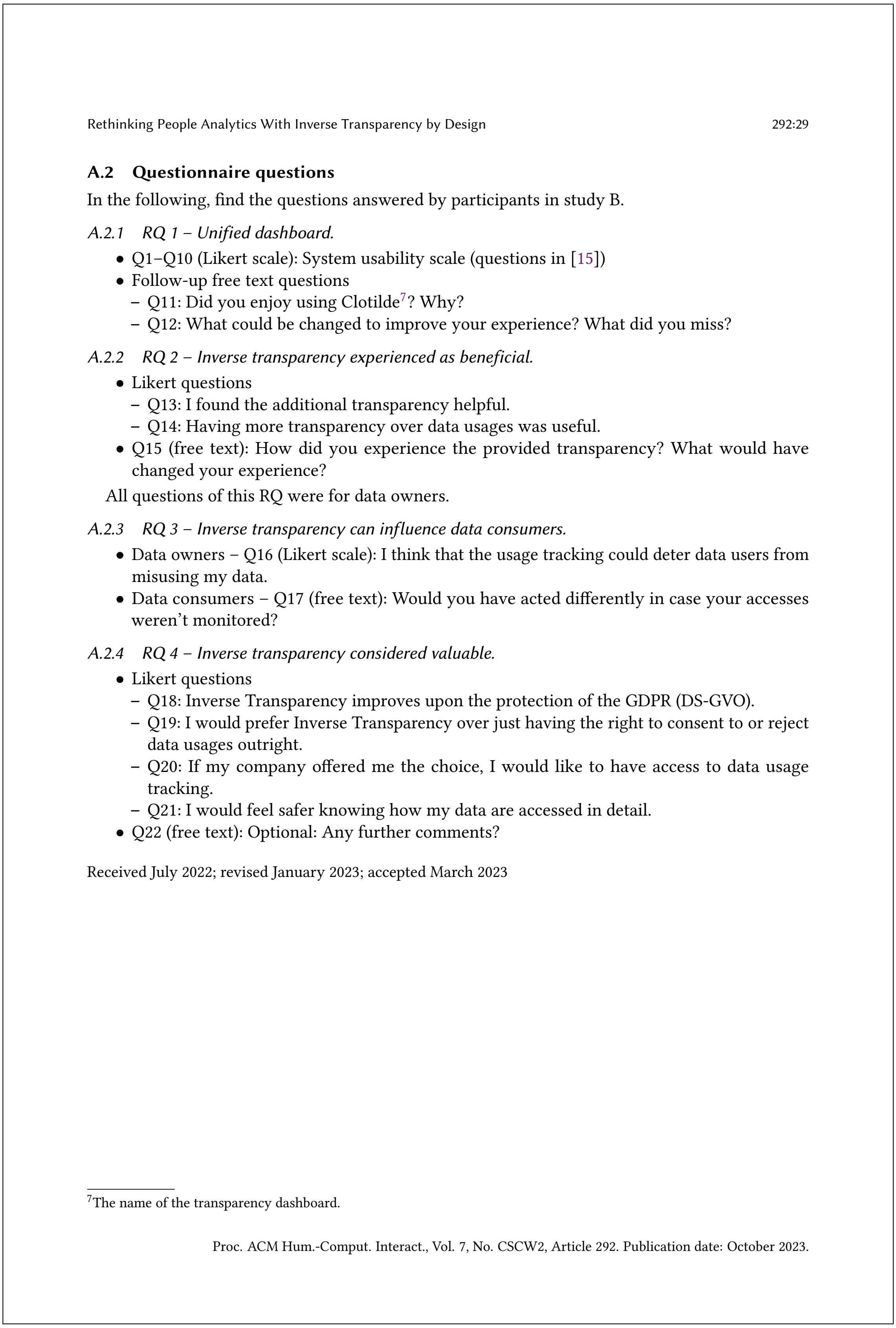
Usage of your data

Filter:

Free text search:
Responsible:
Tool:
Timespan from: mm / dd / yyyy
Timespan to: 10 / 29 / 2021

Responsible	Tool	Kind	Justification	Data types	Timestamp
erick@example.com	confluence	Aggregation	Summarize how many pages were created per user	user_name, pages_created, page_creator	28.10.2021 23:27:05
westermann@example.com	skype	Direct	Open user profile	user_name, full_name, user_email, user_phone	28.10.2021 21:28:40
erick@example.com	slack	Query	List all messages by user	user_name, messages, message_author, message_timestamp	28.10.2021 19:38:16
admin@example.com	skype	Query	List all log in events by user	user_name, login_event	28.10.2021 19:05:49
d_schmidberger@example.com	slack	Aggregation	Summarize the number of emoji reactions per person	user_name, emoji_reactions, reaction_timestamp	28.10.2021 19:01:14
marlene@example.com	skype	Direct	Open user profile	user_name, full_name	28.10.2021 18:59:36

Fig. 12. Screenshot of the transparency dashboard used in study B. The top half summarizes the data usages recorded in the last seven days. The bottom presents a detailed list of the logged data usages for data owners to inspect. The email addresses and usage data shown are artificial and for illustration purposes only.



A.2 Questionnaire questions

In the following, find the questions answered by participants in study B.

A.2.1 RQ 1 – Unified dashboard.

- Q1–Q10 (Likert scale): System usability scale (questions in [15])
- Follow-up free text questions
  - Q11: Did you enjoy using Clotilde[7]? Why?
  - Q12: What could be changed to improve your experience? What did you miss?

A.2.2 RQ 2 – Inverse transparency experienced as beneficial.

- Likert questions
  - Q13: I found the additional transparency helpful.
  - Q14: Having more transparency over data usages was useful.
- Q15 (free text): How did you experience the provided transparency? What would have changed your experience?

All questions of this RQ were for data owners.

A.2.3 RQ 3 – Inverse transparency can influence data consumers.

- Data owners – Q16 (Likert scale): I think that the usage tracking could deter data users from misusing my data.
- Data consumers – Q17 (free text): Would you have acted differently in case your accesses weren't monitored?

A.2.4 RQ 4 – Inverse transparency considered valuable.

- Likert questions
  - Q18: Inverse Transparency improves upon the protection of the GDPR (DS-GVO).
  - Q19: I would prefer Inverse Transparency over just having the right to consent to or reject data usages outright.
  - Q20: If my company offered me the choice, I would like to have access to data usage tracking.
  - Q21: I would feel safer knowing how my data are accessed in detail.
- Q22 (free text): Optional: Any further comments?

Received July 2022; revised January 2023; accepted March 2023

[7] The name of the transparency dashboard.

5. Decentralized Inverse Transparency With Blockchain

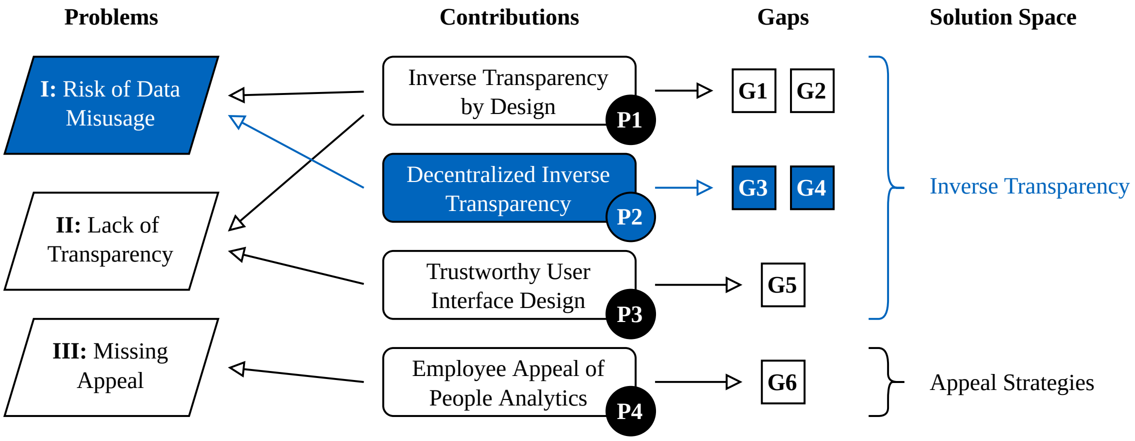


Figure 5.1.: Relationship of our second contribution to the big picture.

Summary. The following summary is partially adapted from our paper [see 2].

- *Problem:* Employees face an elevated *risk of data misusage* (problem **I**) due to the increasing value of their collected data in the workplace. The concept of *inverse transparency* (see Section 2.5) aims to give them oversight to promote data consumers' accountability. Sensitive data require additional protection, though. We imagine data owners storing these data themselves and only making them available on request. This ensures that they know the data consumer and can hold them personally accountable. Due to the increased risks, we aim for a solution that is not dependent on a TTP, which we refer to as *decentralized*. Instead, the data are sent directly by the data owner to the data consumer to be processed on the consumer's system. This *data exchange* needs to be recorded and stored securely.

- *Solution:* To exchange data in a decentralized context while preserving inverse transparency, we need to establish an agreed-upon and non-repudiable timeline of events. We consider the problem of data *access* transparency as a first step towards full data *usage* transparency. As no TTP can serve as an arbiter, we therefore require (1) a system for peer-to-peer non-repudiable data exchange (see also Section 2.8) and (2) a distributed log that enables data owners to access the transparency data.

- *Gaps:* Protocols for non-repudiable data exchange without a TTP exist, but we lack concrete algorithms for the required identity verification and time-asymmetric encryption (**G3**). For secure usage logs, meanwhile, existing applicable solutions are not GDPR-compliant, which is a requirement in our legal context, or weaken the provided security guarantees to achieve compliance (**G4**).
- *Contribution:* We present KOVACS, a decentralized data exchange and usage logging system for inverse transparency. As our main contributions, we provide algorithms to enable non-repudiable data exchange, generate private pseudonyms with unlinkability and proof of ownership, and preserve GDPR compliance while securely storing the logs. We implement and evaluate KOVACS, providing analyses of security and GDPR compliance, as well as performance and scalability benchmarks. We find that KOVACS can fulfill the required properties. It scales linearly and exhibits reasonable exchange duration and storage size.
- *Limitations:* Optimizing for security means that other properties are necessarily not optimized for. We find that the practicality of KOVACS is especially limited regarding the exchange duration, which takes at least seven seconds for a data consumer but can increase more than tenfold in the worst case. Furthermore, if data are accessed, the respective data owner's system must be online, which limits availability.

Author Contributions. VZ developed the initial research idea. GLD researched and summarized the initial state of the art. VZ built on this work to systematically summarize related works. VZ developed the first theoretical solution concept in discussion with GLD. GLD implemented and benchmarked the system under the close guidance of VZ. VZ conceived and implemented the framing of the work. AP provided feedback on the contribution of the work in relation to related works. Finally, VZ and GLD wrote the manuscript in close discussion with AP.

Copyright Note. © 2023 The authors. Reproduced with permission from ACM. Valentin Zieglmeier, Gabriel Loyola Daiqui, and Alexander Pretschner. "Decentralized Inverse Transparency With Blockchain." *Distributed Ledger Technologies: Research and Practice* 2.3, Article 17 (2023).

On the following pages, the published paper is reprinted in accordance with the ACM author rights to reuse their own work in a dissertation. The version of record can be found with the following DOI: `10.1145/3592624`.

Decentralized Inverse Transparency with Blockchain

VALENTIN ZIEGLMEIER, GABRIEL LOYOLA DAIQUI, and ALEXANDER PRETSCHNER,
Technical University of Munich, Germany

Employee data can be used to facilitate work, but their misusage may pose risks for individuals. *Inverse transparency* therefore aims to track all usages of personal data, allowing individuals to monitor them to ensure accountability for potential misusage. This necessitates a trusted log to establish an agreed-upon and non-repudiable timeline of events. The unique properties of blockchain facilitate this by providing immutability and availability. For power asymmetric environments such as the workplace, permissionless blockchain is especially beneficial as no trusted third party is required. Yet, two issues remain: (1) In a decentralized environment, no arbiter can facilitate and attest to data exchanges. Simple peer-to-peer sharing of data, conversely, lacks the required non-repudiation. (2) With data governed by privacy legislation such as the GDPR, the core advantage of immutability becomes a liability. After a rightful request, an individual's personal data need to be rectified or deleted, which is impossible in an immutable blockchain.

To solve these issues, we present Kovacs, a decentralized data exchange and usage logging system for inverse transparency built on blockchain. Its new-usage protocol ensures non-repudiation, and therefore accountability, for inverse transparency. Its one-time pseudonym generation algorithm guarantees unlinkability and enables proof of ownership, which allows data subjects to exercise their legal rights regarding their personal data. With our implementation, we show the viability of our solution. The decentralized communication impacts performance and scalability, but exchange duration and storage size are still reasonable. More importantly, the provided information security meets high requirements. We conclude that Kovacs realizes decentralized inverse transparency through secure and GDPR-compliant use of permissionless blockchain.

CCS Concepts: • **Computer systems organization → Peer-to-peer architectures;** • **Security and privacy → Distributed systems security**; *Privacy-preserving protocols*; *Cryptography*;

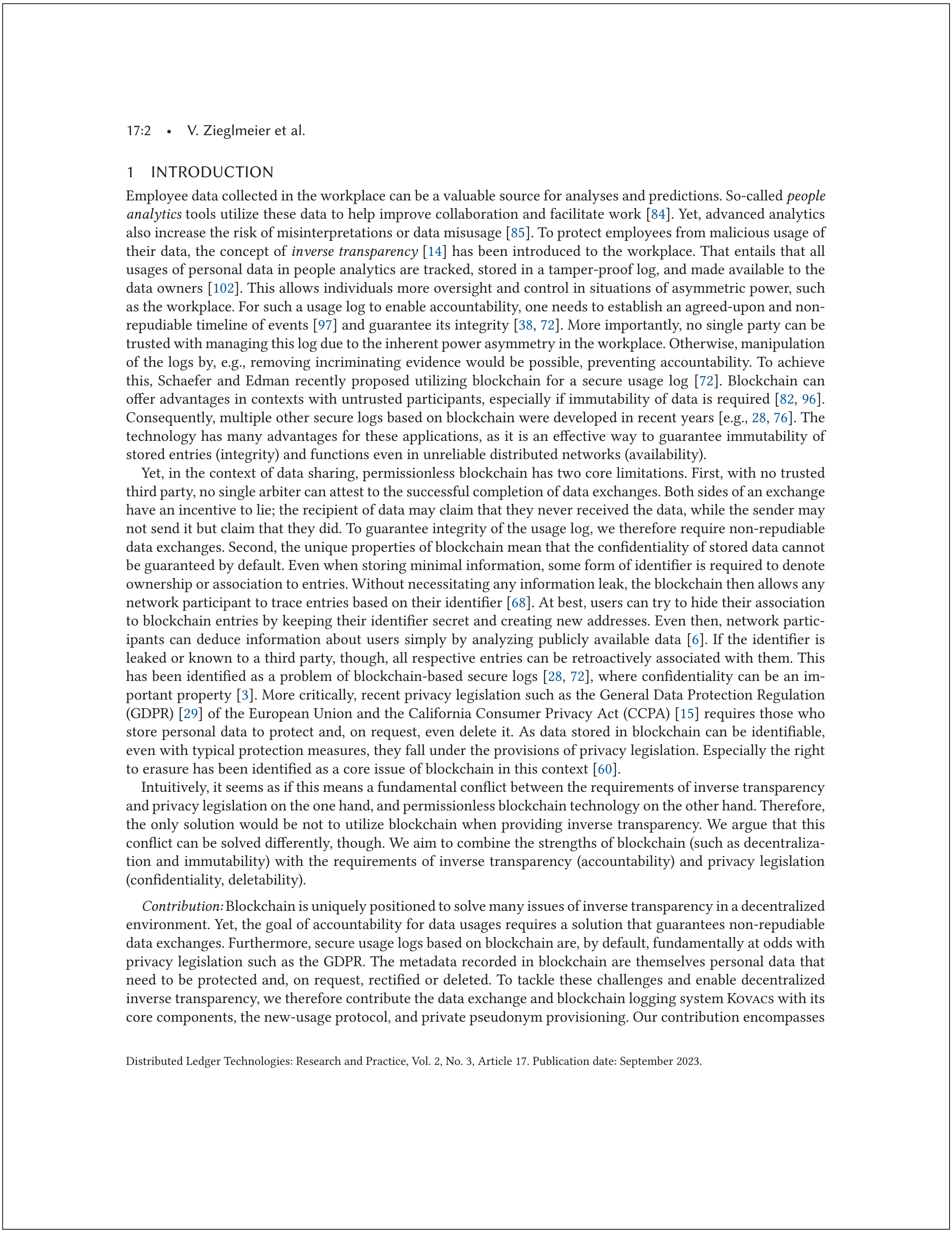
Additional Key Words and Phrases: Decentralization, non-repudiation, accountability, privacy, anonymity

ACM Reference format:
Valentin Zieglmeier, Gabriel Loyola Daiqui, and Alexander Pretschner. 2023. Decentralized Inverse Transparency with Blockchain. *Distrib. Ledger Technol.* 2, 3, Article 17 (September 2023), 28 pages.
https://doi.org/10.1145/3592624

This work was supported by the German Federal Ministry of Education and Research (BMBF) under grant no. 5091121.

Authors' address: V. Zieglmeier, G. Loyola Daiqui, and A. Pretschner, Technical University of Munich, TUM School of Computation, Information and Technology, Chair of Software and Systems Engineering, Boltzmannstr. 3, 85748 Garching, Germany; emails: {valentin.zieglmeier, gabriel.loyola-daiqui, alexander.pretschner}@tum.de.

2769-6472/2023/09-ART17 $15.00
https://doi.org/10.1145/3592624



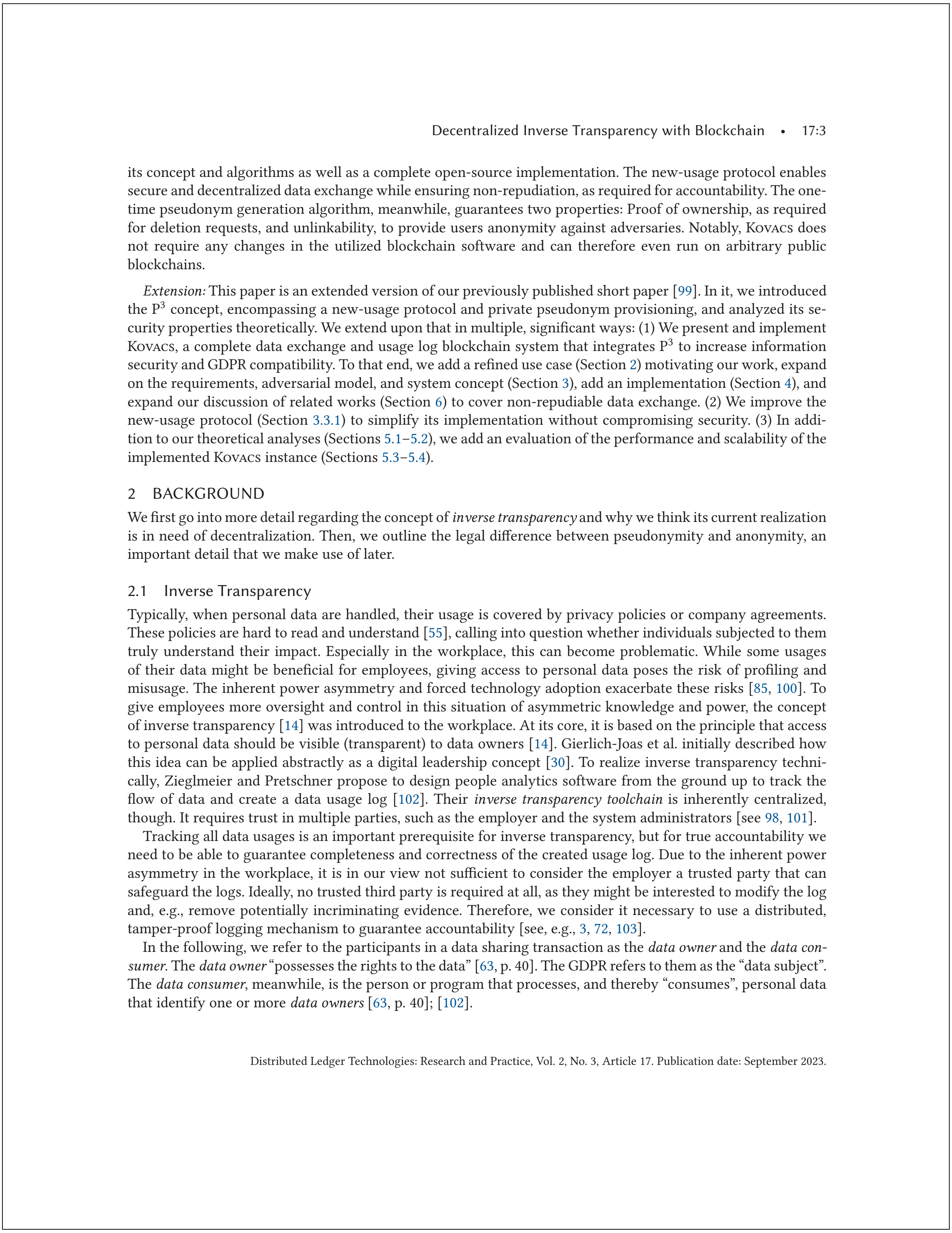
1 INTRODUCTION

Employee data collected in the workplace can be a valuable source for analyses and predictions. So-called *people analytics* tools utilize these data to help improve collaboration and facilitate work [84]. Yet, advanced analytics also increase the risk of misinterpretations or data misusage [85]. To protect employees from malicious usage of their data, the concept of *inverse transparency* [14] has been introduced to the workplace. That entails that all usages of personal data in people analytics are tracked, stored in a tamper-proof log, and made available to the data owners [102]. This allows individuals more oversight and control in situations of asymmetric power, such as the workplace. For such a usage log to enable accountability, one needs to establish an agreed-upon and non-repudiable timeline of events [97] and guarantee its integrity [38, 72]. More importantly, no single party can be trusted with managing this log due to the inherent power asymmetry in the workplace. Otherwise, manipulation of the logs by, e.g., removing incriminating evidence would be possible, preventing accountability. To achieve this, Schaefer and Edman recently proposed utilizing blockchain for a secure usage log [72]. Blockchain can offer advantages in contexts with untrusted participants, especially if immutability of data is required [82, 96]. Consequently, multiple other secure logs based on blockchain were developed in recent years [e.g., 28, 76]. The technology has many advantages for these applications, as it is an effective way to guarantee immutability of stored entries (integrity) and functions even in unreliable distributed networks (availability).

Yet, in the context of data sharing, permissionless blockchain has two core limitations. First, with no trusted third party, no single arbiter can attest to the successful completion of data exchanges. Both sides of an exchange have an incentive to lie; the recipient of data may claim that they never received the data, while the sender may not send it but claim that they did. To guarantee integrity of the usage log, we therefore require non-repudiable data exchanges. Second, the unique properties of blockchain mean that the confidentiality of stored data cannot be guaranteed by default. Even when storing minimal information, some form of identifier is required to denote ownership or association to entries. Without necessitating any information leak, the blockchain then allows any network participant to trace entries based on their identifier [68]. At best, users can try to hide their association to blockchain entries by keeping their identifier secret and creating new addresses. Even then, network participants can deduce information about users simply by analyzing publicly available data [6]. If the identifier is leaked or known to a third party, though, all respective entries can be retroactively associated with them. This has been identified as a problem of blockchain-based secure logs [28, 72], where confidentiality can be an important property [3]. More critically, recent privacy legislation such as the General Data Protection Regulation (GDPR) [29] of the European Union and the California Consumer Privacy Act (CCPA) [15] requires those who store personal data to protect and, on request, even delete it. As data stored in blockchain can be identifiable, even with typical protection measures, they fall under the provisions of privacy legislation. Especially the right to erasure has been identified as a core issue of blockchain in this context [60].

Intuitively, it seems as if this means a fundamental conflict between the requirements of inverse transparency and privacy legislation on the one hand, and permissionless blockchain technology on the other hand. Therefore, the only solution would be not to utilize blockchain when providing inverse transparency. We argue that this conflict can be solved differently, though. We aim to combine the strengths of blockchain (such as decentralization and immutability) with the requirements of inverse transparency (accountability) and privacy legislation (confidentiality, deletability).

Contribution: Blockchain is uniquely positioned to solve many issues of inverse transparency in a decentralized environment. Yet, the goal of accountability for data usages requires a solution that guarantees non-repudiable data exchanges. Furthermore, secure usage logs based on blockchain are, by default, fundamentally at odds with privacy legislation such as the GDPR. The metadata recorded in blockchain are themselves personal data that need to be protected and, on request, rectified or deleted. To tackle these challenges and enable decentralized inverse transparency, we therefore contribute the data exchange and blockchain logging system Kovacs with its core components, the new-usage protocol, and private pseudonym provisioning. Our contribution encompasses



its concept and algorithms as well as a complete open-source implementation. The new-usage protocol enables secure and decentralized data exchange while ensuring non-repudiation, as required for accountability. The one-time pseudonym generation algorithm, meanwhile, guarantees two properties: Proof of ownership, as required for deletion requests, and unlinkability, to provide users anonymity against adversaries. Notably, Kovacs does not require any changes in the utilized blockchain software and can therefore even run on arbitrary public blockchains.

Extension: This paper is an extended version of our previously published short paper [99]. In it, we introduced the P^3 concept, encompassing a new-usage protocol and private pseudonym provisioning, and analyzed its security properties theoretically. We extend upon that in multiple, significant ways: (1) We present and implement Kovacs, a complete data exchange and usage log blockchain system that integrates P^3 to increase information security and GDPR compatibility. To that end, we add a refined use case (Section 2) motivating our work, expand on the requirements, adversarial model, and system concept (Section 3), add an implementation (Section 4), and expand our discussion of related works (Section 6) to cover non-repudiable data exchange. (2) We improve the new-usage protocol (Section 3.3.1) to simplify its implementation without compromising security. (3) In addition to our theoretical analyses (Sections 5.1–5.2), we add an evaluation of the performance and scalability of the implemented Kovacs instance (Sections 5.3–5.4).

2 BACKGROUND

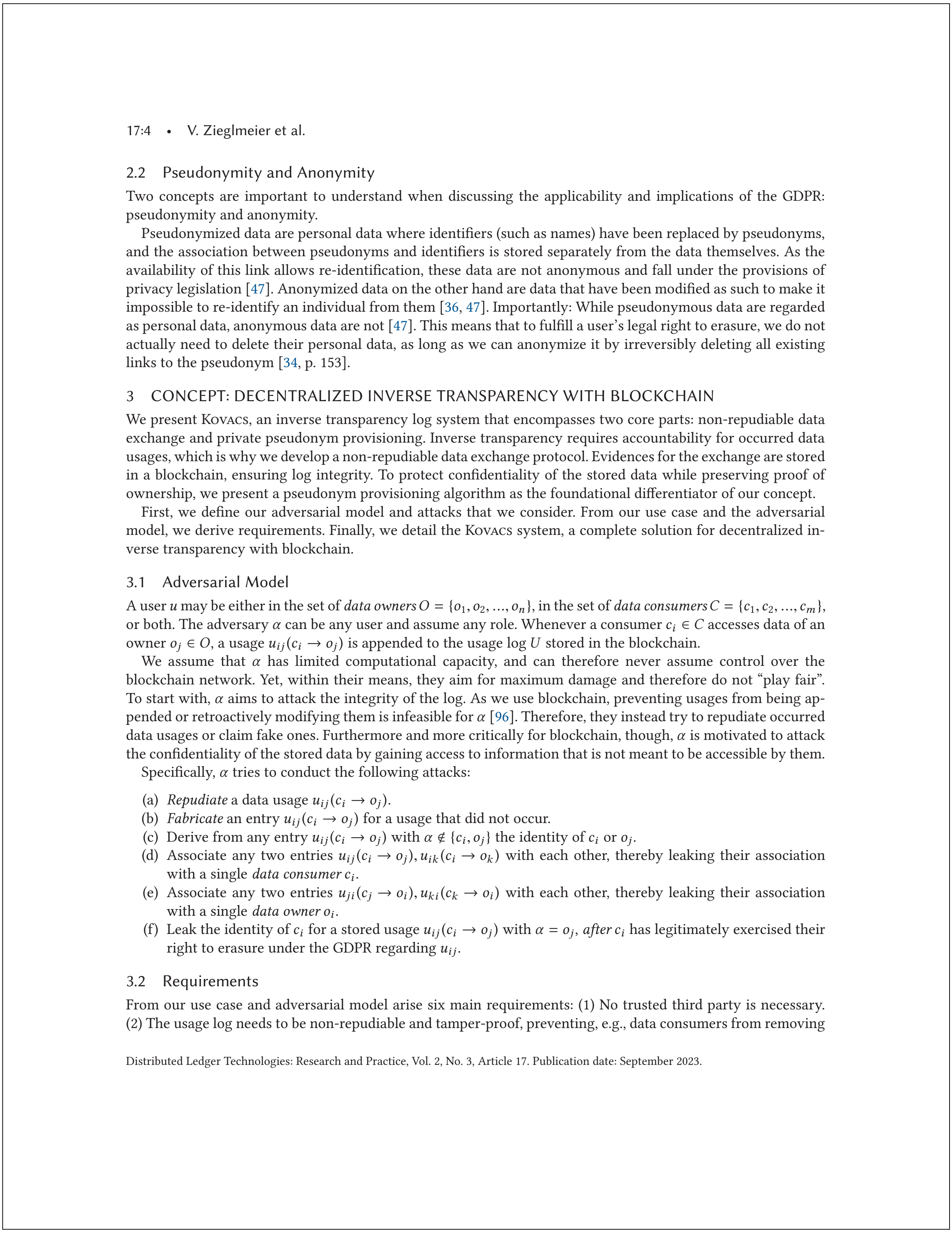
We first go into more detail regarding the concept of *inverse transparency* and why we think its current realization is in need of decentralization. Then, we outline the legal difference between pseudonymity and anonymity, an important detail that we make use of later.

2.1 Inverse Transparency

Typically, when personal data are handled, their usage is covered by privacy policies or company agreements. These policies are hard to read and understand [55], calling into question whether individuals subjected to them truly understand their impact. Especially in the workplace, this can become problematic. While some usages of their data might be beneficial for employees, giving access to personal data poses the risk of profiling and misusage. The inherent power asymmetry and forced technology adoption exacerbate these risks [85, 100]. To give employees more oversight and control in this situation of asymmetric knowledge and power, the concept of inverse transparency [14] was introduced to the workplace. At its core, it is based on the principle that access to personal data should be visible (transparent) to data owners [14]. Gierlich-Joas et al. initially described how this idea can be applied abstractly as a digital leadership concept [30]. To realize inverse transparency technically, Zieglmeier and Pretschner propose to design people analytics software from the ground up to track the flow of data and create a data usage log [102]. Their *inverse transparency toolchain* is inherently centralized, though. It requires trust in multiple parties, such as the employer and the system administrators [see 98, 101].

Tracking all data usages is an important prerequisite for inverse transparency, but for true accountability we need to be able to guarantee completeness and correctness of the created usage log. Due to the inherent power asymmetry in the workplace, it is in our view not sufficient to consider the employer a trusted party that can safeguard the logs. Ideally, no trusted third party is required at all, as they might be interested to modify the log and, e.g., remove potentially incriminating evidence. Therefore, we consider it necessary to use a distributed, tamper-proof logging mechanism to guarantee accountability [see, e.g., 3, 72, 103].

In the following, we refer to the participants in a data sharing transaction as the *data owner* and the *data consumer*. The *data owner* "possesses the rights to the data" [63, p. 40]. The GDPR refers to them as the "data subject". The *data consumer*, meanwhile, is the person or program that processes, and thereby "consumes", personal data that identify one or more *data owners* [63, p. 40]; [102].



2.2 Pseudonymity and Anonymity

Two concepts are important to understand when discussing the applicability and implications of the GDPR: pseudonymity and anonymity.

Pseudonymized data are personal data where identifiers (such as names) have been replaced by pseudonyms, and the association between pseudonyms and identifiers is stored separately from the data themselves. As the availability of this link allows re-identification, these data are not anonymous and fall under the provisions of privacy legislation [47]. Anonymized data on the other hand are data that have been modified as such to make it impossible to re-identify an individual from them [36, 47]. Importantly: While pseudonymous data are regarded as personal data, anonymous data are not [47]. This means that to fulfill a user's legal right to erasure, we do not actually need to delete their personal data, as long as we can anonymize it by irreversibly deleting all existing links to the pseudonym [34, p. 153].

3 CONCEPT: DECENTRALIZED INVERSE TRANSPARENCY WITH BLOCKCHAIN

We present Kovacs, an inverse transparency log system that encompasses two core parts: non-repudiable data exchange and private pseudonym provisioning. Inverse transparency requires accountability for occurred data usages, which is why we develop a non-repudiable data exchange protocol. Evidences for the exchange are stored in a blockchain, ensuring log integrity. To protect confidentiality of the stored data while preserving proof of ownership, we present a pseudonym provisioning algorithm as the foundational differentiator of our concept.

First, we define our adversarial model and attacks that we consider. From our use case and the adversarial model, we derive requirements. Finally, we detail the Kovacs system, a complete solution for decentralized inverse transparency with blockchain.

3.1 Adversarial Model

A user u may be either in the set of *data owners* $O = \{o_1, o_2, \ldots, o_n\}$, in the set of *data consumers* $C = \{c_1, c_2, \ldots, c_m\}$, or both. The adversary α can be any user and assume any role. Whenever a consumer $c_i \in C$ accesses data of an owner $o_j \in O$, a usage $u_{ij}(c_i \rightarrow o_j)$ is appended to the usage log U stored in the blockchain.

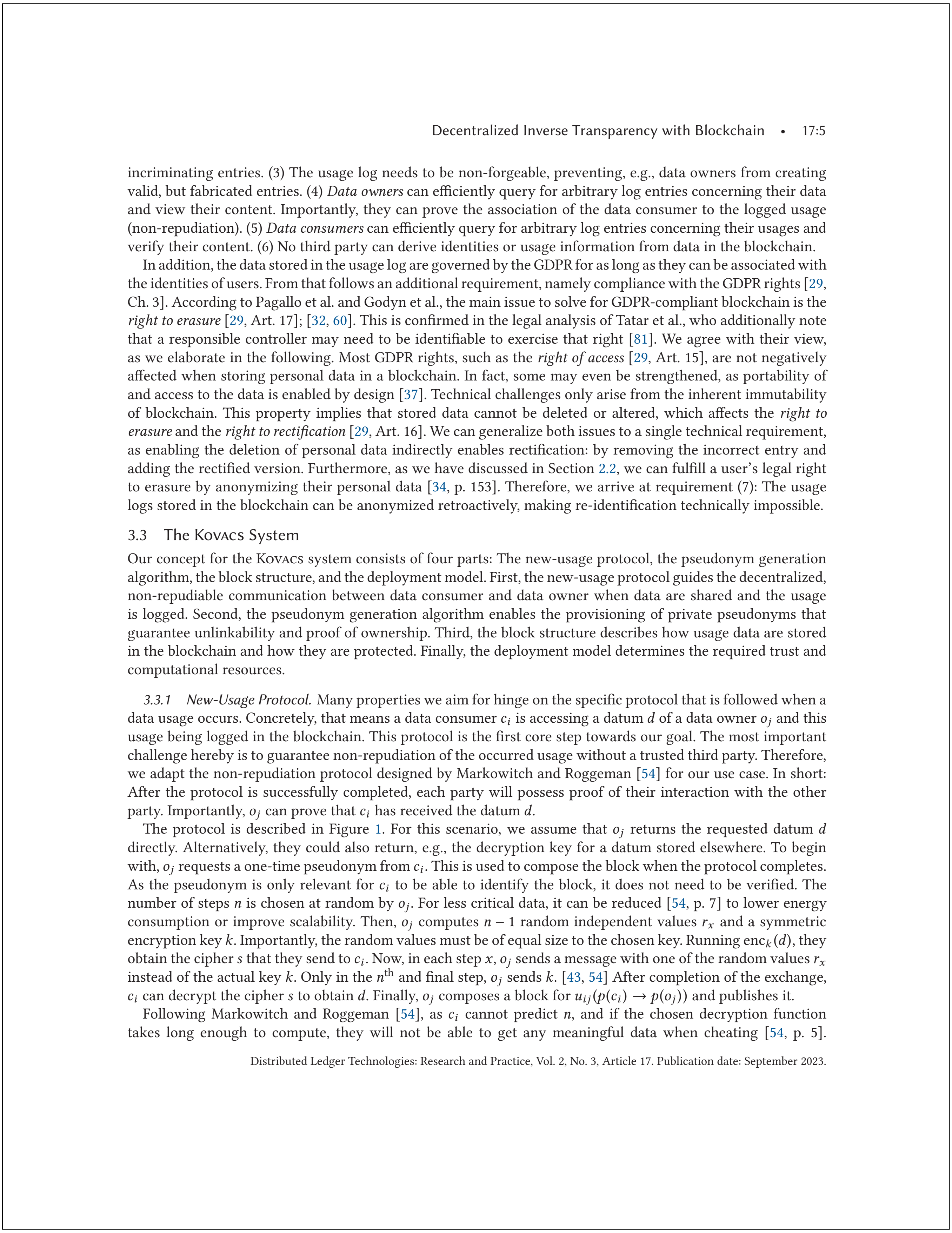
We assume that α has limited computational capacity, and can therefore never assume control over the blockchain network. Yet, within their means, they aim for maximum damage and therefore do not "play fair". To start with, α aims to attack the integrity of the log. As we use blockchain, preventing usages from being appended or retroactively modifying them is infeasible for α [96]. Therefore, they instead try to repudiate occurred data usages or claim fake ones. Furthermore and more critically for blockchain, though, α is motivated to attack the confidentiality of the stored data by gaining access to information that is not meant to be accessible by them.

Specifically, α tries to conduct the following attacks:

(a) *Repudiate* a data usage $u_{ij}(c_i \rightarrow o_j)$.
(b) *Fabricate* an entry $u_{ij}(c_i \rightarrow o_j)$ for a usage that did not occur.
(c) Derive from any entry $u_{ij}(c_i \rightarrow o_j)$ with $\alpha \notin \{c_i, o_j\}$ the identity of c_i or o_j.
(d) Associate any two entries $u_{ij}(c_i \rightarrow o_j), u_{ik}(c_i \rightarrow o_k)$ with each other, thereby leaking their association with a single *data consumer* c_i.
(e) Associate any two entries $u_{ji}(c_j \rightarrow o_i), u_{ki}(c_k \rightarrow o_i)$ with each other, thereby leaking their association with a single *data owner* o_i.
(f) Leak the identity of c_i for a stored usage $u_{ij}(c_i \rightarrow o_j)$ with $\alpha = o_j$, *after* c_i has legitimately exercised their right to erasure under the GDPR regarding u_{ij}.

3.2 Requirements

From our use case and adversarial model arise six main requirements: (1) No trusted third party is necessary. (2) The usage log needs to be non-repudiable and tamper-proof, preventing, e.g., data consumers from removing



incriminating entries. (3) The usage log needs to be non-forgeable, preventing, e.g., data owners from creating valid, but fabricated entries. (4) *Data owners* can efficiently query for arbitrary log entries concerning their data and view their content. Importantly, they can prove the association of the data consumer to the logged usage (non-repudiation). (5) *Data consumers* can efficiently query for arbitrary log entries concerning their usages and verify their content. (6) No third party can derive identities or usage information from data in the blockchain.

In addition, the data stored in the usage log are governed by the GDPR for as long as they can be associated with the identities of users. From that follows an additional requirement, namely compliance with the GDPR rights [29, Ch. 3]. According to Pagallo et al. and Godyn et al., the main issue to solve for GDPR-compliant blockchain is the *right to erasure* [29, Art. 17]; [32, 60]. This is confirmed in the legal analysis of Tatar et al., who additionally note that a responsible controller may need to be identifiable to exercise that right [81]. We agree with their view, as we elaborate in the following. Most GDPR rights, such as the *right of access* [29, Art. 15], are not negatively affected when storing personal data in a blockchain. In fact, some may even be strengthened, as portability of and access to the data is enabled by design [37]. Technical challenges only arise from the inherent immutability of blockchain. This property implies that stored data cannot be deleted or altered, which affects the *right to erasure* and the *right to rectification* [29, Art. 16]. We can generalize both issues to a single technical requirement, as enabling the deletion of personal data indirectly enables rectification: by removing the incorrect entry and adding the rectified version. Furthermore, as we have discussed in Section 2.2, we can fulfill a user's legal right to erasure by anonymizing their personal data [34, p. 153]. Therefore, we arrive at requirement (7): The usage logs stored in the blockchain can be anonymized retroactively, making re-identification technically impossible.

3.3 The Kovacs System

Our concept for the Kovacs system consists of four parts: The new-usage protocol, the pseudonym generation algorithm, the block structure, and the deployment model. First, the new-usage protocol guides the decentralized, non-repudiable communication between data consumer and data owner when data are shared and the usage is logged. Second, the pseudonym generation algorithm enables the provisioning of private pseudonyms that guarantee unlinkability and proof of ownership. Third, the block structure describes how usage data are stored in the blockchain and how they are protected. Finally, the deployment model determines the required trust and computational resources.

3.3.1 New-Usage Protocol. Many properties we aim for hinge on the specific protocol that is followed when a data usage occurs. Concretely, that means a data consumer c_i is accessing a datum d of a data owner o_j and this usage being logged in the blockchain. This protocol is the first core step towards our goal. The most important challenge hereby is to guarantee non-repudiation of the occurred usage without a trusted third party. Therefore, we adapt the non-repudiation protocol designed by Markowitch and Roggeman [54] for our use case. In short: After the protocol is successfully completed, each party will possess proof of their interaction with the other party. Importantly, o_j can prove that c_i has received the datum d.

The protocol is described in Figure 1. For this scenario, we assume that o_j returns the requested datum d directly. Alternatively, they could also return, e.g., the decryption key for a datum stored elsewhere. To begin with, o_j requests a one-time pseudonym from c_i. This is used to compose the block when the protocol completes. As the pseudonym is only relevant for c_i to be able to identify the block, it does not need to be verified. The number of steps n is chosen at random by o_j. For less critical data, it can be reduced [54, p. 7] to lower energy consumption or improve scalability. Then, o_j computes $n-1$ random independent values r_x and a symmetric encryption key k. Importantly, the random values must be of equal size to the chosen key. Running $\text{enc}_k(d)$, they obtain the cipher s that they send to c_i. Now, in each step x, o_j sends a message with one of the random values r_x instead of the actual key k. Only in the n^{th} and final step, o_j sends k. [43, 54] After completion of the exchange, c_i can decrypt the cipher s to obtain d. Finally, o_j composes a block for $u_{ij}(p(c_i) \rightarrow p(o_j))$ and publishes it.

Following Markowitch and Roggeman [54], as c_i cannot predict n, and if the chosen decryption function takes long enough to compute, they will not be able to get any meaningful data when cheating [54, p. 5].

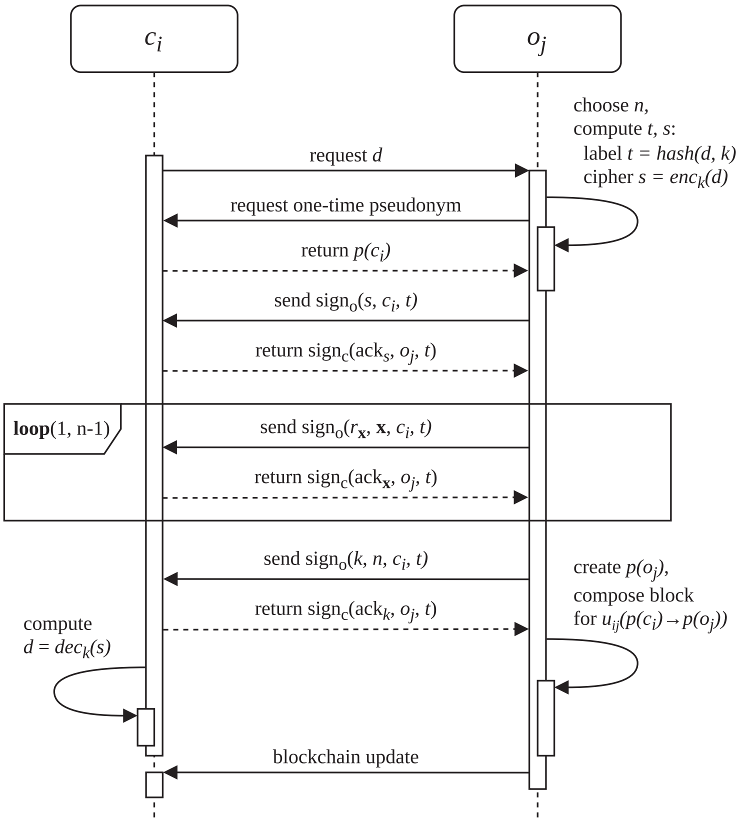
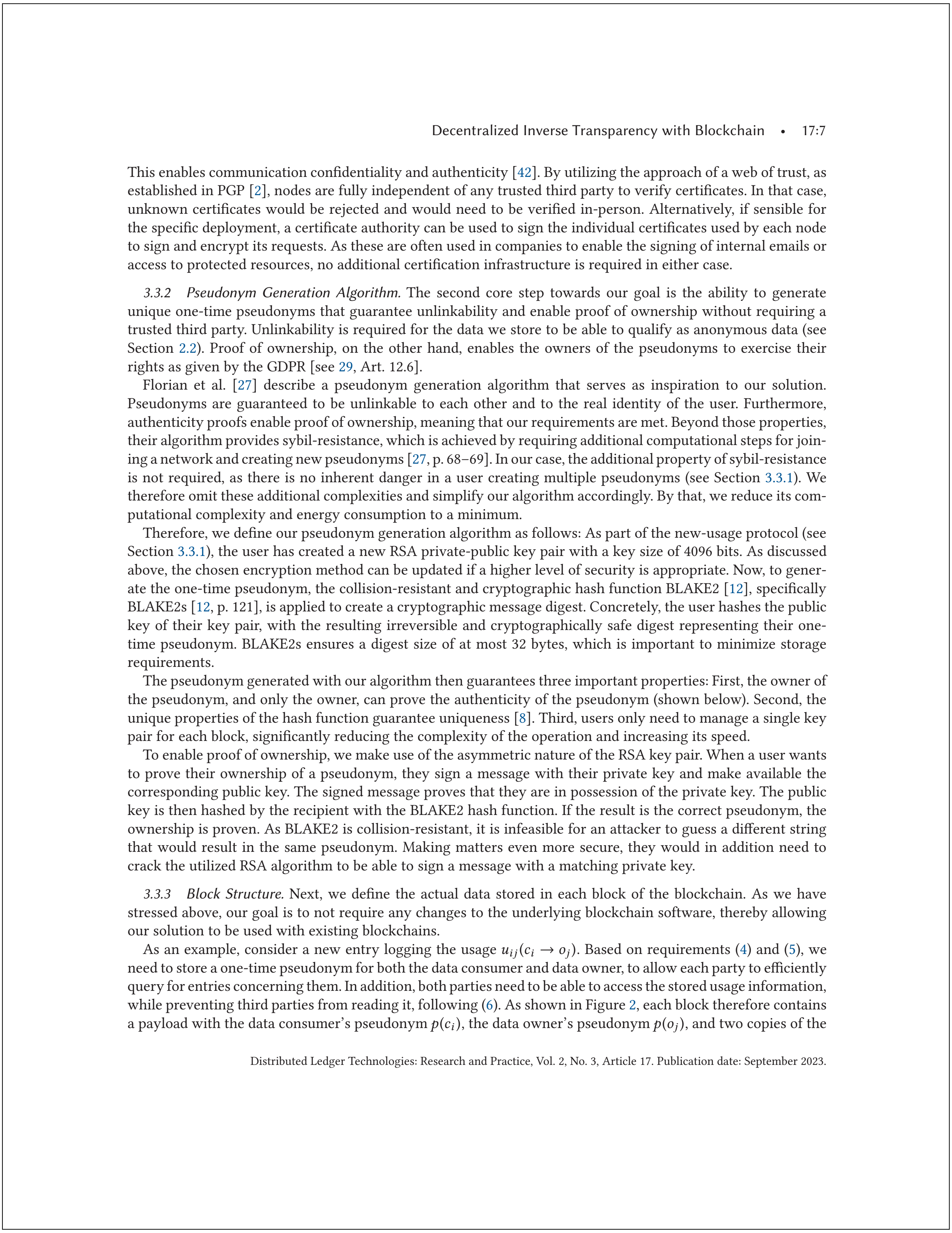

c_i
o_j
choose n,
compute t, s:
label t = $hash(d, k)$
cipher s = $enc_k(d)$
request d
request one-time pseudonym
return $p(c_i)$
send $\text{sign}_o(s, c_i, t)$
return $\text{sign}_c(\text{ack}_s, o_j, t)$
loop(1, n-1)
send $\text{sign}_o(r_\mathbf{x}, \mathbf{x}, c_i, t)$
return $\text{sign}_c(\text{ack}_\mathbf{x}, o_j, t)$
send $\text{sign}_o(k, n, c_i, t)$
return $\text{sign}_c(\text{ack}_k, o_j, t)$
create $p(o_j)$,
compose block
for $u_{ij}(p(c_i) \rightarrow p(o_j))$
compute
$d = dec_k(s)$
blockchain update

Fig. 1. The new-usage protocol, adapted from [43, 54, 75]. c_i requests access to the datum d from u_j. After receiving pseudonym $p(c_i)$, o_j sends the cipher s and the label t, which is a hash of datum and key serving as an identifier for the transaction. The receipt of s is acknowledged by c_i. Then, the iterative non-repudiation protocol is performed, with a random number of steps n unknown to c_i [see 54, pp. 5–6]. In each step x, o_j sends a random independent value r_x that must have the same size as the key k [43, p. 1610]. Only in the last message, o_j sends the actual decryption key k. Each message must be acknowledged by c_i quicker than the decryption could be performed. After receiving ack_k from c_i, o_j ends the exchange. This leads to c_i timing out, indicating that the exchange was completed. Thus, c_i can proceed to compute the datum d by decrypting s with k. Lastly, the block is composed and published. Both parties store their respective non-repudiation evidence on their own machine.

Only after having received the last message, they can decrypt the cipher [54, 75]. Now, each party holds non-repudiation evidence of the interaction. For o_j, this is $\{\, \text{sign}_c(\text{ack}_s, o_j, t), \text{sign}_c(\text{ack}_k, o_j, t) \,\}$, for c_i it is $\{\, \text{sign}_o(s, c_i, t), \text{sign}_o(k, c_i, t) \,\}$ [43, p. 1610].

This protocol depends on the nodes being able to verify the authenticity of requests and, importantly, being protected against man-in-the-middle or eavesdropper attacks. Therefore, each request is signed by the sender. We do not aim to reinvent the wheel here, instead relying on the established HTTP over TLS standard [69].



This enables communication confidentiality and authenticity [42]. By utilizing the approach of a web of trust, as established in PGP [2], nodes are fully independent of any trusted third party to verify certificates. In that case, unknown certificates would be rejected and would need to be verified in-person. Alternatively, if sensible for the specific deployment, a certificate authority can be used to sign the individual certificates used by each node to sign and encrypt its requests. As these are often used in companies to enable the signing of internal emails or access to protected resources, no additional certification infrastructure is required in either case.

3.3.2 Pseudonym Generation Algorithm. The second core step towards our goal is the ability to generate unique one-time pseudonyms that guarantee unlinkability and enable proof of ownership without requiring a trusted third party. Unlinkability is required for the data we store to be able to qualify as anonymous data (see Section 2.2). Proof of ownership, on the other hand, enables the owners of the pseudonyms to exercise their rights as given by the GDPR [see 29, Art. 12.6].

Florian et al. [27] describe a pseudonym generation algorithm that serves as inspiration to our solution. Pseudonyms are guaranteed to be unlinkable to each other and to the real identity of the user. Furthermore, authenticity proofs enable proof of ownership, meaning that our requirements are met. Beyond those properties, their algorithm provides sybil-resistance, which is achieved by requiring additional computational steps for joining a network and creating new pseudonyms [27, p. 68–69]. In our case, the additional property of sybil-resistance is not required, as there is no inherent danger in a user creating multiple pseudonyms (see Section 3.3.1). We therefore omit these additional complexities and simplify our algorithm accordingly. By that, we reduce its computational complexity and energy consumption to a minimum.

Therefore, we define our pseudonym generation algorithm as follows: As part of the new-usage protocol (see Section 3.3.1), the user has created a new RSA private-public key pair with a key size of 4096 bits. As discussed above, the chosen encryption method can be updated if a higher level of security is appropriate. Now, to generate the one-time pseudonym, the collision-resistant and cryptographic hash function BLAKE2 [12], specifically BLAKE2s [12, p. 121], is applied to create a cryptographic message digest. Concretely, the user hashes the public key of their key pair, with the resulting irreversible and cryptographically safe digest representing their one-time pseudonym. BLAKE2s ensures a digest size of at most 32 bytes, which is important to minimize storage requirements.

The pseudonym generated with our algorithm then guarantees three important properties: First, the owner of the pseudonym, and only the owner, can prove the authenticity of the pseudonym (shown below). Second, the unique properties of the hash function guarantee uniqueness [8]. Third, users only need to manage a single key pair for each block, significantly reducing the complexity of the operation and increasing its speed.

To enable proof of ownership, we make use of the asymmetric nature of the RSA key pair. When a user wants to prove their ownership of a pseudonym, they sign a message with their private key and make available the corresponding public key. The signed message proves that they are in possession of the private key. The public key is then hashed by the recipient with the BLAKE2 hash function. If the result is the correct pseudonym, the ownership is proven. As BLAKE2 is collision-resistant, it is infeasible for an attacker to guess a different string that would result in the same pseudonym. Making matters even more secure, they would in addition need to crack the utilized RSA algorithm to be able to sign a message with a matching private key.

3.3.3 Block Structure. Next, we define the actual data stored in each block of the blockchain. As we have stressed above, our goal is to not require any changes to the underlying blockchain software, thereby allowing our solution to be used with existing blockchains.

As an example, consider a new entry logging the usage $u_{ij}(c_i \rightarrow o_j)$. Based on requirements (4) and (5), we need to store a one-time pseudonym for both the data consumer and data owner, to allow each party to efficiently query for entries concerning them. In addition, both parties need to be able to access the stored usage information, while preventing third parties from reading it, following (6). As shown in Figure 2, each block therefore contains a payload with the data consumer's pseudonym $p(c_i)$, the data owner's pseudonym $p(o_j)$, and two copies of the

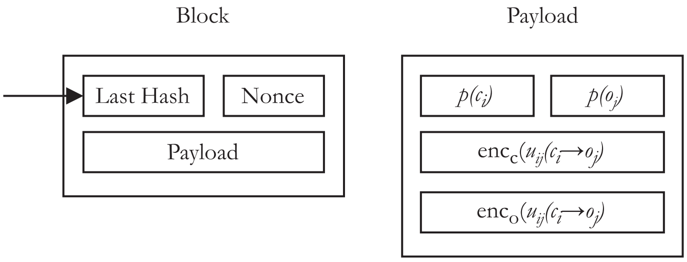

Block
Payload
Last Hash
Nonce
Payload
$p(c_i)$
$p(o_j)$
$\mathrm{enc}_{\mathrm{c}}(u_{ij}(c_i \rightarrow o_j))$
$\mathrm{enc}_{\mathrm{o}}(u_{ij}(c_i \rightarrow o_j))$

Fig. 2. The components of each block (adapted from [58]), for the exemplary usage $u_{ij}(c_i \rightarrow o_j)$. We replace the transaction history utilized in Bitcoin with a generic payload. The payload consists of the pseudonyms $p(c_i)$ and $p(o_j)$ of the data consumer and data owner, respectively. Furthermore, the logged usage is stored twice, once encrypted for the *consumer* and once for the *owner*.

usage request. This does not contain any identifiable information and only consists of the type of datum requested and a justification. To provide confidentiality, each copy is encrypted with an encryption function *enc*(), once with the *owner*'s and once with the *consumer*'s one-time public key. Importantly, this key is not shared publicly and instead stored securely with the private key. The chosen encryption should be regularly updated, but we require asymmetric (public-key) encryption [19]. As of this point, we recommend RSA [70] with a key size of 4096 bits [41, 46]. Notably, though, each individual can choose their preferred key size, and therefore security level, themselves.

Besides allowing both parties to read the usage log, storing it twice is also important because the block is created by o_j (see Section 3.3.1). c_i then needs to be able to verify the validity of the block. In case o_j manipulates the stored entry, c_i can utilize their copy of the usage and the non-repudiation evidence (see Section 3.3.1) to defend themselves against the faked evidence.

3.3.4 Deployment Model. Finally, as we have hinted at above, the chosen deployment highly influences the privacy and security guarantees that can be given. Our concept can be flexibly adapted and supports both centralized and peer-to-peer architectures. As we aim to not be dependent on any trusted third party, though, our deployment architecture is fully decentralized.

The central component of the deployment is the Kovacs system that handles data exchange, pseudonym provisioning, key management, and block creation. Each node in the peer-to-peer network runs its own Kovacs instance as well as a private non-repudiation store to store its RSA key pairs, pseudonyms, and non-repudiation evidences (see Sections 3.3.1 and 3.3.2). The data exchange does not require an intermediary and utilizes the peer-to-peer network. As the usage log blockchain is permissionless, it is shared between all nodes that want to participate in the network. That means the architecture is fully decentralized, with each node communicating directly with other nodes both for data exchanges and blockchain updates (see Figure 3).

The management of a large number of keys as required by our approach can itself become a privacy risk. In theory, the nodes would not necessarily need to store their key pairs and used pseudonyms at all to reduce their attack surface. Yet, these are important to enable requirements (4), (5), and (7). To be able to query for entries concerning their usages, users need to know which pseudonyms belong to them. In theory, they could iterate all blocks in the blockchain and simply try to decrypt their content, but this quickly becomes infeasible. Additionally, to exercise their GDPR-awarded rights, users need to prove their ownership of a pseudonym, requiring them to be in possession of the respective private key for the specific block (see Section 3.3.2). Otherwise, anyone could claim to be the owner of the encrypted data and, e.g., request its deletion. Still, each user can choose on their own how to manage their keys. If they prefer the maximum level of security, they are free to, e.g., delete new keys immediately after usage.

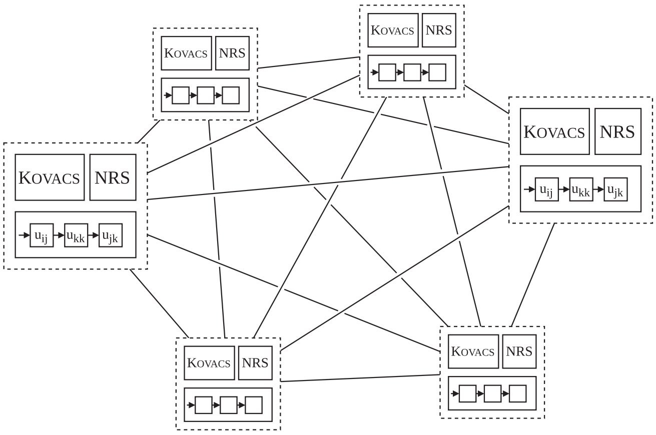


Fig. 3. Deployment in a fully decentralized peer-to-peer architecture. Each node runs its own Kovacs instance and private non-repudiation store NRS [see also 97], while the blockchain copy is shared within the network [see also 58].

4 IMPLEMENTATION

Our implementation of Kovacs represents one possible manifestation of our concept. Some important facets have to be chosen for the concrete implementation and deployment scenario, namely the concrete blockchain, realizations of the underlying algorithms, as well as the integration with a complete network. Along those facets, we therefore describe how we implement Kovacs for our proof of concept and evaluation. The source code of the final tool is available on GitHub under the MIT license.[1]

4.1 Blockchain

First, the consensus mechanism and concrete blockchain for the implementation need to be chosen. We aim for blockchain-agnosticism, therefore the concrete choices are considered an implementation detail. We only outline our considerations and final choices below to ensure transparency and reproducibility regarding our evaluation.

4.1.1 Consensus Mechanism. The choice of consensus mechanism was made between the commonly known **proof of work (PoW)** [58], **proof of stake (PoS)** [40], and **proof of authority (PoA)** [93]. We arrive at the choice by process of elimination. We cannot use PoA, as it would violate our requirements by requiring a central authority that creates and signs blocks [7, 50]. If we choose PoS, we need to have a currency that can be staked. However, a usage log has no concept of "currency". Accordingly, we use PoW as the consensus algorithm for our implementation.

4.1.2 Blockchain Implementation. We do not depend on smart contracts, which made our selection of a concrete blockchain more flexible. Potential choices included Bitcoin [58] and Ethereum [92]. The widely used Hyperledger Fabric [5] was not considered, as it is based on PoA-based consensus. Between the two, Ethereum offers multiple advantages for our use case, namely a configurable hash difficulty and support for the creation of private chains [31]. Therefore, we use it for our implementation. As the client, we use the official implementation *Go Ethereum* (Geth).

[1] https://github.com/tum-i4/kovacs.

Note that, at the time of implementation, Ethereum still used PoW consensus. Recently, Ethereum switched to PoS consensus [78]. When setting up a private network, it would therefore have to be configured to use PoW instead. Alternatively, a different PoW-based blockchain can be used.

4.2 Algorithms

With the blockchain in place, we turn to the implementation choices for the core algorithms of the Kovacs system. Broadly, the algorithms are, of course, just implementations of our concept. Yet, certain aspects may be realized in different ways depending on the application, which may impact, e.g., security or performance. Therefore, in the following, we deliberate the concrete implementation choices for the identity verification, time-asymmetric encryption, block composition, and fake chatter algorithms.

4.2.1 Identity Verification. In order to enable the traceability of data accesses, which is fundamental for inverse transparency to enable accountability [102], the data owner must know the data consumer's real identity. This enables attribution of a data usage to the responsible party. Thus, nodes need to be able to request and verify each other's identities.

We have noted in Section 3.3.1 that fully decentralized identity verification is possible by utilizing, e.g., a web of trust [2]. Yet, in our work with industry partners, we found that most companies rely on **institutional identity providers (IdP)** to realize **single sign-on (SSO)** for company-internal identity verification. As the use case of inverse transparency is specifically tailored to the company-internal context, we therefore show how to integrate an institutional IdP for identity verification. To minimize the risk of confidentiality attacks, we apply the concept of *self-sovereign identity* [see, e.g., 49, 64]. That means that each party is issued a *verifiable credential* [79] from the IdP, which it can present to exchange participants directly (see Figure 4).

Enabling self-sovereign identity requires only small adaptations to the IdP and can otherwise utilize existing authentication infrastructure. Following [57], each party's Kovacs node is issued a unique verifiable credential by the IdP on request. To trigger this, it logs in and sends a public key to be associated with their identity. The IdP then creates the verifiable credential containing of the user's IdP ID and their public key (the *claims*), and its own signature verifying the authenticity of the credential (the *proof*), and returns it to the requester [see also 79,

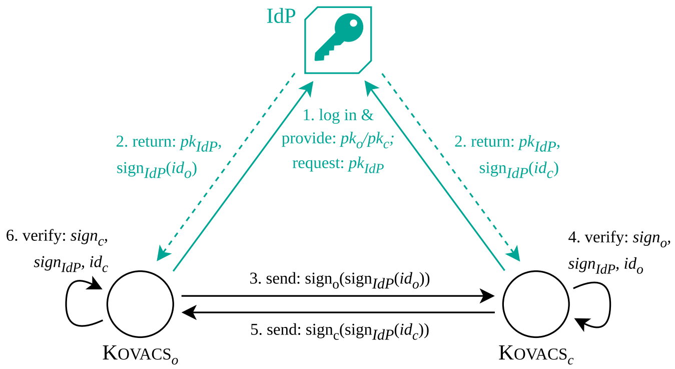
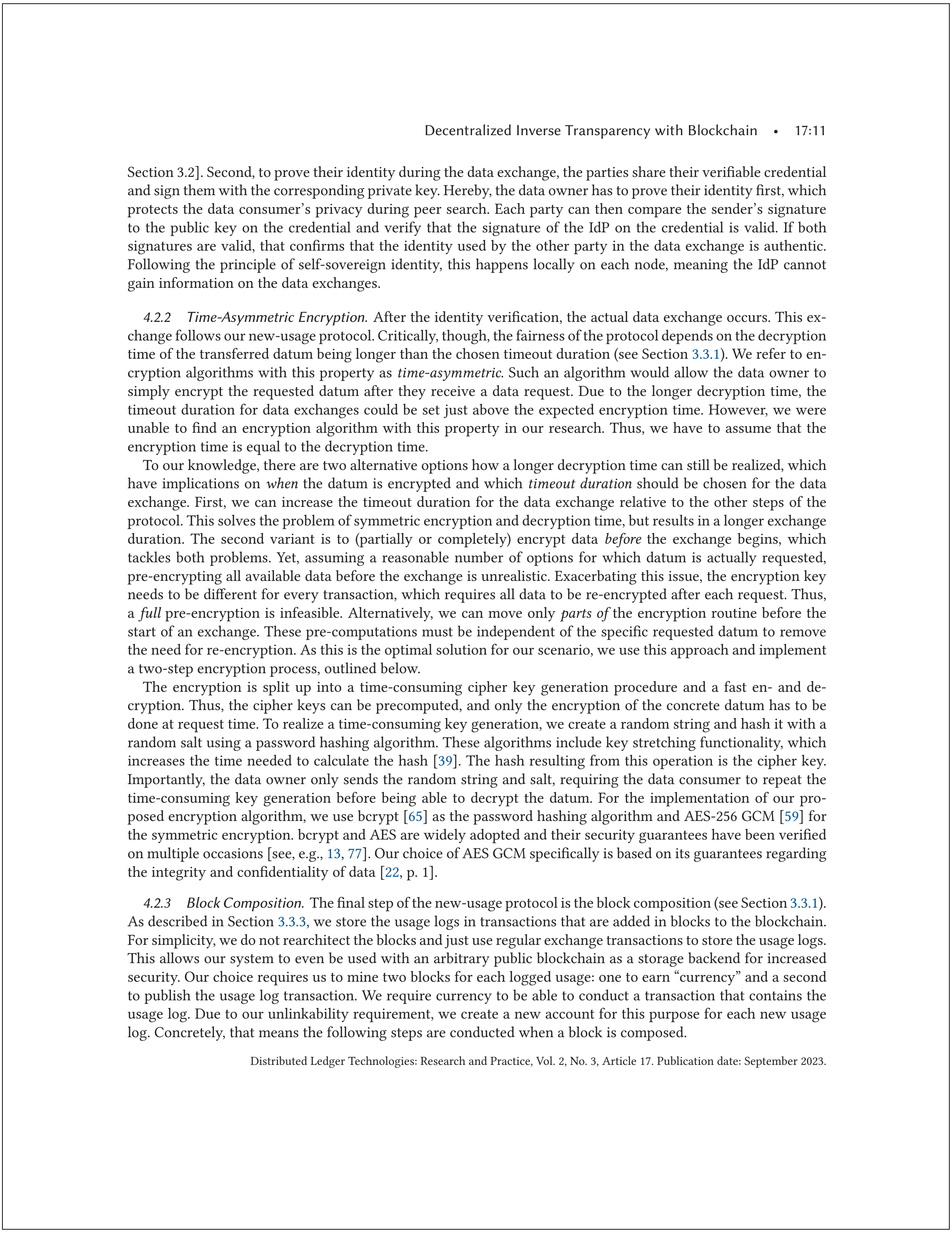

IdP
1. log in & provide: pk_o/pk_c; request: pk_{IdP}
2. return: pk_{IdP}, sign$_{IdP}$(id_o)
2. return: pk_{IdP}, sign$_{IdP}$(id_c)
6. verify: $sign_c$, $sign_{IdP}$, id_c
4. verify: $sign_o$, $sign_{IdP}$, id_o
3. send: sign$_o$(sign$_{IdP}$(id_o))
5. send: sign$_c$(sign$_{IdP}$(id_c))
Kovacs$_o$
Kovacs$_c$

Fig. 4. Identity verification steps for a data owner o and a data consumer c, utilizing an external IdP to issue verifiable credentials [following 57]; [see also 64, Chap. 3]. The Kovacs nodes of o and c log into the IdP once (1) and send their public key to receive the corresponding verifiable credentials id_o and id_c as well as its public key pk_{IdP} (2). During future data exchanges, o first provides their identity information to c (3), which c verifies locally (4). If the identity is confirmed, c also provides their identity information (5), which is verified as well by o (6).



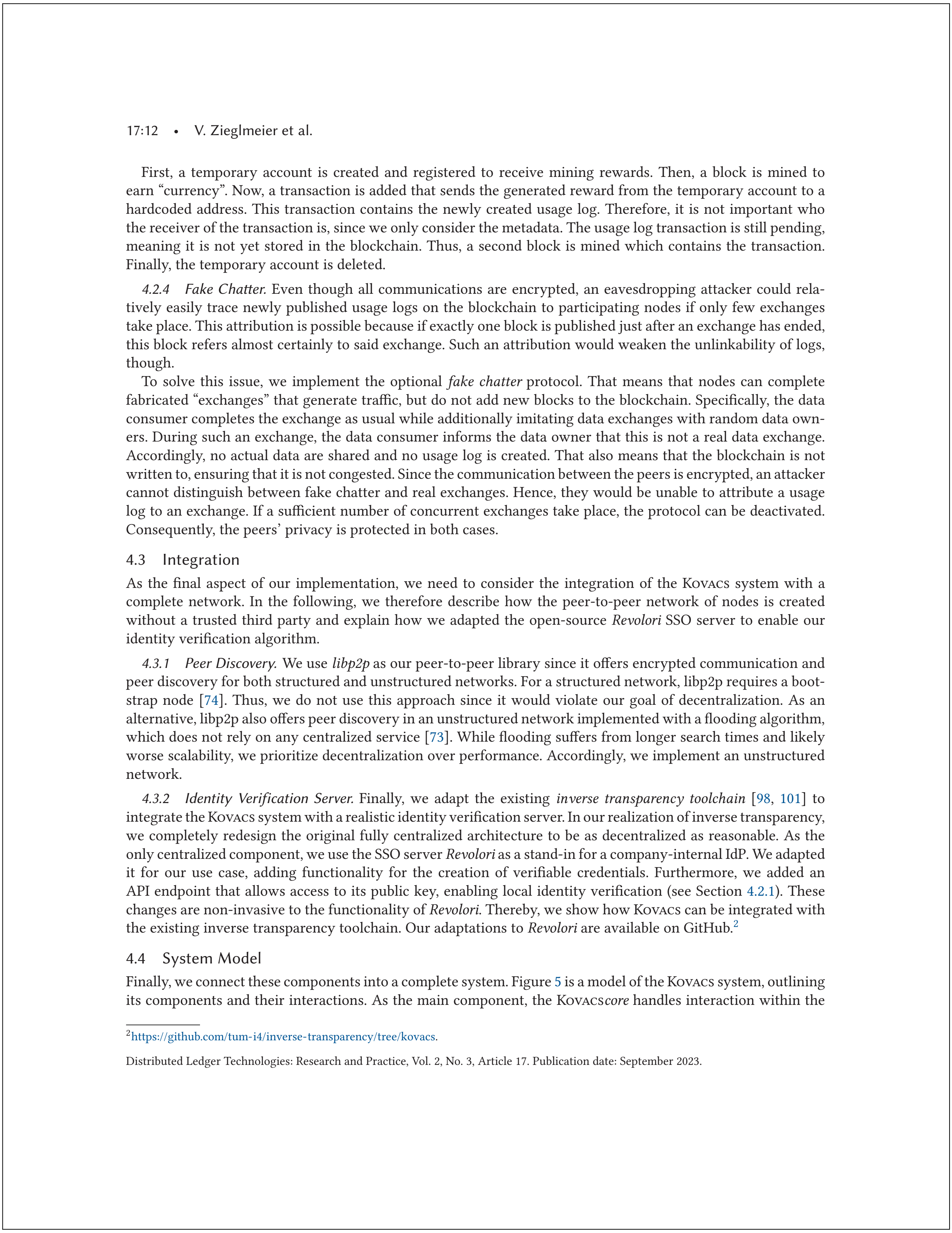
Section 3.2]. Second, to prove their identity during the data exchange, the parties share their verifiable credential and sign them with the corresponding private key. Hereby, the data owner has to prove their identity first, which protects the data consumer's privacy during peer search. Each party can then compare the sender's signature to the public key on the credential and verify that the signature of the IdP on the credential is valid. If both signatures are valid, that confirms that the identity used by the other party in the data exchange is authentic. Following the principle of self-sovereign identity, this happens locally on each node, meaning the IdP cannot gain information on the data exchanges.

4.2.2 Time-Asymmetric Encryption. After the identity verification, the actual data exchange occurs. This exchange follows our new-usage protocol. Critically, though, the fairness of the protocol depends on the decryption time of the transferred datum being longer than the chosen timeout duration (see Section 3.3.1). We refer to encryption algorithms with this property as *time-asymmetric*. Such an algorithm would allow the data owner to simply encrypt the requested datum after they receive a data request. Due to the longer decryption time, the timeout duration for data exchanges could be set just above the expected encryption time. However, we were unable to find an encryption algorithm with this property in our research. Thus, we have to assume that the encryption time is equal to the decryption time.

To our knowledge, there are two alternative options how a longer decryption time can still be realized, which have implications on *when* the datum is encrypted and which *timeout duration* should be chosen for the data exchange. First, we can increase the timeout duration for the data exchange relative to the other steps of the protocol. This solves the problem of symmetric encryption and decryption time, but results in a longer exchange duration. The second variant is to (partially or completely) encrypt data *before* the exchange begins, which tackles both problems. Yet, assuming a reasonable number of options for which datum is actually requested, pre-encrypting all available data before the exchange is unrealistic. Exacerbating this issue, the encryption key needs to be different for every transaction, which requires all data to be re-encrypted after each request. Thus, a *full* pre-encryption is infeasible. Alternatively, we can move only *parts of* the encryption routine before the start of an exchange. These pre-computations must be independent of the specific requested datum to remove the need for re-encryption. As this is the optimal solution for our scenario, we use this approach and implement a two-step encryption process, outlined below.

The encryption is split up into a time-consuming cipher key generation procedure and a fast en- and decryption. Thus, the cipher keys can be precomputed, and only the encryption of the concrete datum has to be done at request time. To realize a time-consuming key generation, we create a random string and hash it with a random salt using a password hashing algorithm. These algorithms include key stretching functionality, which increases the time needed to calculate the hash [39]. The hash resulting from this operation is the cipher key. Importantly, the data owner only sends the random string and salt, requiring the data consumer to repeat the time-consuming key generation before being able to decrypt the datum. For the implementation of our proposed encryption algorithm, we use bcrypt [65] as the password hashing algorithm and AES-256 GCM [59] for the symmetric encryption. bcrypt and AES are widely adopted and their security guarantees have been verified on multiple occasions [see, e.g., 13, 77]. Our choice of AES GCM specifically is based on its guarantees regarding the integrity and confidentiality of data [22, p. 1].

4.2.3 Block Composition. The final step of the new-usage protocol is the block composition (see Section 3.3.1). As described in Section 3.3.3, we store the usage logs in transactions that are added in blocks to the blockchain. For simplicity, we do not rearchitect the blocks and just use regular exchange transactions to store the usage logs. This allows our system to even be used with an arbitrary public blockchain as a storage backend for increased security. Our choice requires us to mine two blocks for each logged usage: one to earn "currency" and a second to publish the usage log transaction. We require currency to be able to conduct a transaction that contains the usage log. Due to our unlinkability requirement, we create a new account for this purpose for each new usage log. Concretely, that means the following steps are conducted when a block is composed.



First, a temporary account is created and registered to receive mining rewards. Then, a block is mined to earn "currency". Now, a transaction is added that sends the generated reward from the temporary account to a hardcoded address. This transaction contains the newly created usage log. Therefore, it is not important who the receiver of the transaction is, since we only consider the metadata. The usage log transaction is still pending, meaning it is not yet stored in the blockchain. Thus, a second block is mined which contains the transaction. Finally, the temporary account is deleted.

4.2.4 Fake Chatter. Even though all communications are encrypted, an eavesdropping attacker could relatively easily trace newly published usage logs on the blockchain to participating nodes if only few exchanges take place. This attribution is possible because if exactly one block is published just after an exchange has ended, this block refers almost certainly to said exchange. Such an attribution would weaken the unlinkability of logs, though.

To solve this issue, we implement the optional *fake chatter* protocol. That means that nodes can complete fabricated "exchanges" that generate traffic, but do not add new blocks to the blockchain. Specifically, the data consumer completes the exchange as usual while additionally imitating data exchanges with random data owners. During such an exchange, the data consumer informs the data owner that this is not a real data exchange. Accordingly, no actual data are shared and no usage log is created. That also means that the blockchain is not written to, ensuring that it is not congested. Since the communication between the peers is encrypted, an attacker cannot distinguish between fake chatter and real exchanges. Hence, they would be unable to attribute a usage log to an exchange. If a sufficient number of concurrent exchanges take place, the protocol can be deactivated. Consequently, the peers' privacy is protected in both cases.

4.3 Integration

As the final aspect of our implementation, we need to consider the integration of the Kovacs system with a complete network. In the following, we therefore describe how the peer-to-peer network of nodes is created without a trusted third party and explain how we adapted the open-source *Revolori* SSO server to enable our identity verification algorithm.

4.3.1 Peer Discovery. We use *libp2p* as our peer-to-peer library since it offers encrypted communication and peer discovery for both structured and unstructured networks. For a structured network, libp2p requires a bootstrap node [74]. Thus, we do not use this approach since it would violate our goal of decentralization. As an alternative, libp2p also offers peer discovery in an unstructured network implemented with a flooding algorithm, which does not rely on any centralized service [73]. While flooding suffers from longer search times and likely worse scalability, we prioritize decentralization over performance. Accordingly, we implement an unstructured network.

4.3.2 Identity Verification Server. Finally, we adapt the existing *inverse transparency toolchain* [98, 101] to integrate the Kovacs system with a realistic identity verification server. In our realization of inverse transparency, we completely redesign the original fully centralized architecture to be as decentralized as reasonable. As the only centralized component, we use the SSO server *Revolori* as a stand-in for a company-internal IdP. We adapted it for our use case, adding functionality for the creation of verifiable credentials. Furthermore, we added an API endpoint that allows access to its public key, enabling local identity verification (see Section 4.2.1). These changes are non-invasive to the functionality of *Revolori*. Thereby, we show how Kovacs can be integrated with the existing inverse transparency toolchain. Our adaptations to *Revolori* are available on GitHub.[2]

4.4 System Model

Finally, we connect these components into a complete system. Figure 5 is a model of the Kovacs system, outlining its components and their interactions. As the main component, the Kovacs*core* handles interaction within the

[2] https://github.com/tum-i4/inverse-transparency/tree/kovacs.

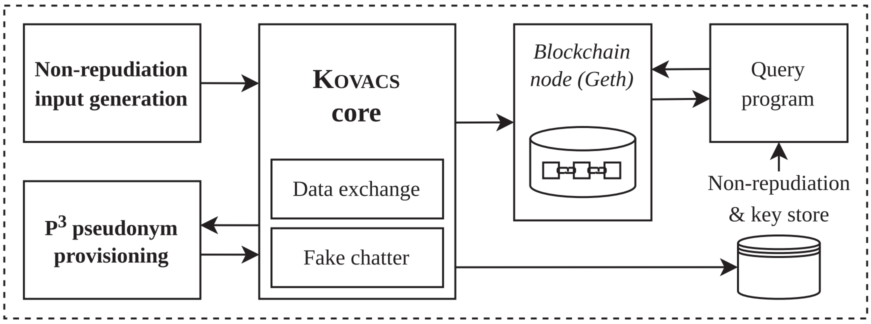

Non-repudiation input generation
KOVACS core
Data exchange
Fake chatter
P3 pseudonym provisioning
Blockchain node (Geth)
Query program
Non-repudiation & key store

Fig. 5. A KOVACS node. The arrows represent communication and data flow between the components. Bold names denote the three logic components: KOVACS core, non-repudiation input generation, and P^3 pseudonym provisioning. The name of the Geth node is italicized to denote it is a third party tool.

node and with the peer-to-peer network (see Section 3.3.4). For data exchanges it first requests input data from the *non-repudiation input generation* component, which generates the parameters for the new-usage protocol (see Section 3.3.1). Furthermore, it requests one-time pseudonyms from the P^3*pseudonym provisioning* component (see Section 3.3.2). The *query program*, finally, is utilized by the user to access the stored usage log. To do so, it loads all used pseudonyms and the decryption keys for the stored data from the *non-repudiation & key store.* Then, it queries the blockchain via the *blockchain node* to retrieve the requested usage log entries.

5 EVALUATION

To evaluate the KOVACS system, we first analyze the security of the concept and any implementations based on it. Second, we assess the GDPR compliance, specifically focusing on the data stored in the blockchain. Third, we benchmark the performance of our implementation for the most time critical operations. Finally, we measure its scalability for increasing numbers of log entries.

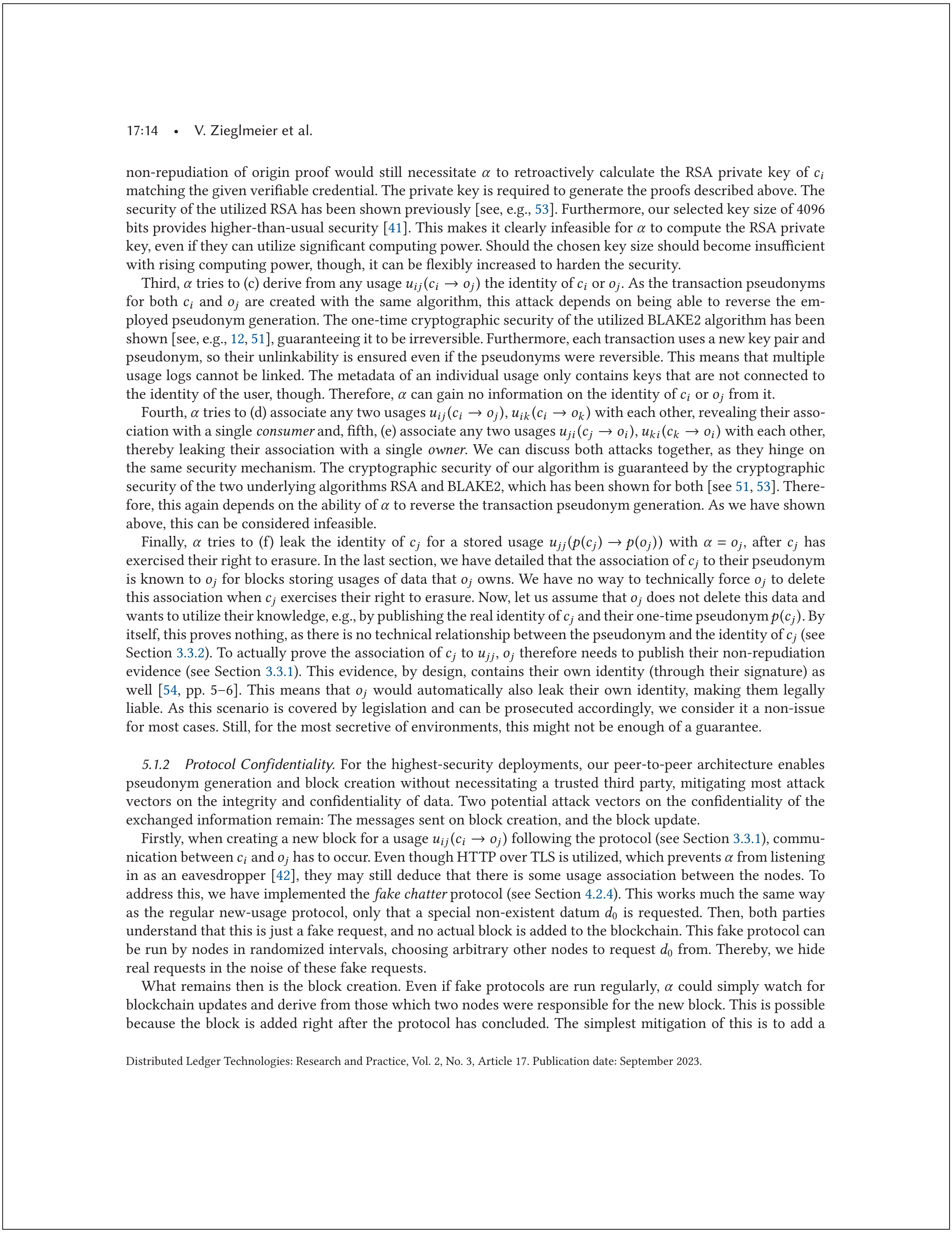
5.1 Security Analysis

We begin by analyzing the security of our system based on two core aspects: its robustness against attacks and the protocol confidentiality.

5.1.1 Robustness Against Attacks. We have described an adversarial model in Section 3.1, with adversary α trying to subvert the integrity and confidentiality of the stored logs. Specifically, they try to conduct six attacks. For each, we analyze the robustness of our approach against the attack and potential implications.

First, α tries to (a) repudiate a logged usage $u_{ij}(c_i \rightarrow o_j)$ after receiving the datum. To prevent this, we need to guarantee *non-repudiation of receipt* [97]. Our new-usage protocol is built on the Markowitch and Roggeman non-repudiation protocol, inheriting its security properties. As shown by Markowitch and Roggeman [54] and later confirmed by Aldini and Gorrieri [4], the protocol can guarantee non-repudiation of receipt based on the chosen success parameter θ, which influences the choice of the number of rounds n. As θ is chosen by the data owner, they can configure the protocol to make it infeasible for α to repudiate a logged usage, preventing α to repudiate the usage. We discuss the performance implications of this in Section 5.3.1.

Second, α tries to (b) fabricate an entry $u_{ij}(c_i \rightarrow o_j)$ for a usage that did not occur, incriminating c_i. This requires them to successfully fabricate a corresponding *non-repudiation of origin* evidence. In our case, this is $\{\ \text{sign}_c(\text{ack}_s,\ o_j, t),\ \text{sign}_c(\text{ack}_k,\ o_j, t)\ \}$ (see Section 3.3.1). For this attack to be feasible, let us assume that α previously exchanged data with c_i, thereby receiving their verifiable credential. Given this, fabricating the



non-repudiation of origin proof would still necessitate α to retroactively calculate the RSA private key of c_i matching the given verifiable credential. The private key is required to generate the proofs described above. The security of the utilized RSA has been shown previously [see, e.g., 53]. Furthermore, our selected key size of 4096 bits provides higher-than-usual security [41]. This makes it clearly infeasible for α to compute the RSA private key, even if they can utilize significant computing power. Should the chosen key size should become insufficient with rising computing power, though, it can be flexibly increased to harden the security.

Third, α tries to (c) derive from any usage $u_{ij}(c_i \rightarrow o_j)$ the identity of c_i or o_j. As the transaction pseudonyms for both c_i and o_j are created with the same algorithm, this attack depends on being able to reverse the employed pseudonym generation. The one-time cryptographic security of the utilized BLAKE2 algorithm has been shown [see, e.g., 12, 51], guaranteeing it to be irreversible. Furthermore, each transaction uses a new key pair and pseudonym, so their unlinkability is ensured even if the pseudonyms were reversible. This means that multiple usage logs cannot be linked. The metadata of an individual usage only contains keys that are not connected to the identity of the user, though. Therefore, α can gain no information on the identity of c_i or o_j from it.

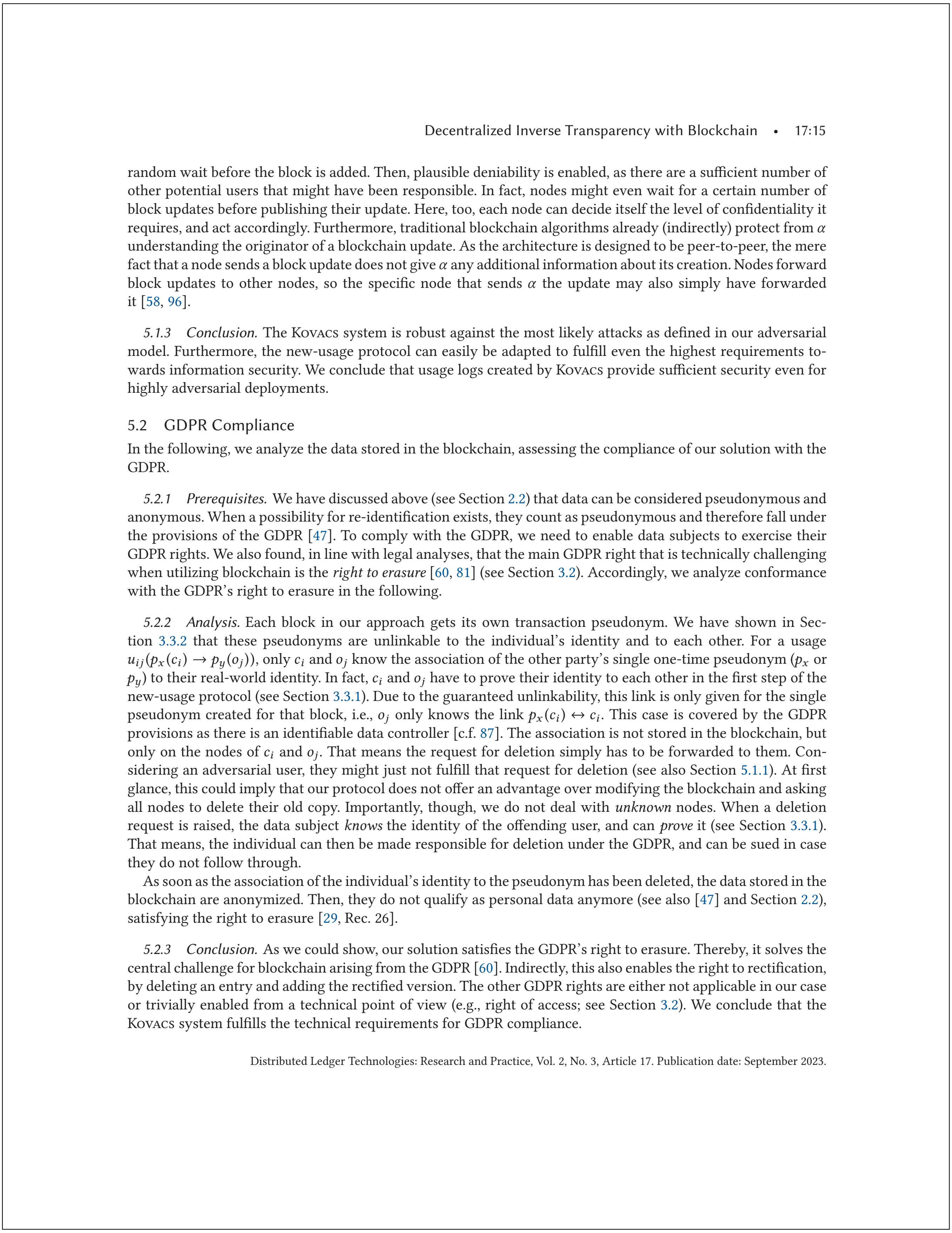
Fourth, α tries to (d) associate any two usages $u_{ij}(c_i \rightarrow o_j)$, $u_{ik}(c_i \rightarrow o_k)$ with each other, revealing their association with a single *consumer* and, fifth, (e) associate any two usages $u_{ji}(c_j \rightarrow o_i)$, $u_{ki}(c_k \rightarrow o_i)$ with each other, thereby leaking their association with a single *owner*. We can discuss both attacks together, as they hinge on the same security mechanism. The cryptographic security of our algorithm is guaranteed by the cryptographic security of the two underlying algorithms RSA and BLAKE2, which has been shown for both [see 51, 53]. Therefore, this again depends on the ability of α to reverse the transaction pseudonym generation. As we have shown above, this can be considered infeasible.

Finally, α tries to (f) leak the identity of c_j for a stored usage $u_{jj}(p(c_j) \rightarrow p(o_j))$ with $\alpha = o_j$, after c_j has exercised their right to erasure. In the last section, we have detailed that the association of c_j to their pseudonym is known to o_j for blocks storing usages of data that o_j owns. We have no way to technically force o_j to delete this association when c_j exercises their right to erasure. Now, let us assume that o_j does not delete this data and wants to utilize their knowledge, e.g., by publishing the real identity of c_j and their one-time pseudonym $p(c_j)$. By itself, this proves nothing, as there is no technical relationship between the pseudonym and the identity of c_j (see Section 3.3.2). To actually prove the association of c_j to u_{jj}, o_j therefore needs to publish their non-repudiation evidence (see Section 3.3.1). This evidence, by design, contains their own identity (through their signature) as well [54, pp. 5–6]. This means that o_j would automatically also leak their own identity, making them legally liable. As this scenario is covered by legislation and can be prosecuted accordingly, we consider it a non-issue for most cases. Still, for the most secretive of environments, this might not be enough of a guarantee.

5.1.2 Protocol Confidentiality. For the highest-security deployments, our peer-to-peer architecture enables pseudonym generation and block creation without necessitating a trusted third party, mitigating most attack vectors on the integrity and confidentiality of data. Two potential attack vectors on the confidentiality of the exchanged information remain: The messages sent on block creation, and the block update.

Firstly, when creating a new block for a usage $u_{ij}(c_i \rightarrow o_j)$ following the protocol (see Section 3.3.1), communication between c_i and o_j has to occur. Even though HTTP over TLS is utilized, which prevents α from listening in as an eavesdropper [42], they may still deduce that there is some usage association between the nodes. To address this, we have implemented the *fake chatter* protocol (see Section 4.2.4). This works much the same way as the regular new-usage protocol, only that a special non-existent datum d_0 is requested. Then, both parties understand that this is just a fake request, and no actual block is added to the blockchain. This fake protocol can be run by nodes in randomized intervals, choosing arbitrary other nodes to request d_0 from. Thereby, we hide real requests in the noise of these fake requests.

What remains then is the block creation. Even if fake protocols are run regularly, α could simply watch for blockchain updates and derive from those which two nodes were responsible for the new block. This is possible because the block is added right after the protocol has concluded. The simplest mitigation of this is to add a



random wait before the block is added. Then, plausible deniability is enabled, as there are a sufficient number of other potential users that might have been responsible. In fact, nodes might even wait for a certain number of block updates before publishing their update. Here, too, each node can decide itself the level of confidentiality it requires, and act accordingly. Furthermore, traditional blockchain algorithms already (indirectly) protect from α understanding the originator of a blockchain update. As the architecture is designed to be peer-to-peer, the mere fact that a node sends a block update does not give α any additional information about its creation. Nodes forward block updates to other nodes, so the specific node that sends α the update may also simply have forwarded it [58, 96].

5.1.3 Conclusion. The Kovacs system is robust against the most likely attacks as defined in our adversarial model. Furthermore, the new-usage protocol can easily be adapted to fulfill even the highest requirements towards information security. We conclude that usage logs created by Kovacs provide sufficient security even for highly adversarial deployments.

5.2 GDPR Compliance

In the following, we analyze the data stored in the blockchain, assessing the compliance of our solution with the GDPR.

5.2.1 Prerequisites. We have discussed above (see Section 2.2) that data can be considered pseudonymous and anonymous. When a possibility for re-identification exists, they count as pseudonymous and therefore fall under the provisions of the GDPR [47]. To comply with the GDPR, we need to enable data subjects to exercise their GDPR rights. We also found, in line with legal analyses, that the main GDPR right that is technically challenging when utilizing blockchain is the *right to erasure* [60, 81] (see Section 3.2). Accordingly, we analyze conformance with the GDPR's right to erasure in the following.

5.2.2 Analysis. Each block in our approach gets its own transaction pseudonym. We have shown in Section 3.3.2 that these pseudonyms are unlinkable to the individual's identity and to each other. For a usage $u_{ij}(p_x(c_i) \to p_y(o_j))$, only c_i and o_j know the association of the other party's single one-time pseudonym (p_x or p_y) to their real-world identity. In fact, c_i and o_j have to prove their identity to each other in the first step of the new-usage protocol (see Section 3.3.1). Due to the guaranteed unlinkability, this link is only given for the single pseudonym created for that block, i.e., o_j only knows the link $p_x(c_i) \leftrightarrow c_i$. This case is covered by the GDPR provisions as there is an identifiable data controller [c.f. 87]. The association is not stored in the blockchain, but only on the nodes of c_i and o_j. That means the request for deletion simply has to be forwarded to them. Considering an adversarial user, they might just not fulfill that request for deletion (see also Section 5.1.1). At first glance, this could imply that our protocol does not offer an advantage over modifying the blockchain and asking all nodes to delete their old copy. Importantly, though, we do not deal with *unknown* nodes. When a deletion request is raised, the data subject *knows* the identity of the offending user, and can *prove* it (see Section 3.3.1). That means, the individual can then be made responsible for deletion under the GDPR, and can be sued in case they do not follow through.

As soon as the association of the individual's identity to the pseudonym has been deleted, the data stored in the blockchain are anonymized. Then, they do not qualify as personal data anymore (see also [47] and Section 2.2), satisfying the right to erasure [29, Rec. 26].

5.2.3 Conclusion. As we could show, our solution satisfies the GDPR's right to erasure. Thereby, it solves the central challenge for blockchain arising from the GDPR [60]. Indirectly, this also enables the right to rectification, by deleting an entry and adding the rectified version. The other GDPR rights are either not applicable in our case or trivially enabled from a technical point of view (e.g., right of access; see Section 3.2). We conclude that the Kovacs system fulfills the technical requirements for GDPR compliance.

5.3 Performance

After the theoretical analysis, we assess the performance of our Kovacs implementation. All measurements were taken on an eight-core Ryzen 7 5800X with 16GB of DDR4 RAM running at 3600 MHz. The timeout duration of the non-repudiation protocol was set to three seconds. Thus, the decryption time needed to be at least three seconds, which was achieved with a bcrypt hash difficulty of 16. All nodes were run in Docker containers. Our adapted *Revolori* SSO was run on the same system, with nginx acting as a reverse proxy to allow the nodes to connect to it using the host system's network address. Geth was configured to be a full node, meaning that each node has a copy of the entire blockchain. Unless otherwise stated, the network was run with 50 peers.

We evaluate two concrete scenarios, namely the exchange duration, which differs for data owner and data consumer, as well as the log entry append time, which is only relevant for the data owner.

5.3.1 Exchange Duration. To analyze the exchange duration, we first need to define its beginning and end. From the perspective of a data consumer, the exchange begins with the start of their node and ends with them receiving and decrypting the requested datum. From the perspective of a data owner, the exchange begins when a data consumer connects to their node and ends after the new-usage protocol is completed and the usage log entry is appended to the blockchain.

Measured over 1,500 runs, we find that the median exchange durations for data owner and consumer are 6.4 and 10.6 seconds, respectively (see Figure 6), averaging 6.6 and 12.2 seconds. Thus, the exchange is approximately 5.6 seconds longer for the data consumer. This additional time for the data consumer is mainly spent waiting to time out and decrypting the received datum after the exchange has ended. To contextualize these results, we calculate the additional time effort for our approach compared to simple, repudiable peer-to-peer data sharing. For that, we measure the time for the additional required steps of our exchange, namely setup steps and peer search (≈ 3.1 s, only data consumer), the new-usage protocol (≈ 4.2 s for data owner, ≈ 9.0 s for consumer), and account and mining operations relating to the blockchain (≈ 2.5 s, only data owner). The identity verification is negligible with < 0.5 milliseconds. This shows that the largest proportion of time is spent on the new-usage protocol. Depending on network latency, this can rise further.

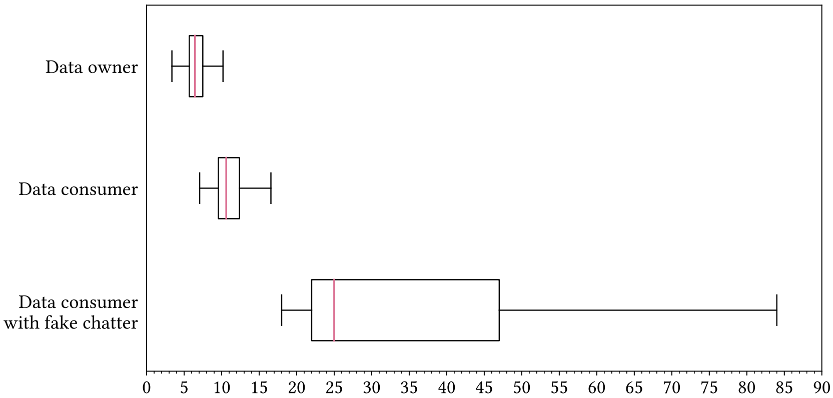


Fig. 6. Duration (in seconds) of an exchange from the perspective of the data owner, the data consumer, and the data consumer with fake chatter enabled (each measured over 1,500 runs). The medians of approximately 6.4 seconds for the data owner, 10.6 seconds for the data consumer, and 25.0 seconds for the data consumer with fake chatter are marked with pink lines. The whiskers extend to $1.5 \times IQR$, meaning they show the 25th and 75th percentiles. Outliers beyond the whiskers are not shown.

As described in Section 4.2.4, we implement the optional fake chatter protocol to increase the confidentiality of interactions. Naturally, it also reduces the performance of data exchanges, though. In our implementation, only the data consumer performs fake chatter, meaning the exchange duration is only affected for them. Figure 6 accordingly also shows the exchange duration from the perspective of the data consumer with fake chatter enabled. We aim to give an upper bound of the required time and therefore allowed fake chatter to trigger between 12 and up to 321 additional fake connections per exchange. This number is intentionally very high and can be lowered in practice. With this, we find that an exchange with fake chatter takes approximately 2.4× longer for the data consumer than without it, resulting in a median exchange duration of 25.0 seconds.

5.3.2 Appending a Log Entry. After the exchange is completed, the data owner appends the newly created usage log entry to the blockchain. Measured over 1,500 runs, we find that the median blockchain append takes 2.3 seconds (see Figure 7). In some cases, we measured durations of up to 11.9 seconds, but these are rare. We also find that the first blockchain append is consistently significantly slower than the median. We did not investigate further, as we attribute this slowdown to initiation overhead of the Geth client.

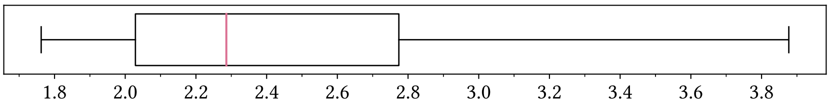


Fig. 7. Duration (in seconds) of a blockchain append, measured over 1,500 runs. The median of approximately 2.3 s is marked with a pink line. The whiskers extend to $1.5 \times IQR$, meaning they show the 25th and 75th percentiles. Outliers beyond the whiskers are not shown.

To understand which individual operations take up most time, we can roughly split the blockchain append into three steps: (1) Mining, to earn currency and store the transaction containing the usage log, (2) creating and unlocking an account to create a transaction, and (3) creating the transaction that will store the usage log. Since we use a PoW consensus algorithm, one could expect mining to be the main reason for the slow blockchain append. However, our benchmarks reveal that mining accounts for less than half of the time spent (on average 1.01 s). The remaining time is spent creating and unlocking the account (on average 1.47 s), which is performed by Geth. Finally, the time spent creating a transaction is negligible at about 2 milliseconds.

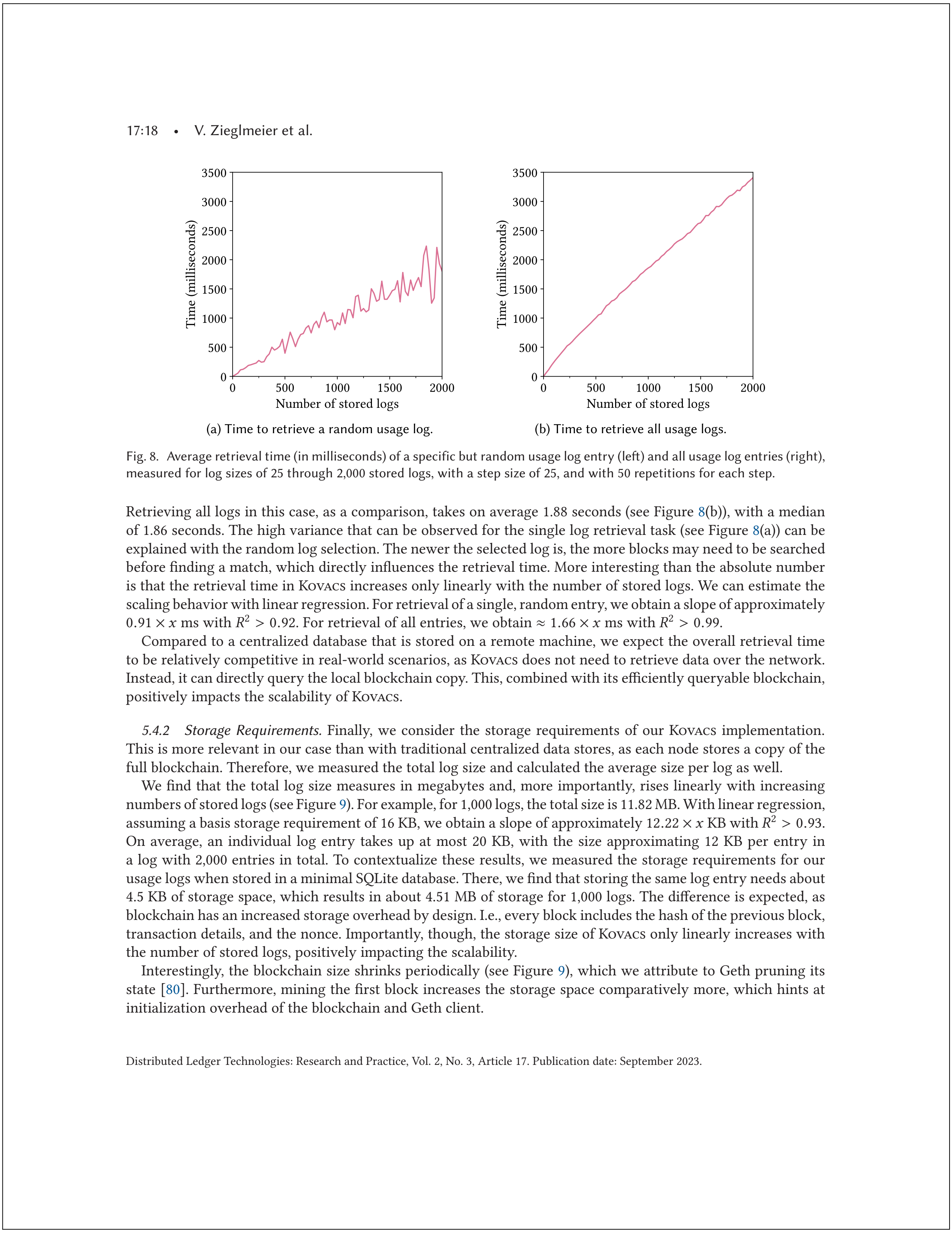
5.3.3 Conclusion. Both the exchange duration and blockchain append times of Kovacs are notably slow, being measured in seconds. While this is expected, as we focus on increased security, we need to note that this is a clear trade-off. For scenarios with lower security requirements, the significantly increased exchange duration specifically can be disqualifying. Yet, given our primary goals of decentralization and high security, we consider the time taken to still be reasonable. Furthermore, this time does not increase even if data from multiple data owners need to be requested, as requests can be parallelized. That means that these operations do not impact scalability.

5.4 Scalability

Finally, we evaluate the scalability of our Kovacs implementation. As above, all measurements were taken on an eight-core Ryzen 7 5800X with 16GB of DDR4 RAM running at 3600 MHz, and all nodes were run in Docker containers. Concretely, we evaluate the log retrieval time as well as the storage requirements for increasing numbers of log entries.

5.4.1 Retrieving Usage Logs. One of the advantages of Kovacs compared to other blockchain-based secure usage logs is that searching through log entries does not require decrypting each block. Therefore, one would expect good scaling behavior when retrieving usage logs. To evaluate this, we measured both retrieval of a single, random usage log, and retrieval of all usage logs. We ran each step 50 times and averaged the results.

Overall, we find that Kovacs returns the result in hundreds of milliseconds in both cases, with, e.g., retrieval of one log out of 1,000 stored logs taking on average 0.97 seconds (see Figure 8(a)), with a median of 0.92 seconds.



3500
3000
2500
2000
1500
1000
500
0
Time (milliseconds)
0
500
1000
1500
2000
Number of stored logs

(a) Time to retrieve a random usage log.

3500
3000
2500
2000
1500
1000
500
0
Time (milliseconds)
0
500
1000
1500
2000
Number of stored logs

(b) Time to retrieve all usage logs.

Fig. 8. Average retrieval time (in milliseconds) of a specific but random usage log entry (left) and all usage log entries (right), measured for log sizes of 25 through 2,000 stored logs, with a step size of 25, and with 50 repetitions for each step.

Retrieving all logs in this case, as a comparison, takes on average 1.88 seconds (see Figure 8(b)), with a median of 1.86 seconds. The high variance that can be observed for the single log retrieval task (see Figure 8(a)) can be explained with the random log selection. The newer the selected log is, the more blocks may need to be searched before finding a match, which directly influences the retrieval time. More interesting than the absolute number is that the retrieval time in Kovacs increases only linearly with the number of stored logs. We can estimate the scaling behavior with linear regression. For retrieval of a single, random entry, we obtain a slope of approximately $0.91 \times x$ ms with $R^2 > 0.92$. For retrieval of all entries, we obtain $\approx 1.66 \times x$ ms with $R^2 > 0.99$.

Compared to a centralized database that is stored on a remote machine, we expect the overall retrieval time to be relatively competitive in real-world scenarios, as Kovacs does not need to retrieve data over the network. Instead, it can directly query the local blockchain copy. This, combined with its efficiently queryable blockchain, positively impacts the scalability of Kovacs.

5.4.2 Storage Requirements. Finally, we consider the storage requirements of our Kovacs implementation. This is more relevant in our case than with traditional centralized data stores, as each node stores a copy of the full blockchain. Therefore, we measured the total log size and calculated the average size per log as well.

We find that the total log size measures in megabytes and, more importantly, rises linearly with increasing numbers of stored logs (see Figure 9). For example, for 1,000 logs, the total size is 11.82 MB. With linear regression, assuming a basis storage requirement of 16 KB, we obtain a slope of approximately $12.22 \times x$ KB with $R^2 > 0.93$. On average, an individual log entry takes up at most 20 KB, with the size approximating 12 KB per entry in a log with 2,000 entries in total. To contextualize these results, we measured the storage requirements for our usage logs when stored in a minimal SQLite database. There, we find that storing the same log entry needs about 4.5 KB of storage space, which results in about 4.51 MB of storage for 1,000 logs. The difference is expected, as blockchain has an increased storage overhead by design. I.e., every block includes the hash of the previous block, transaction details, and the nonce. Importantly, though, the storage size of Kovacs only linearly increases with the number of stored logs, positively impacting the scalability.

Interestingly, the blockchain size shrinks periodically (see Figure 9), which we attribute to Geth pruning its state [80]. Furthermore, mining the first block increases the storage space comparatively more, which hints at initialization overhead of the blockchain and Geth client.

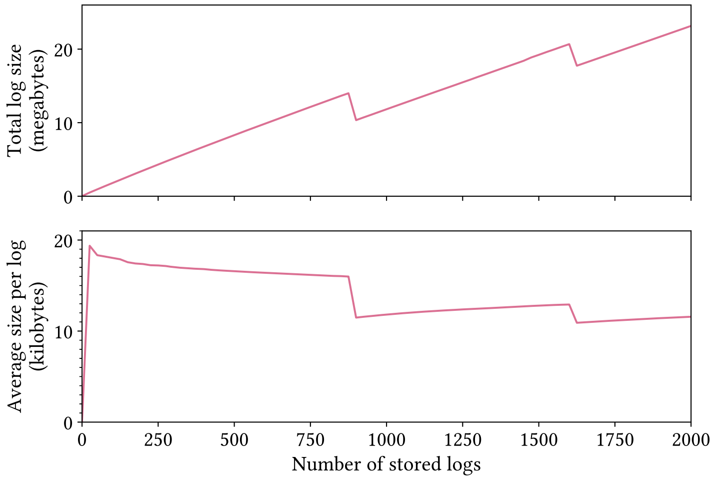


Fig. 9. The storage requirements of the blockchain log. The top graph shows the total log size, which increases linearly. The bottom graph shows the average size per log, which decreases for larger numbers of logs stored in the blockchain. For increasing log sizes, a single entry converges to a size of 12 KB.

5.4.3 Conclusion. KOVACS shows good scaling behavior. The log retrieval time rises only linearly for increasing numbers of log entries, both when retrieving a specific entry and all logs entries. In absolute terms, it is relatively low, especially considering that no additional network requests are necessary. Similarly, the storage requirements are, while larger than a minimal database, very low considering the additional metadata that need to be stored. This means that storing a full blockchain copy is unproblematic for individual nodes. We conclude that KOVACS is sufficiently scalable for real-world usage and usable even in environments with hundreds of participants.

6 RELATED WORK

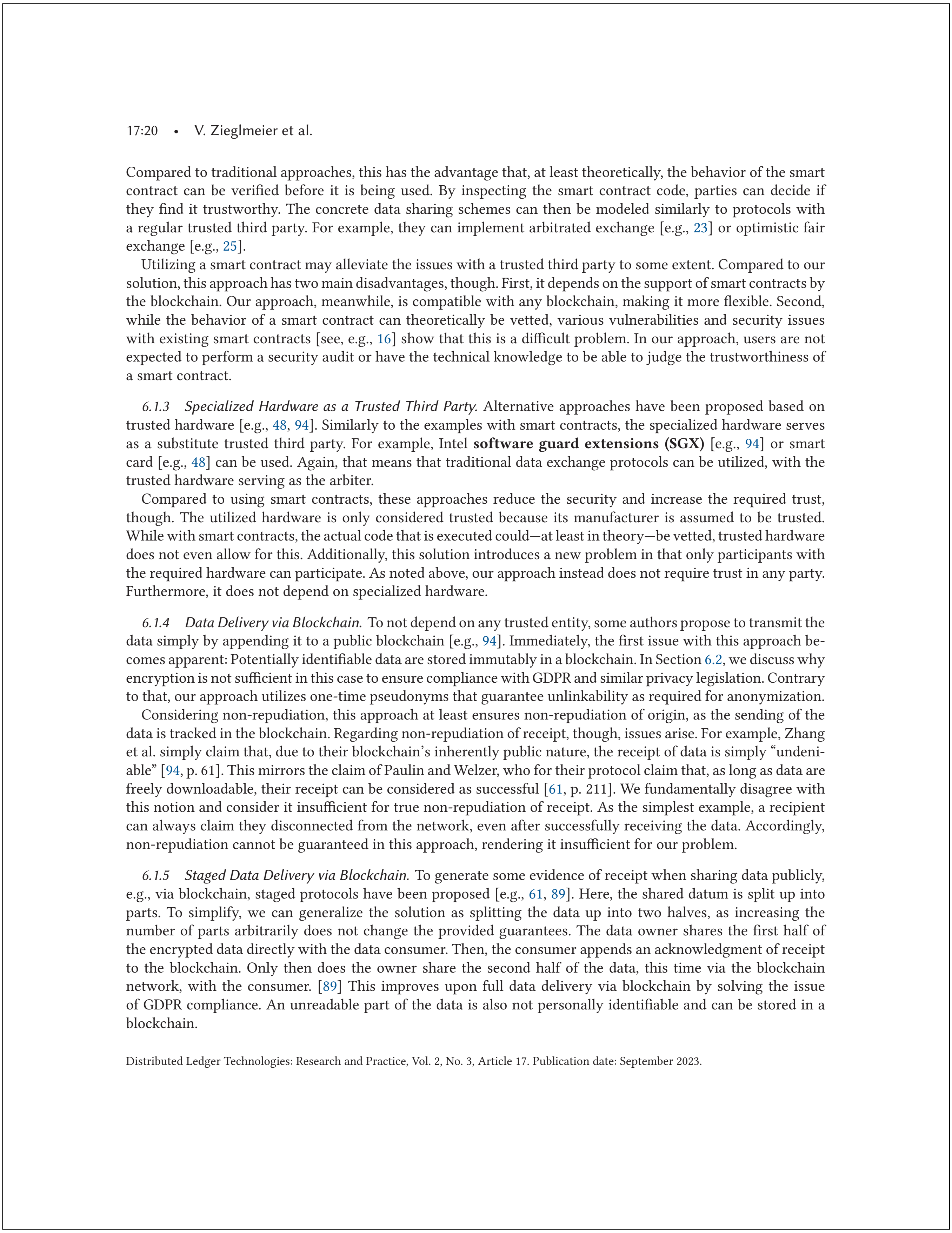
We solve two challenges of decentralized inverse transparency: non-repudiable data exchange and GDPR-compliant use of blockchain. In the following, we discuss alternative solution approaches.

6.1 Non-Repudiable Data Exchange

In decentralized scenarios, achieving non-repudiation in data exchange becomes a challenge, as we have noted. Various alternative proposals to solve this exist, of which we discuss notable examples in the following. The overview provided by Wang et al. [89, Section 1.2] and the work by Kremer et al. [43] serve as partial foundations and help confirm our research.

6.1.1 Protocols Requiring a Trusted Third Party. Many data exchange protocols exist that require a trusted third party. This includes traditional protocols based on arbitrated exchange [e.g., 1, 17], timing-based protocols [e.g., 66, 95], and optimistic fair exchange protocols [e.g., 10, 44]. This category of protocols does, by design, not fit our requirement of decentralization. Therefore, they are on principle not relevant in our scenario. Contrary to these approaches, we solve the issue of non-repudiable data exchange without a trusted third party.

6.1.2 Smart Contracts as a Trusted Third Party. To benefit from the existence of a trusted third party without necessitating the same trust, some authors propose utilizing a smart contract to fulfill that role [e.g., 23, 25].



Compared to traditional approaches, this has the advantage that, at least theoretically, the behavior of the smart contract can be verified before it is being used. By inspecting the smart contract code, parties can decide if they find it trustworthy. The concrete data sharing schemes can then be modeled similarly to protocols with a regular trusted third party. For example, they can implement arbitrated exchange [e.g., 23] or optimistic fair exchange [e.g., 25].

Utilizing a smart contract may alleviate the issues with a trusted third party to some extent. Compared to our solution, this approach has two main disadvantages, though. First, it depends on the support of smart contracts by the blockchain. Our approach, meanwhile, is compatible with any blockchain, making it more flexible. Second, while the behavior of a smart contract can theoretically be vetted, various vulnerabilities and security issues with existing smart contracts [see, e.g., 16] show that this is a difficult problem. In our approach, users are not expected to perform a security audit or have the technical knowledge to be able to judge the trustworthiness of a smart contract.

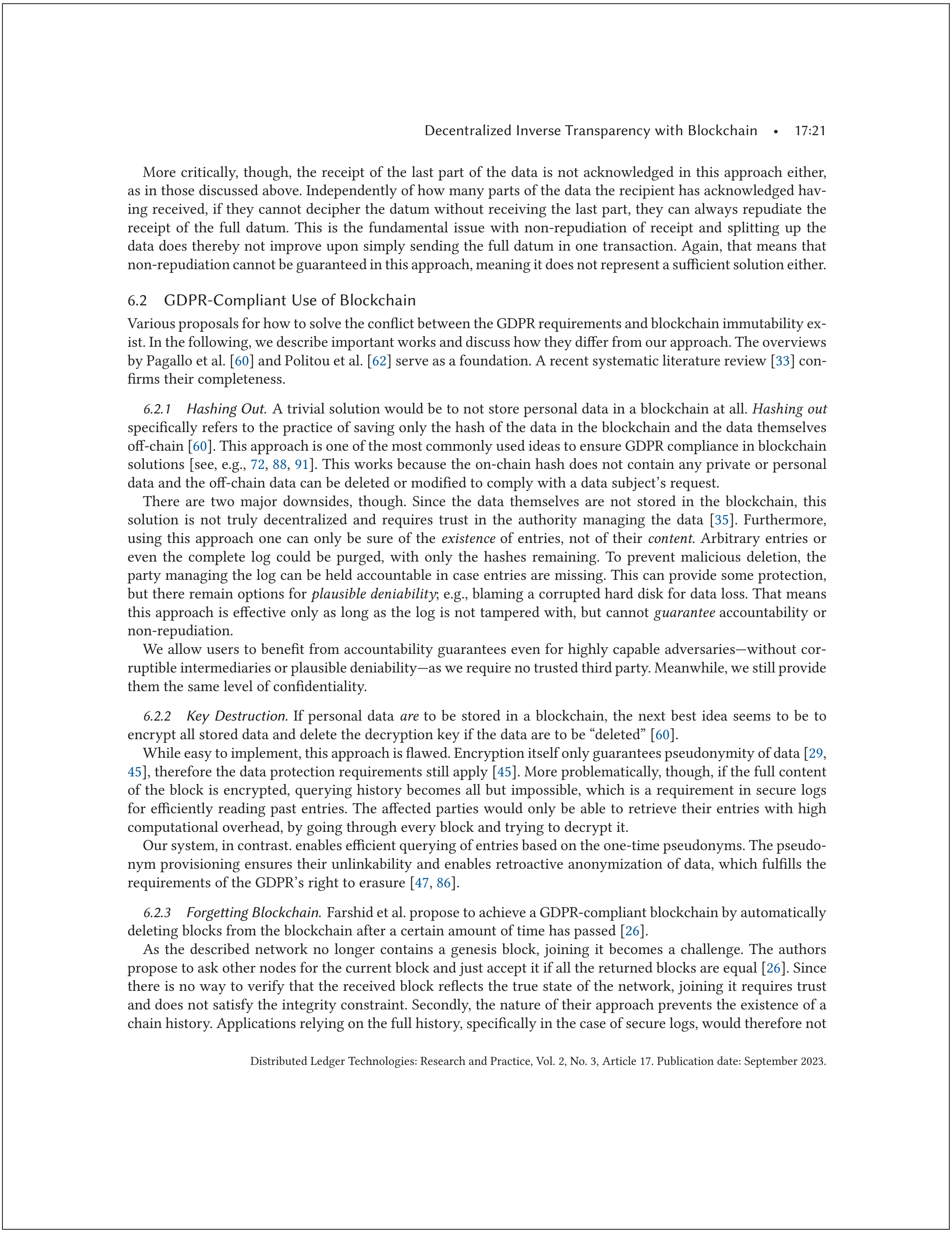
6.1.3 Specialized Hardware as a Trusted Third Party. Alternative approaches have been proposed based on trusted hardware [e.g., 48, 94]. Similarly to the examples with smart contracts, the specialized hardware serves as a substitute trusted third party. For example, Intel **software guard extensions (SGX)** [e.g., 94] or smart card [e.g., 48] can be used. Again, that means that traditional data exchange protocols can be utilized, with the trusted hardware serving as the arbiter.

Compared to using smart contracts, these approaches reduce the security and increase the required trust, though. The utilized hardware is only considered trusted because its manufacturer is assumed to be trusted. While with smart contracts, the actual code that is executed could—at least in theory—be vetted, trusted hardware does not even allow for this. Additionally, this solution introduces a new problem in that only participants with the required hardware can participate. As noted above, our approach instead does not require trust in any party. Furthermore, it does not depend on specialized hardware.

6.1.4 Data Delivery via Blockchain. To not depend on any trusted entity, some authors propose to transmit the data simply by appending it to a public blockchain [e.g., 94]. Immediately, the first issue with this approach becomes apparent: Potentially identifiable data are stored immutably in a blockchain. In Section 6.2, we discuss why encryption is not sufficient in this case to ensure compliance with GDPR and similar privacy legislation. Contrary to that, our approach utilizes one-time pseudonyms that guarantee unlinkability as required for anonymization.

Considering non-repudiation, this approach at least ensures non-repudiation of origin, as the sending of the data is tracked in the blockchain. Regarding non-repudiation of receipt, though, issues arise. For example, Zhang et al. simply claim that, due to their blockchain's inherently public nature, the receipt of data is simply "undeniable" [94, p. 61]. This mirrors the claim of Paulin and Welzer, who for their protocol claim that, as long as data are freely downloadable, their receipt can be considered as successful [61, p. 211]. We fundamentally disagree with this notion and consider it insufficient for true non-repudiation of receipt. As the simplest example, a recipient can always claim they disconnected from the network, even after successfully receiving the data. Accordingly, non-repudiation cannot be guaranteed in this approach, rendering it insufficient for our problem.

6.1.5 Staged Data Delivery via Blockchain. To generate some evidence of receipt when sharing data publicly, e.g., via blockchain, staged protocols have been proposed [e.g., 61, 89]. Here, the shared datum is split up into parts. To simplify, we can generalize the solution as splitting the data up into two halves, as increasing the number of parts arbitrarily does not change the provided guarantees. The data owner shares the first half of the encrypted data directly with the data consumer. Then, the consumer appends an acknowledgment of receipt to the blockchain. Only then does the owner share the second half of the data, this time via the blockchain network, with the consumer. [89] This improves upon full data delivery via blockchain by solving the issue of GDPR compliance. An unreadable part of the data is also not personally identifiable and can be stored in a blockchain.



More critically, though, the receipt of the last part of the data is not acknowledged in this approach either, as in those discussed above. Independently of how many parts of the data the recipient has acknowledged having received, if they cannot decipher the datum without receiving the last part, they can always repudiate the receipt of the full datum. This is the fundamental issue with non-repudiation of receipt and splitting up the data does thereby not improve upon simply sending the full datum in one transaction. Again, that means that non-repudiation cannot be guaranteed in this approach, meaning it does not represent a sufficient solution either.

6.2 GDPR-Compliant Use of Blockchain

Various proposals for how to solve the conflict between the GDPR requirements and blockchain immutability exist. In the following, we describe important works and discuss how they differ from our approach. The overviews by Pagallo et al. [60] and Politou et al. [62] serve as a foundation. A recent systematic literature review [33] confirms their completeness.

6.2.1 Hashing Out. A trivial solution would be to not store personal data in a blockchain at all. *Hashing out* specifically refers to the practice of saving only the hash of the data in the blockchain and the data themselves off-chain [60]. This approach is one of the most commonly used ideas to ensure GDPR compliance in blockchain solutions [see, e.g., 72, 88, 91]. This works because the on-chain hash does not contain any private or personal data and the off-chain data can be deleted or modified to comply with a data subject's request.

There are two major downsides, though. Since the data themselves are not stored in the blockchain, this solution is not truly decentralized and requires trust in the authority managing the data [35]. Furthermore, using this approach one can only be sure of the *existence* of entries, not of their *content.* Arbitrary entries or even the complete log could be purged, with only the hashes remaining. To prevent malicious deletion, the party managing the log can be held accountable in case entries are missing. This can provide some protection, but there remain options for *plausible deniability*; e.g., blaming a corrupted hard disk for data loss. That means this approach is effective only as long as the log is not tampered with, but cannot *guarantee* accountability or non-repudiation.

We allow users to benefit from accountability guarantees even for highly capable adversaries—without corruptible intermediaries or plausible deniability—as we require no trusted third party. Meanwhile, we still provide them the same level of confidentiality.

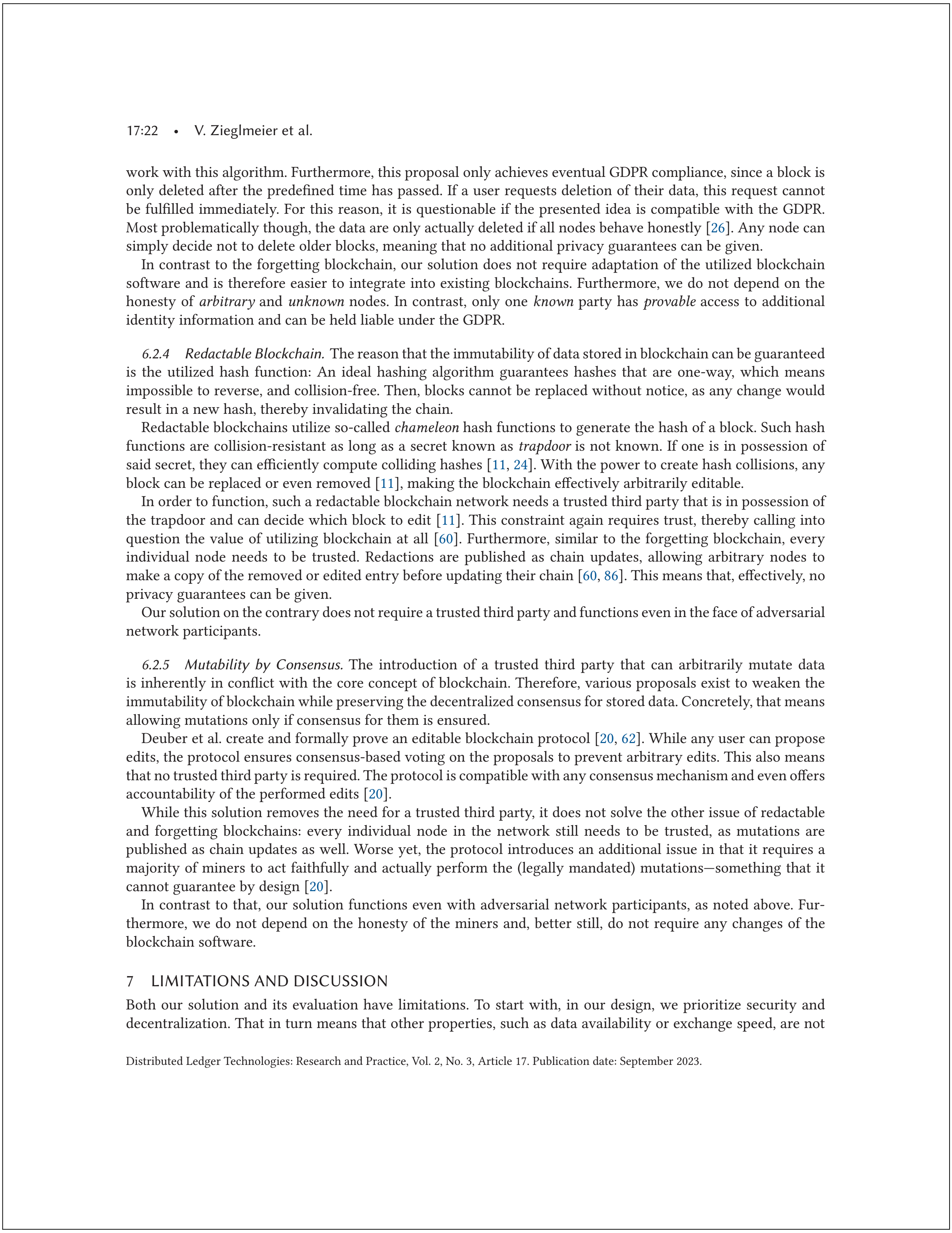
6.2.2 Key Destruction. If personal data *are* to be stored in a blockchain, the next best idea seems to be to encrypt all stored data and delete the decryption key if the data are to be "deleted" [60].

While easy to implement, this approach is flawed. Encryption itself only guarantees pseudonymity of data [29, 45], therefore the data protection requirements still apply [45]. More problematically, though, if the full content of the block is encrypted, querying history becomes all but impossible, which is a requirement in secure logs for efficiently reading past entries. The affected parties would only be able to retrieve their entries with high computational overhead, by going through every block and trying to decrypt it.

Our system, in contrast. enables efficient querying of entries based on the one-time pseudonyms. The pseudonym provisioning ensures their unlinkability and enables retroactive anonymization of data, which fulfills the requirements of the GDPR's right to erasure [47, 86].

6.2.3 Forgetting Blockchain. Farshid et al. propose to achieve a GDPR-compliant blockchain by automatically deleting blocks from the blockchain after a certain amount of time has passed [26].

As the described network no longer contains a genesis block, joining it becomes a challenge. The authors propose to ask other nodes for the current block and just accept it if all the returned blocks are equal [26]. Since there is no way to verify that the received block reflects the true state of the network, joining it requires trust and does not satisfy the integrity constraint. Secondly, the nature of their approach prevents the existence of a chain history. Applications relying on the full history, specifically in the case of secure logs, would therefore not



work with this algorithm. Furthermore, this proposal only achieves eventual GDPR compliance, since a block is only deleted after the predefined time has passed. If a user requests deletion of their data, this request cannot be fulfilled immediately. For this reason, it is questionable if the presented idea is compatible with the GDPR. Most problematically though, the data are only actually deleted if all nodes behave honestly [26]. Any node can simply decide not to delete older blocks, meaning that no additional privacy guarantees can be given.

In contrast to the forgetting blockchain, our solution does not require adaptation of the utilized blockchain software and is therefore easier to integrate into existing blockchains. Furthermore, we do not depend on the honesty of *arbitrary* and *unknown* nodes. In contrast, only one *known* party has *provable* access to additional identity information and can be held liable under the GDPR.

6.2.4 Redactable Blockchain. The reason that the immutability of data stored in blockchain can be guaranteed is the utilized hash function: An ideal hashing algorithm guarantees hashes that are one-way, which means impossible to reverse, and collision-free. Then, blocks cannot be replaced without notice, as any change would result in a new hash, thereby invalidating the chain.

Redactable blockchains utilize so-called *chameleon* hash functions to generate the hash of a block. Such hash functions are collision-resistant as long as a secret known as *trapdoor* is not known. If one is in possession of said secret, they can efficiently compute colliding hashes [11, 24]. With the power to create hash collisions, any block can be replaced or even removed [11], making the blockchain effectively arbitrarily editable.

In order to function, such a redactable blockchain network needs a trusted third party that is in possession of the trapdoor and can decide which block to edit [11]. This constraint again requires trust, thereby calling into question the value of utilizing blockchain at all [60]. Furthermore, similar to the forgetting blockchain, every individual node needs to be trusted. Redactions are published as chain updates, allowing arbitrary nodes to make a copy of the removed or edited entry before updating their chain [60, 86]. This means that, effectively, no privacy guarantees can be given.

Our solution on the contrary does not require a trusted third party and functions even in the face of adversarial network participants.

6.2.5 Mutability by Consensus. The introduction of a trusted third party that can arbitrarily mutate data is inherently in conflict with the core concept of blockchain. Therefore, various proposals exist to weaken the immutability of blockchain while preserving the decentralized consensus for stored data. Concretely, that means allowing mutations only if consensus for them is ensured.

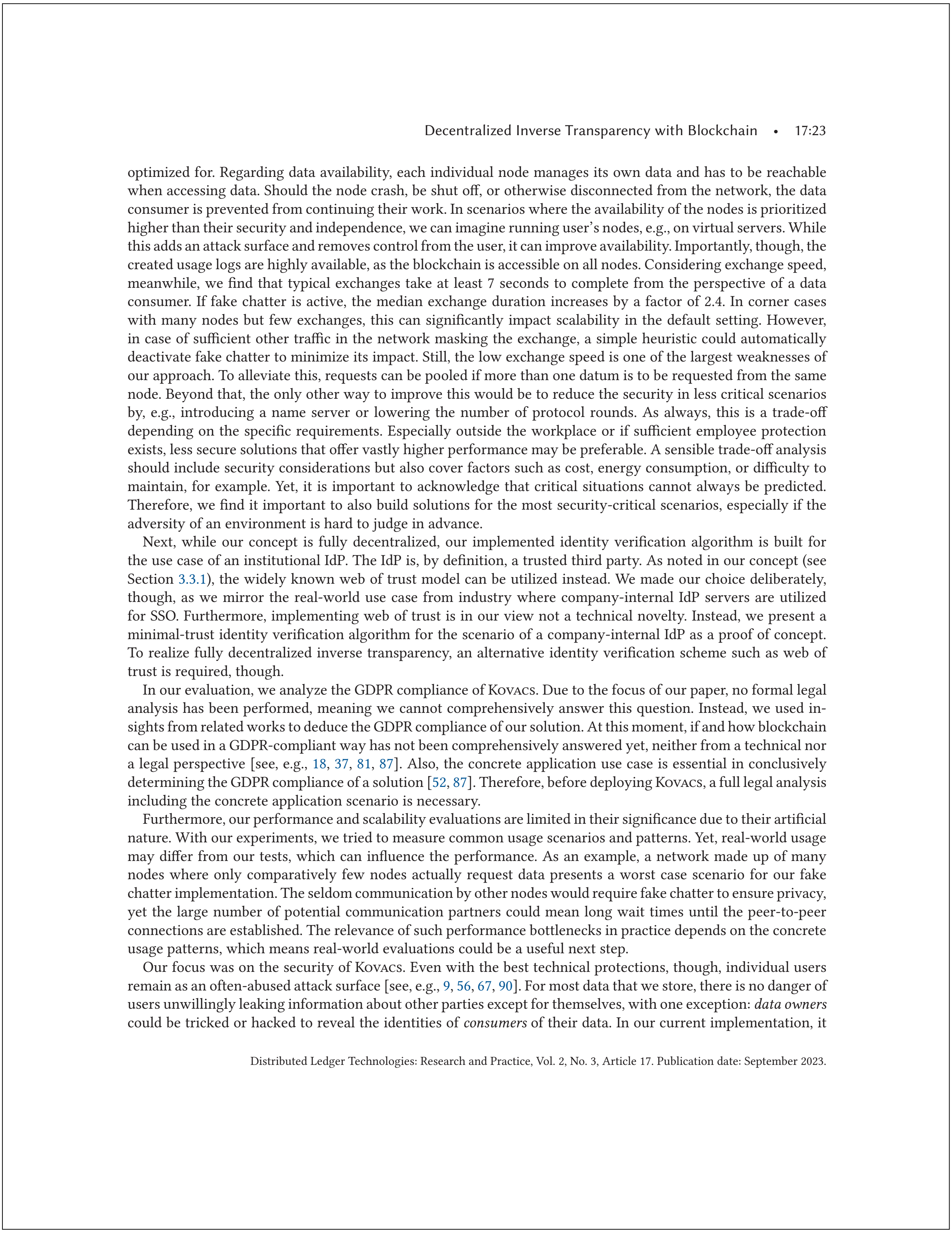
Deuber et al. create and formally prove an editable blockchain protocol [20, 62]. While any user can propose edits, the protocol ensures consensus-based voting on the proposals to prevent arbitrary edits. This also means that no trusted third party is required. The protocol is compatible with any consensus mechanism and even offers accountability of the performed edits [20].

While this solution removes the need for a trusted third party, it does not solve the other issue of redactable and forgetting blockchains: every individual node in the network still needs to be trusted, as mutations are published as chain updates as well. Worse yet, the protocol introduces an additional issue in that it requires a majority of miners to act faithfully and actually perform the (legally mandated) mutations—something that it cannot guarantee by design [20].

In contrast to that, our solution functions even with adversarial network participants, as noted above. Furthermore, we do not depend on the honesty of the miners and, better still, do not require any changes of the blockchain software.

7 LIMITATIONS AND DISCUSSION

Both our solution and its evaluation have limitations. To start with, in our design, we prioritize security and decentralization. That in turn means that other properties, such as data availability or exchange speed, are not



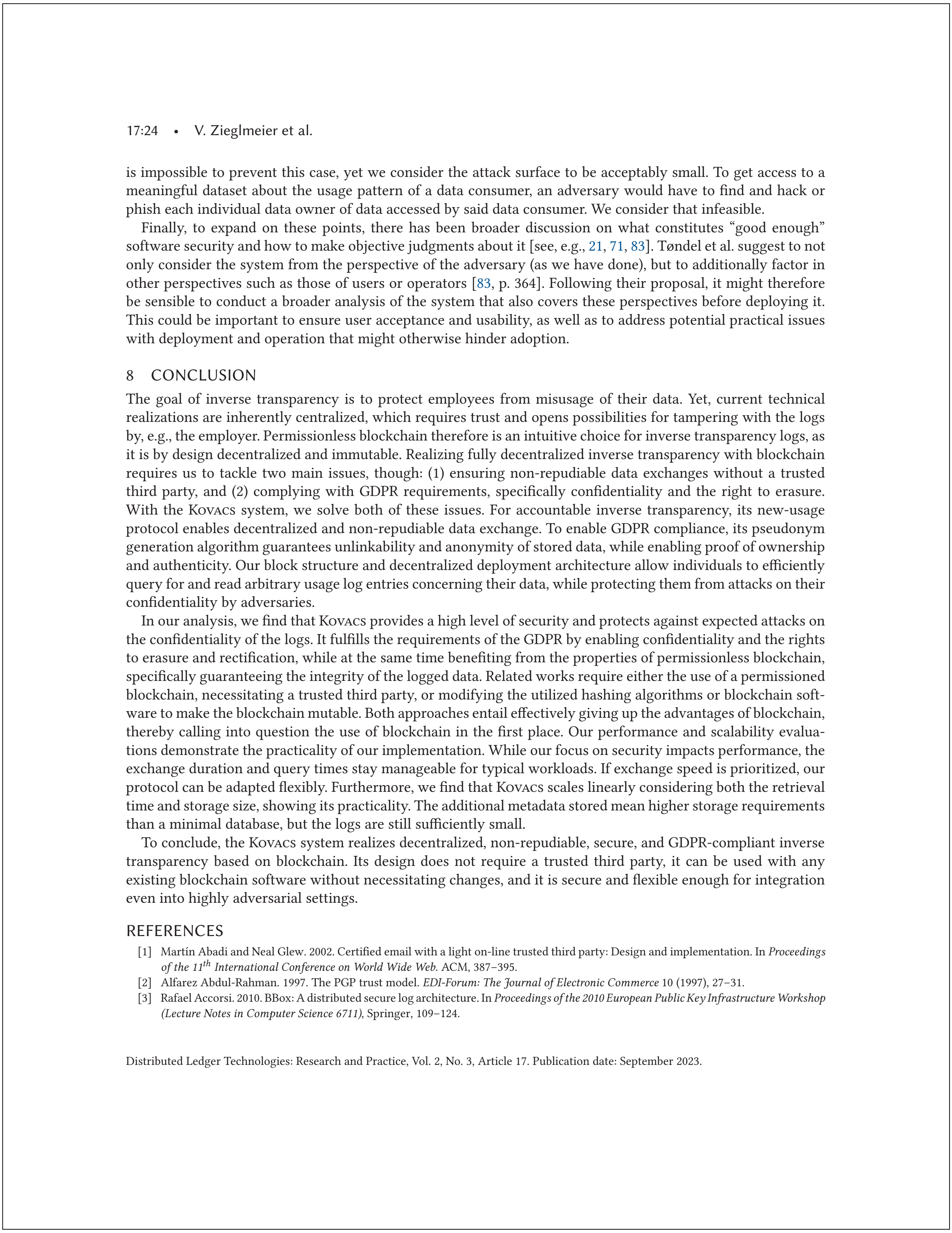
optimized for. Regarding data availability, each individual node manages its own data and has to be reachable when accessing data. Should the node crash, be shut off, or otherwise disconnected from the network, the data consumer is prevented from continuing their work. In scenarios where the availability of the nodes is prioritized higher than their security and independence, we can imagine running user's nodes, e.g., on virtual servers. While this adds an attack surface and removes control from the user, it can improve availability. Importantly, though, the created usage logs are highly available, as the blockchain is accessible on all nodes. Considering exchange speed, meanwhile, we find that typical exchanges take at least 7 seconds to complete from the perspective of a data consumer. If fake chatter is active, the median exchange duration increases by a factor of 2.4. In corner cases with many nodes but few exchanges, this can significantly impact scalability in the default setting. However, in case of sufficient other traffic in the network masking the exchange, a simple heuristic could automatically deactivate fake chatter to minimize its impact. Still, the low exchange speed is one of the largest weaknesses of our approach. To alleviate this, requests can be pooled if more than one datum is to be requested from the same node. Beyond that, the only other way to improve this would be to reduce the security in less critical scenarios by, e.g., introducing a name server or lowering the number of protocol rounds. As always, this is a trade-off depending on the specific requirements. Especially outside the workplace or if sufficient employee protection exists, less secure solutions that offer vastly higher performance may be preferable. A sensible trade-off analysis should include security considerations but also cover factors such as cost, energy consumption, or difficulty to maintain, for example. Yet, it is important to acknowledge that critical situations cannot always be predicted. Therefore, we find it important to also build solutions for the most security-critical scenarios, especially if the adversity of an environment is hard to judge in advance.

Next, while our concept is fully decentralized, our implemented identity verification algorithm is built for the use case of an institutional IdP. The IdP is, by definition, a trusted third party. As noted in our concept (see Section 3.3.1), the widely known web of trust model can be utilized instead. We made our choice deliberately, though, as we mirror the real-world use case from industry where company-internal IdP servers are utilized for SSO. Furthermore, implementing web of trust is in our view not a technical novelty. Instead, we present a minimal-trust identity verification algorithm for the scenario of a company-internal IdP as a proof of concept. To realize fully decentralized inverse transparency, an alternative identity verification scheme such as web of trust is required, though.

In our evaluation, we analyze the GDPR compliance of Kovacs. Due to the focus of our paper, no formal legal analysis has been performed, meaning we cannot comprehensively answer this question. Instead, we used insights from related works to deduce the GDPR compliance of our solution. At this moment, if and how blockchain can be used in a GDPR-compliant way has not been comprehensively answered yet, neither from a technical nor a legal perspective [see, e.g., 18, 37, 81, 87]. Also, the concrete application use case is essential in conclusively determining the GDPR compliance of a solution [52, 87]. Therefore, before deploying Kovacs, a full legal analysis including the concrete application scenario is necessary.

Furthermore, our performance and scalability evaluations are limited in their significance due to their artificial nature. With our experiments, we tried to measure common usage scenarios and patterns. Yet, real-world usage may differ from our tests, which can influence the performance. As an example, a network made up of many nodes where only comparatively few nodes actually request data presents a worst case scenario for our fake chatter implementation. The seldom communication by other nodes would require fake chatter to ensure privacy, yet the large number of potential communication partners could mean long wait times until the peer-to-peer connections are established. The relevance of such performance bottlenecks in practice depends on the concrete usage patterns, which means real-world evaluations could be a useful next step.

Our focus was on the security of Kovacs. Even with the best technical protections, though, individual users remain as an often-abused attack surface [see, e.g., 9, 56, 67, 90]. For most data that we store, there is no danger of users unwillingly leaking information about other parties except for themselves, with one exception: *data owners* could be tricked or hacked to reveal the identities of *consumers* of their data. In our current implementation, it



is impossible to prevent this case, yet we consider the attack surface to be acceptably small. To get access to a meaningful dataset about the usage pattern of a data consumer, an adversary would have to find and hack or phish each individual data owner of data accessed by said data consumer. We consider that infeasible.

Finally, to expand on these points, there has been broader discussion on what constitutes "good enough" software security and how to make objective judgments about it [see, e.g., 21, 71, 83]. Tøndel et al. suggest to not only consider the system from the perspective of the adversary (as we have done), but to additionally factor in other perspectives such as those of users or operators [83, p. 364]. Following their proposal, it might therefore be sensible to conduct a broader analysis of the system that also covers these perspectives before deploying it. This could be important to ensure user acceptance and usability, as well as to address potential practical issues with deployment and operation that might otherwise hinder adoption.

8 CONCLUSION

The goal of inverse transparency is to protect employees from misusage of their data. Yet, current technical realizations are inherently centralized, which requires trust and opens possibilities for tampering with the logs by, e.g., the employer. Permissionless blockchain therefore is an intuitive choice for inverse transparency logs, as it is by design decentralized and immutable. Realizing fully decentralized inverse transparency with blockchain requires us to tackle two main issues, though: (1) ensuring non-repudiable data exchanges without a trusted third party, and (2) complying with GDPR requirements, specifically confidentiality and the right to erasure. With the Kovacs system, we solve both of these issues. For accountable inverse transparency, its new-usage protocol enables decentralized and non-repudiable data exchange. To enable GDPR compliance, its pseudonym generation algorithm guarantees unlinkability and anonymity of stored data, while enabling proof of ownership and authenticity. Our block structure and decentralized deployment architecture allow individuals to efficiently query for and read arbitrary usage log entries concerning their data, while protecting them from attacks on their confidentiality by adversaries.

In our analysis, we find that Kovacs provides a high level of security and protects against expected attacks on the confidentiality of the logs. It fulfills the requirements of the GDPR by enabling confidentiality and the rights to erasure and rectification, while at the same time benefiting from the properties of permissionless blockchain, specifically guaranteeing the integrity of the logged data. Related works require either the use of a permissioned blockchain, necessitating a trusted third party, or modifying the utilized hashing algorithms or blockchain software to make the blockchain mutable. Both approaches entail effectively giving up the advantages of blockchain, thereby calling into question the use of blockchain in the first place. Our performance and scalability evaluations demonstrate the practicality of our implementation. While our focus on security impacts performance, the exchange duration and query times stay manageable for typical workloads. If exchange speed is prioritized, our protocol can be adapted flexibly. Furthermore, we find that Kovacs scales linearly considering both the retrieval time and storage size, showing its practicality. The additional metadata stored mean higher storage requirements than a minimal database, but the logs are still sufficiently small.

To conclude, the Kovacs system realizes decentralized, non-repudiable, secure, and GDPR-compliant inverse transparency based on blockchain. Its design does not require a trusted third party, it can be used with any existing blockchain software without necessitating changes, and it is secure and flexible enough for integration even into highly adversarial settings.

REFERENCES

[1] Martín Abadi and Neal Glew. 2002. Certified email with a light on-line trusted third party: Design and implementation. In *Proceedings of the 11th International Conference on World Wide Web*. ACM, 387–395.

[2] Alfarez Abdul-Rahman. 1997. The PGP trust model. *EDI-Forum: The Journal of Electronic Commerce* 10 (1997), 27–31.

[3] Rafael Accorsi. 2010. BBox: A distributed secure log architecture. In *Proceedings of the 2010 European Public Key Infrastructure Workshop (Lecture Notes in Computer Science 6711)*, Springer, 109–124.

[4] Alessandro Aldini and Roberto Gorrieri. 2002. Security analysis of a probabilistic non-repudiation protocol. In *Proceedings of the 2nd Joint International Workshop von Process Algebra and Probabilistic Methods, Performance Modeling and Verification (Lecture Notes in Computer Science 2399)*, Springer, 17–36.

[5] Elli Androulaki, Artem Barger, Vita Bortnikov, Christian Cachin, Konstantinos Christidis, Angelo De Caro, David Enyeart, Christopher Ferris, Gennady Laventman, Yacov Manevich, Srinivasan Muralidharan, Chet Murthy, Binh Nguyen, Manish Sethi, Gari Singh, Keith Smith, Alessandro Sorniotti, Chrysoula Stathakopoulou, Marko Vukolić, Sharon Weed Cocco, and Jason Yellick. 2018. Hyperledger Fabric: A distributed operating system for permissioned blockchains. In *Proceedings of the 13th European Conference on Computer Systems*. ACM, Article 30, 15 pages.

[6] Elli Androulaki, Ghassan O. Karame, Marc Roeschlin, Tobias Scherer, and Srdjan Capkun. 2013. Evaluating user privacy in Bitcoin. In *Proceedings of the 17th International Conference on Financial Cryptography and Data Security (Lecture Notes in Computer Science 7859)*, Springer, 34–51.

[7] Stefano De Angelis, Leonardo Aniello, Roberto Baldoni, Federico Lombardi, Andrea Margheri, and Vladimiro Sassone. 2018. PBFT vs proof-of-authority: Applying the CAP theorem to permissioned blockchain. In *Proceedings of the 2nd Italian Conference on Cyber Security*.

[8] Benny Applebaum, Naama Haramaty-Krasne, Yuval Ishai, Eyal Kushilevitz, and Vinod Vaikuntanathan. 2017. Low-complexity cryptographic hash functions. In *Proceedings of the 8th Innovations in Theoretical Computer Science Conference (Leibniz International Proceedings in Informatics)*. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Article 7, 31 pages.

[9] Iván Arce. 2003. The weakest link revisited. *IEEE Security & Privacy* 1, 2 (2003), 72–76.

[10] Nadarajah Asokan, Matthias Schunter, and Michael Waidner. 1997. Optimistic protocols for fair exchange. In *Proceedings of the 4th ACM Conference on Computer and Communications Security*. ACM, 7–17.

[11] Giuseppe Ateniese, Bernardo Magri, Daniele Venturi, and Ewerton Andrade. 2017. Redactable blockchain – or – Rewriting history in Bitcoin and friends. In *Proceedings of the 2nd IEEE European Symposium on Security and Privacy*. IEEE, 111–126.

[12] Jean-Philippe Aumasson, Samuel Neves, Zooko Wilcox-O'Hearn, and Christian Winnerlein. 2013. BLAKE2: Simpler, smaller, fast as MD5. In *Proceedings of the 11th International Conference on Applied Cryptography and Network Security (Lecture Notes in Computer Science 7954)*, Springer, 119–135.

[13] Toras Pangidoan Batubara, Syahril Efendi, and Erna Budhiarti Nababan. 2021. Analysis performance BCRYPT algorithm to improve password security from brute force. *Journal of Physics: Conference Series* 1811, 1, Article 012129 (2021).

[14] David Brin. 1998. *The Transparent Society*. Basic Books.

[15] California Consumer Privacy Act. 2018. An act to add title 1.81.5 (commencing with section 1798.100) to part 4 of division 3 of the civil Code, relating to privacy (California Consumer Privacy Act of 2018). *Assembly Bill* 375 (2018), 1–24.

[16] Huashan Chen, Marcus Pendleton, Laurent Njilla, and Shouhuai Xu. 2020. A survey on Ethereum systems security: Vulnerabilities, attacks, and defenses. *Comput. Surveys* 53, 3 (2020), 1–43.

[17] Tom Coffey and Puneet Saidha. 1996. Non-repudiation with mandatory proof of receipt. *ACM SIGCOMM Computer Communication Review* 26, 1 (1996), 6–17.

[18] Pedro Garcia de Pesquera Villagran. 2022. Blockchain technology and the general data protection regulation: An inevitable conflict?*Amsterdam Law Forum* 14 (2022), 61–64.

[19] Hans Delfs and Helmut Knebl. 2007. Public-key cryptography. In *Introduction to Cryptography*. Springer, Chapter 3, 33–80.

[20] Dominic Deuber, Bernardo Magri, and Sri Aravinda Krishnan Thyagarajan. 2019. Redactable blockchain in the permissionless setting. In *Proceedings of the 2019 IEEE Symposium on Security and Privacy*. IEEE, 124–138.

[21] John B. Dickson. 2011. Software security: Is ok good enough?. In *Proceedings of the 1st ACM Conference on Data and Application Security and Privacy*. ACM, 25–26.

[22] Morris Dworkin. 2007. *Recommendation for Block Cipher Modes of Operation: Galois/counter Mode (GCM) and GMAC*. NIST Special Publication 800-38D. National Institute of Standards and Technology.

[23] Stefan Dziembowski, Lisa Eckey, and Sebastian Faust. 2018. FairSwap: How to fairly exchange digital goods. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 967–984.

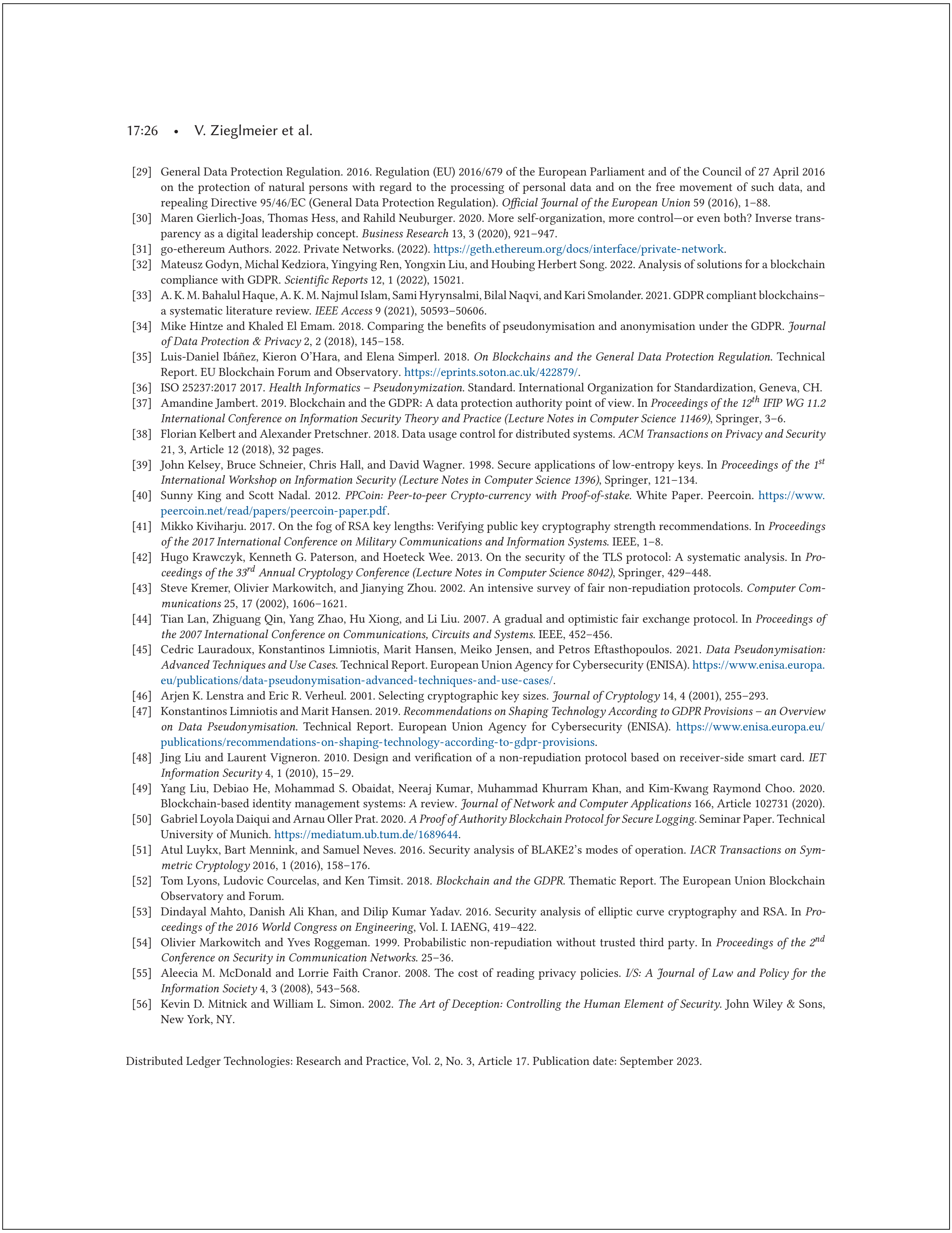
[24] Nico Döttling, Sanjam Garg, Yuval Ishai, Giulio Malavolta, Tamer Mour, and Rafail Ostrovsky. 2019. Trapdoor hash functions and their applications. In *Proceedings of the 39th Annual International Cryptology Conference (Lecture Notes in Computer Science 11694)*, Springer, 3–32.

[25] Lisa Eckey, Sebastian Faust, and Benjamin Schlosser. 2020. OptiSwap: Fast optimistic fair exchange. In *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security*. ACM, 543–557.

[26] Simon Farshid, Andreas Reitz, and Peter Roßbach. 2019. Design of a forgetting blockchain: A possible way to accomplish GDPR compatibility. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*. 7087–7095.

[27] Martin Florian, Johannes Walter, and Ingmar Baumgart. 2015. Sybil-resistant pseudonymization and pseudonym change without trusted third parties. In *Proceedings of the 14th ACM Workshop on Privacy in the Electronic Society*. ACM, 65–74.

[28] Chunpeng Ge, Siwei Sun, and Pawel Szalachowski. 2019. Permissionless blockchains and secure logging. In *Proceedings of the 2019 IEEE International Conference on Blockchain and Cryptocurrency*. IEEE, 56–60.



[29] General Data Protection Regulation. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union* 59 (2016), 1–88.

[30] Maren Gierlich-Joas, Thomas Hess, and Rahild Neuburger. 2020. More self-organization, more control—or even both? Inverse transparency as a digital leadership concept. *Business Research* 13, 3 (2020), 921–947.

[31] go-ethereum Authors. 2022. Private Networks. (2022). https://geth.ethereum.org/docs/interface/private-network.

[32] Mateusz Godyn, Michal Kedziora, Yingying Ren, Yongxin Liu, and Houbing Herbert Song. 2022. Analysis of solutions for a blockchain compliance with GDPR. *Scientific Reports* 12, 1 (2022), 15021.

[33] A. K. M. Bahalul Haque, A. K. M. Najmul Islam, Sami Hyrynsalmi, Bilal Naqvi, and Kari Smolander. 2021. GDPR compliant blockchains–a systematic literature review. *IEEE Access* 9 (2021), 50593–50606.

[34] Mike Hintze and Khaled El Emam. 2018. Comparing the benefits of pseudonymisation and anonymisation under the GDPR. *Journal of Data Protection & Privacy* 2, 2 (2018), 145–158.

[35] Luis-Daniel Ibáñez, Kieron O'Hara, and Elena Simperl. 2018. *On Blockchains and the General Data Protection Regulation.* Technical Report. EU Blockchain Forum and Observatory. https://eprints.soton.ac.uk/422879/.

[36] ISO 25237:2017 2017. *Health Informatics – Pseudonymization.* Standard. International Organization for Standardization, Geneva, CH.

[37] Amandine Jambert. 2019. Blockchain and the GDPR: A data protection authority point of view. In *Proceedings of the 12th IFIP WG 11.2 International Conference on Information Security Theory and Practice (Lecture Notes in Computer Science 11469)*, Springer, 3–6.

[38] Florian Kelbert and Alexander Pretschner. 2018. Data usage control for distributed systems. *ACM Transactions on Privacy and Security* 21, 3, Article 12 (2018), 32 pages.

[39] John Kelsey, Bruce Schneier, Chris Hall, and David Wagner. 1998. Secure applications of low-entropy keys. In *Proceedings of the 1st International Workshop on Information Security (Lecture Notes in Computer Science 1396)*, Springer, 121–134.

[40] Sunny King and Scott Nadal. 2012. *PPCoin: Peer-to-peer Crypto-currency with Proof-of-stake.* White Paper. Peercoin. https://www.peercoin.net/read/papers/peercoin-paper.pdf.

[41] Mikko Kiviharju. 2017. On the fog of RSA key lengths: Verifying public key cryptography strength recommendations. In *Proceedings of the 2017 International Conference on Military Communications and Information Systems.* IEEE, 1–8.

[42] Hugo Krawczyk, Kenneth G. Paterson, and Hoeteck Wee. 2013. On the security of the TLS protocol: A systematic analysis. In *Proceedings of the 33rd Annual Cryptology Conference (Lecture Notes in Computer Science 8042)*, Springer, 429–448.

[43] Steve Kremer, Olivier Markowitch, and Jianying Zhou. 2002. An intensive survey of fair non-repudiation protocols. *Computer Communications* 25, 17 (2002), 1606–1621.

[44] Tian Lan, Zhiguang Qin, Yang Zhao, Hu Xiong, and Li Liu. 2007. A gradual and optimistic fair exchange protocol. In *Proceedings of the 2007 International Conference on Communications, Circuits and Systems.* IEEE, 452–456.

[45] Cedric Lauradoux, Konstantinos Limniotis, Marit Hansen, Meiko Jensen, and Petros Eftasthopoulos. 2021. *Data Pseudonymisation: Advanced Techniques and Use Cases.* Technical Report. European Union Agency for Cybersecurity (ENISA). https://www.enisa.europa.eu/publications/data-pseudonymisation-advanced-techniques-and-use-cases/.

[46] Arjen K. Lenstra and Eric R. Verheul. 2001. Selecting cryptographic key sizes. *Journal of Cryptology* 14, 4 (2001), 255–293.

[47] Konstantinos Limniotis and Marit Hansen. 2019. *Recommendations on Shaping Technology According to GDPR Provisions – an Overview on Data Pseudonymisation.* Technical Report. European Union Agency for Cybersecurity (ENISA). https://www.enisa.europa.eu/publications/recommendations-on-shaping-technology-according-to-gdpr-provisions.

[48] Jing Liu and Laurent Vigneron. 2010. Design and verification of a non-repudiation protocol based on receiver-side smart card. *IET Information Security* 4, 1 (2010), 15–29.

[49] Yang Liu, Debiao He, Mohammad S. Obaidat, Neeraj Kumar, Muhammad Khurram Khan, and Kim-Kwang Raymond Choo. 2020. Blockchain-based identity management systems: A review. *Journal of Network and Computer Applications* 166, Article 102731 (2020).

[50] Gabriel Loyola Daiqui and Arnau Oller Prat. 2020. *A Proof of Authority Blockchain Protocol for Secure Logging.* Seminar Paper. Technical University of Munich. https://mediatum.ub.tum.de/1689644.

[51] Atul Luykx, Bart Mennink, and Samuel Neves. 2016. Security analysis of BLAKE2's modes of operation. *IACR Transactions on Symmetric Cryptology* 2016, 1 (2016), 158–176.

[52] Tom Lyons, Ludovic Courcelas, and Ken Timsit. 2018. *Blockchain and the GDPR.* Thematic Report. The European Union Blockchain Observatory and Forum.

[53] Dindayal Mahto, Danish Ali Khan, and Dilip Kumar Yadav. 2016. Security analysis of elliptic curve cryptography and RSA. In *Proceedings of the 2016 World Congress on Engineering*, Vol. I. IAENG, 419–422.

[54] Olivier Markowitch and Yves Roggeman. 1999. Probabilistic non-repudiation without trusted third party. In *Proceedings of the 2nd Conference on Security in Communication Networks.* 25–36.

[55] Aleecia M. McDonald and Lorrie Faith Cranor. 2008. The cost of reading privacy policies. *I/S: A Journal of Law and Policy for the Information Society* 4, 3 (2008), 543–568.

[56] Kevin D. Mitnick and William L. Simon. 2002. *The Art of Deception: Controlling the Human Element of Security.* John Wiley & Sons, New York, NY.

[57] Alexander Mühle, Andreas Grüner, Tatiana Gayvoronskaya, and Christoph Meinel. 2018. A survey on essential components of a self-sovereign identity. *Computer Science Review* 30 (2018), 80–86.

[58] Satoshi Nakamoto. 2008. Bitcoin: A Peer-to-peer Electronic Cash System. (2008).

[59] National Institute of Standards and Technology. 2001. Advanced encryption standard (AES). (2001). DOI: https://doi.org/10.6028/NIST.FIPS.197

[60] Ugo Pagallo, Eleonora Bassi, Marco Crepaldi, and Massimo Durante. 2018. Chronicle of a clash foretold: Blockchains and the GDPR's right to erasure. In *Proceedings of the 31st Annual Conference on Legal Knowledge and Information Systems*. 81–90.

[61] Alois Paulin and Tatjana Welzer. 2013. A universal system for fair non-repudiable certified e-mail without a trusted third party. *Computers & Security* 32 (2013), 207–218.

[62] Eugenia Politou, Fran Casino, Efthimios Alepis, and Constantinos Patsakis. 2019. Blockchain mutability: Challenges and proposed solutions. *IEEE Transactions on Emerging Topics in Computing* 9, 4 (2019), 1972–1986.

[63] Alexander Pretschner, Manuel Hilty, and David Basin. 2006. Distributed usage control. In *Communications of the ACM*. 39–44.

[64] Alex Preukschat and Drummond Reed. 2021. *Self-sovereign Identity*. Manning.

[65] Niels Provos and David Mazieres. 1999. A future-adaptable password scheme. In *Proceedings of the FREENIX Track: 1999 USENIX Annual Technical Conference*. USENIX, 81–91.

[66] Michael O. Rabin. 1983. Transaction protection by beacons. *J. Comput. System Sci.* 27, 2 (1983), 256–267.

[67] Marissa Randazzo, Michelle Keeney, Eileen Kowalski, Dawn Cappelli, and Andrew Moore. 2005. *Insider Threat Study: Illicit Cyber Activity in the Banking and Finance Sector*. Technical Report CMU/SEI-2004-TR-021. Software Engineering Institute, Carnegie Mellon University.

[68] Fergal Reid and Martin Harrigan. 2013. An analysis of anonymity in the Bitcoin system. In *Security and Privacy in Social Networks*, Yaniv Altshuler, Yuval Elovici, Armin B. Cremers, Nadav Aharony, and Alex Pentland (Eds.). Springer, 197–223.

[69] Eric Rescorla. 2000. *HTTP Over TLS*. RFC 2818. RFC Editor. https://www.rfc-editor.org/rfc/rfc2818.txt.

[70] Ronald L. Rivest, Adi Shamir, and Leonard Adleman. 1978. A method for obtaining digital signatures and public-key cryptosystems. *Commun. ACM* 21, 2 (1978), 120–126.

[71] Ravi Sandhu. 2003. Good-enough security. *IEEE Internet Computing* 7, 1 (2003), 66–68.

[72] Christian Schaefer and Christine Edman. 2019. Transparent logging with Hyperledger Fabric. In *Proceedings of the 2019 IEEE International Conference on Blockchain and Cryptocurrency*. IEEE, 65–69.

[73] Richard Schneider. 2021. Multicast DNS (mDNS). (2021). https://github.com/libp2p/specs/blob/master/discovery/mdns.md.

[74] Marten Seemann and Ian Davis. 2021. libp2p DHT Example Implementation. (2021). https://github.com/libp2p/go-libp2p/blob/b7bee3855cb86e50440e23b463605ea874c38787/examples/chat-with-rendezvous/chat.go#L128=.

[75] Adi Shamir. 1979. How to share a secret. *Commun. ACM* 22, 11 (1979), 612–613.

[76] Louis Shekhtman and Erez Waisbard. 2021. EngraveChain: A blockchain-based tamper-proof distributed log system. *Future Internet* 13, 6, Article 143 (2021).

[77] Gurpreet Singh. 2013. A study of encryption algorithms (RSA, DES, 3DES and AES) for information security. *International Journal of Computer Applications* 67, 19 (2013).

[78] Corwin Smith et al. 2023. Proof-of-stake (PoS). (2023). https://ethereum.org/en/developers/docs/consensus-mechanisms/pos/.

[79] Manu Sporny, Dave Longley, and David Chadwick. 2022. *Verifiable Credentials Data Model v1.1*. W3C Recommendation. World Wide Web Consortium. https://www.w3.org/TR/2022/REC-vc-data-model-20220303/.

[80] Péter Szilágyi. 2021. Geth v1.10.0 – Offline Pruning. (2021). https://blog.ethereum.org/2021/03/03/geth-v1-10-0/#offline-pruning.

[81] Unal Tatar, Yasir Gokce, and Brian Nussbaum. 2020. Law versus technology: Blockchain, GDPR, and tough tradeoffs. *Computer Law & Security Review* 38, Article 105454 (2020), 11 pages.

[82] Paul J. Taylor, Tooska Dargahi, Ali Dehghantanha, Reza M. Parizi, and Kim-Kwang Raymond Choo. 2020. A systematic literature review of blockchain cyber security. *Digital Communications and Networks* 6, 2 (2020), 147–156.

[83] Inger Anne Tøndel, Daniela Soares Cruzes, and Martin Gilje Jaatun. 2020. Achieving "good enough" software security: The role of objectivity. In *Proceedings of the 2020 International Conference on Evaluation and Assessment in Software Engineering*. ACM, 360–365.

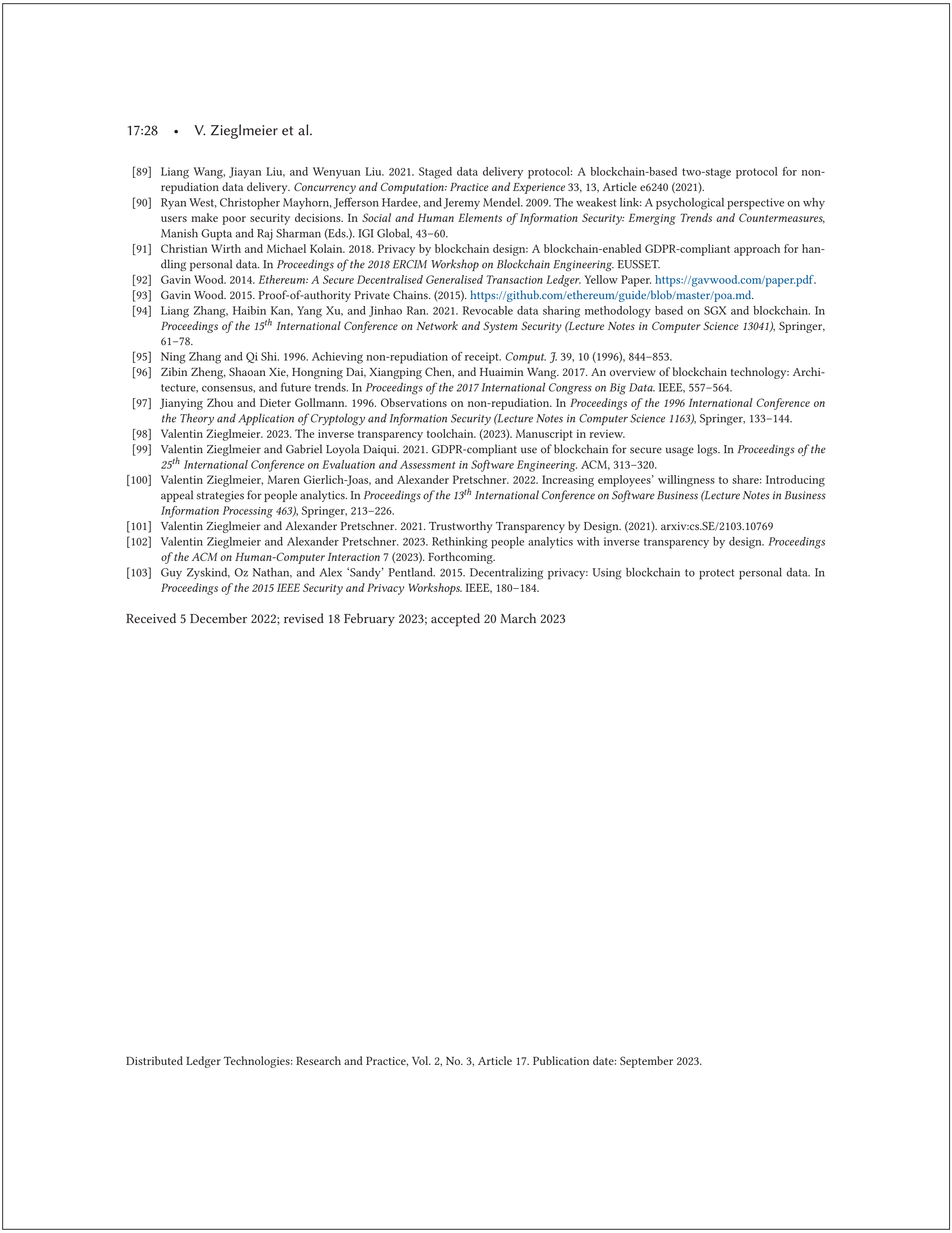
[84] Aizhan Tursunbayeva, Stefano Di Lauro, and Claudia Pagliari. 2018. People analytics—a scoping review of conceptual boundaries and value propositions. *International Journal of Information Management* 43 (2018), 224–247.

[85] Aizhan Tursunbayeva, Claudia Pagliari, Stefano Di Lauro, and Gilda Antonelli. 2021. The ethics of people analytics: Risks, opportunities and recommendations. *Personnel Review* 51, 3 (2021), 900–921.

[86] David van de Giessen. 2019. *Blockchain and the GDPR's Right to Erasure*. Essay. University of Twente.

[87] Patrick Van Eecke and Anne-Gabrielle Haie. 2018. Blockchain and the GDPR: The EU blockchain observatory report. *European Data Protection Law Review* 4/2018, 4 (2018), 531–534.

[88] Laurens Van Hoye, Pieter-Jan Maenhaut, Tim Wauters, Bruno Volckaert, and Filip De Turck. 2019. Logging mechanism for cross-organizational collaborations using Hyperledger Fabric. In *Proceedings of the 2019 IEEE International Conference on Blockchain and Cryptocurrency*. IEEE, 352–359.



[89] Liang Wang, Jiayan Liu, and Wenyuan Liu. 2021. Staged data delivery protocol: A blockchain-based two-stage protocol for non-repudiation data delivery. *Concurrency and Computation: Practice and Experience* 33, 13, Article e6240 (2021).

[90] Ryan West, Christopher Mayhorn, Jefferson Hardee, and Jeremy Mendel. 2009. The weakest link: A psychological perspective on why users make poor security decisions. In *Social and Human Elements of Information Security: Emerging Trends and Countermeasures*, Manish Gupta and Raj Sharman (Eds.). IGI Global, 43–60.

[91] Christian Wirth and Michael Kolain. 2018. Privacy by blockchain design: A blockchain-enabled GDPR-compliant approach for handling personal data. In *Proceedings of the 2018 ERCIM Workshop on Blockchain Engineering*. EUSSET.

[92] Gavin Wood. 2014. *Ethereum: A Secure Decentralised Generalised Transaction Ledger*. Yellow Paper. https://gavwood.com/paper.pdf.

[93] Gavin Wood. 2015. Proof-of-authority Private Chains. (2015). https://github.com/ethereum/guide/blob/master/poa.md.

[94] Liang Zhang, Haibin Kan, Yang Xu, and Jinhao Ran. 2021. Revocable data sharing methodology based on SGX and blockchain. In *Proceedings of the 15th International Conference on Network and System Security (Lecture Notes in Computer Science 13041)*, Springer, 61–78.

[95] Ning Zhang and Qi Shi. 1996. Achieving non-repudiation of receipt. *Comput. J.* 39, 10 (1996), 844–853.

[96] Zibin Zheng, Shaoan Xie, Hongning Dai, Xiangping Chen, and Huaimin Wang. 2017. An overview of blockchain technology: Architecture, consensus, and future trends. In *Proceedings of the 2017 International Congress on Big Data*. IEEE, 557–564.

[97] Jianying Zhou and Dieter Gollmann. 1996. Observations on non-repudiation. In *Proceedings of the 1996 International Conference on the Theory and Application of Cryptology and Information Security (Lecture Notes in Computer Science 1163)*, Springer, 133–144.

[98] Valentin Zieglmeier. 2023. The inverse transparency toolchain. (2023). Manuscript in review.

[99] Valentin Zieglmeier and Gabriel Loyola Daiqui. 2021. GDPR-compliant use of blockchain for secure usage logs. In *Proceedings of the 25th International Conference on Evaluation and Assessment in Software Engineering*. ACM, 313–320.

[100] Valentin Zieglmeier, Maren Gierlich-Joas, and Alexander Pretschner. 2022. Increasing employees' willingness to share: Introducing appeal strategies for people analytics. In *Proceedings of the 13th International Conference on Software Business (Lecture Notes in Business Information Processing 463)*, Springer, 213–226.

[101] Valentin Zieglmeier and Alexander Pretschner. 2021. Trustworthy Transparency by Design. (2021). arxiv:cs.SE/2103.10769

[102] Valentin Zieglmeier and Alexander Pretschner. 2023. Rethinking people analytics with inverse transparency by design. *Proceedings of the ACM on Human-Computer Interaction* 7 (2023). Forthcoming.

[103] Guy Zyskind, Oz Nathan, and Alex 'Sandy' Pentland. 2015. Decentralizing privacy: Using blockchain to protect personal data. In *Proceedings of the 2015 IEEE Security and Privacy Workshops*. IEEE, 180–184.

Received 5 December 2022; revised 18 February 2023; accepted 20 March 2023

6. Designing Trustworthy User Interfaces

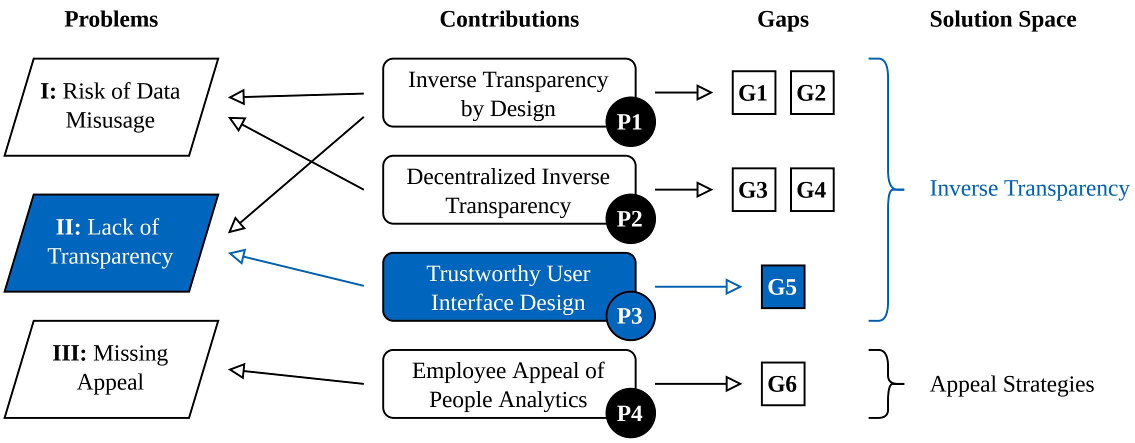


Figure 6.1.: Relationship of our third contribution to the big picture.

Summary. The following summary is partially adapted from our paper [see 3].

- *Problem:* The *lack of transparency* (problem **II**) in PA can be improved by a technical infrastructure that collects trusted transparency information (see Chapter 4), but that does not suffice. The user interface that provides transparency needs to be trusted by individuals as well, as this influences their intention to use it.

- *Solution:* We need to understand which trustworthiness factors in software user interfaces exist, and which concrete measures can increase their trustworthiness.

- *Gap:* Existing research on user trust does not present concrete measures to improve user interface trustworthiness or does not follow a systematic approach (**G5**).

- *Contribution:* Based on a systematic literature review, we contribute a taxonomy of trustworthiness in software user interfaces and concrete measures to increase user interface trustworthiness. The taxonomy is our main contribution. It synthesizes the results of a large number of empirical studies and theoretical conceptualizations. As a smaller contribution, we additionally present a preliminary evaluation to test the applicability of the measures to our transparency dashboard. We find that they can be effective in fostering trust in users.

- *Limitations:* The focus of this paper is on the taxonomy and its summarization of existing knowledge. Previous works have performed extensive empirical evaluations to validate the constructs we summarize. Still, we furthermore performed a small evaluation for our context, which we acknowledge to be preliminary. It exhibits two main limitations. *First*, the study only included twelve participants. *Second*, we followed a within-subject design with no order variation. Therefore, the results cannot be meaningfully generalized. They should be seen in the context of related works that evaluated some of the implemented factors previously.

Author Contributions. VZ developed the initial research idea. Under the guidance of VZ, AML performed the initial systematic literature review. VZ embedded the results in a framework with an additional exploratory literature review. Based on the findings, VZ developed the proof of concept implementation. AML conducted the study in close discussion with VZ. Finally, VZ conceived of the framing of the work and wrote the manuscript.

Designing Trustworthy User Interfaces

Valentin Zieglmeier
Technical University of Munich
Munich, Germany
valentin.zieglmeier@tum.de

Antonia Maria Lehene
Technical University of Munich
Munich, Germany
antonia.lehene@tum.de

ABSTRACT

Interface design can directly influence trustworthiness of a software. Thereby, it affects users' intention to use a tool. Previous research on user trust has not comprehensively addressed user interface design, though. We lack an understanding of what makes interfaces trustworthy (1), as well as actionable measures to improve trustworthiness (2).

We contribute to this by addressing both gaps. Based on a systematic literature review, we give a thorough overview over the theory on user trust and provide a taxonomy of factors influencing user interface trustworthiness. Then, we derive concrete measures to address these factors in interface design. We use the results to create a proof of concept interface. In a preliminary evaluation, we compare a variant designed to elicit trust with one designed to reduce it. Our results show that the measures we apply can be effective in fostering trust in users.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; **Graphical user interfaces**; User studies.

KEYWORDS

User trust, User-centered design, Taxonomy, Systematic literature review, Proof of concept

ACM Reference Format:
Valentin Zieglmeier and Antonia Maria Lehene. 2021. Designing Trustworthy User Interfaces. In *33rd Australian Conference on Human-Computer Interaction (OzCHI '21), November 30–December 2, 2021, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3520495.3520525

1 INTRODUCTION

When creating software tools with user-facing interfaces, one of the central aspects to consider is how they are designed. As this is the component that users are directly exposed to, it can shape their impression of the tool and willingness to use it. How trustworthy the software appears at first glance may play an important role in this. User trust has been shown to influence users' intention to use a software [50]. Furthermore, central trust antecedents, such as credibility, increase intention to use as well [66]. Previous research suggests that trust constructs may even have a higher influence on intention to use than some usability aspects [87]. Beyond the actual service or underlying implementation, the design of the interface has been shown to play an important role in the trustworthiness of a software tool [72]. Yet, to the best of our knowledge, no comprehensive overview over the relevant trust constructs and how they can be concretely addressed with the design of user interfaces exists today.

Our goal is to derive how (initial) user trust can be achieved and improved through user interface design. For this purpose, general influences on user trust for various usage contexts are summarized. Additionally, we analyze the interactions of these factors, as well as how they can be implemented in user interfaces in order to initiate user trust. In a preliminary empirical study, we assess whether a proof of concept design developed according to our findings can in fact increase user trust as we expect it.

Therefore, this work contributes a theoretical overview over user trust formation and factors in software design, summarizes actionable design variants to improve the trustworthiness of a software tool through its user interface, and provides preliminary evaluation results on the effectiveness of the developed software design variants.

2 SYSTEMATIC LITERATURE REVIEW

From literature reviews in the field of user experience and trust [1, 34, 35, 39, 41], we derived central terms used in research on trust in automation. These terms were used to build queries for a systematic literature review covering Scopus and Web of Science. This meant we included works that focus on individual factors related to trust. For each term, we built a query of the form `("user trust") AND (<term> OR <synonyms>) AND NOT ("social media" OR "blockchain")`. For the search terms and synonyms we used, refer to Table 1. We explicitly excluded works with the terms "social media" and "blockchain", as we found that these often consider users' trust in each other, rather than in the respective tool. The search covered title, abstract, and keywords. We limited the results to English language journal articles and conference papers in the area of computer science.

In total, 697 works were found. Through a title and abstract review, we filtered for papers related to our research. This left us with 162 remaining works. These, plus 40 works that we added through snowballing, were read and analyzed. The following sections summarize the results from the most relevant of these works.

3 FUNDAMENTALS: USER TRUST

To better understand our findings, it is important to grasp what the concept of user trust encompasses. Furthermore, it is instructive to consider how trust is built and maintained, as the trust relationship is dynamic. User trust can be defined using different approaches.

OzCHI '21, November 30-December 2, 2021, Melbourne, VIC, Australia

ACM ISBN 978-1-4503-9598-4/21/11…$15.00
https://doi.org/10.1145/3520495.3520525

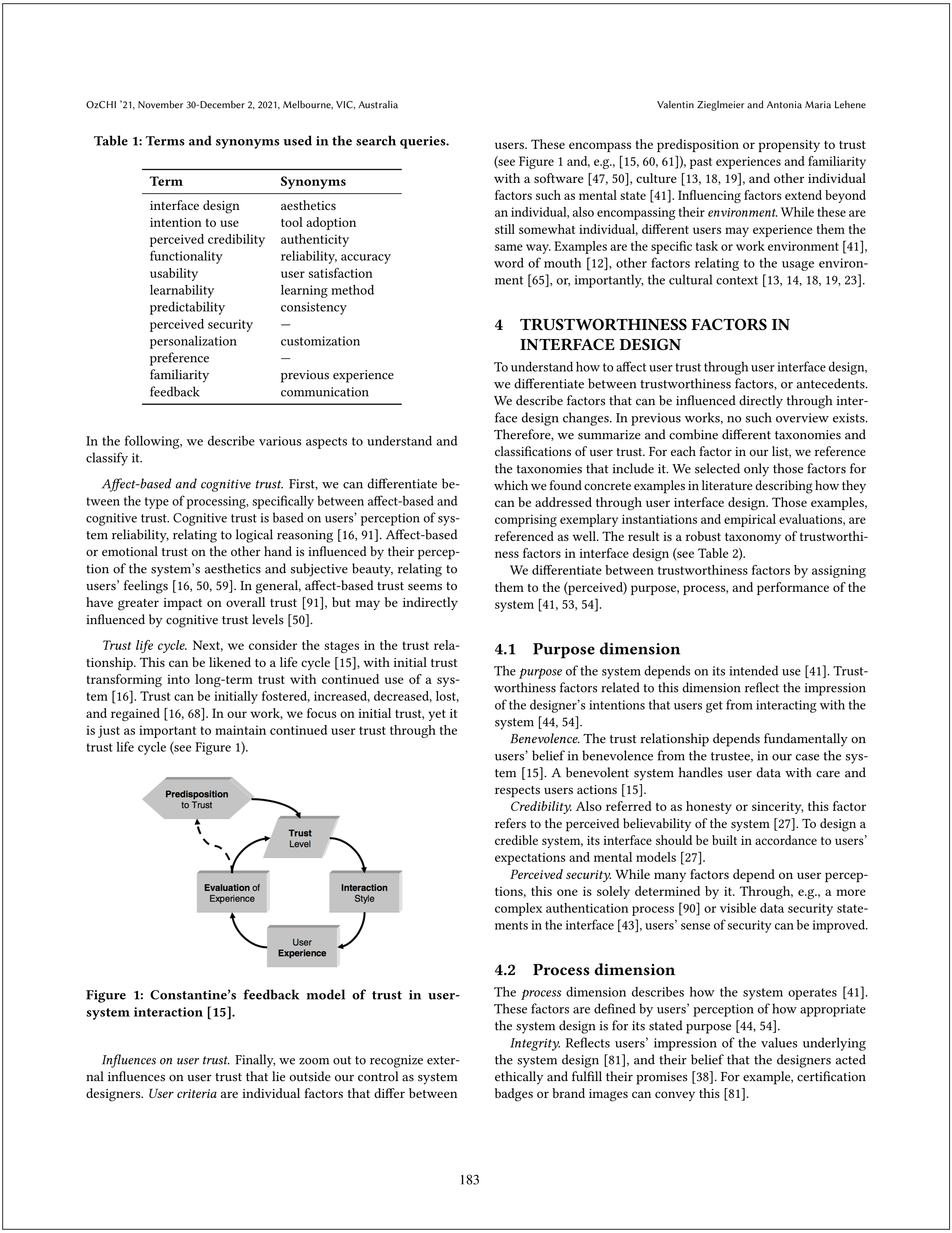


Table 1: Terms and synonyms used in the search queries.

Term	Synonyms
interface design	aesthetics
intention to use	tool adoption
perceived credibility	authenticity
functionality	reliability, accuracy
usability	user satisfaction
learnability	learning method
predictability	consistency
perceived security	—
personalization	customization
preference	—
familiarity	previous experience
feedback	communication

In the following, we describe various aspects to understand and classify it.

Affect-based and cognitive trust. First, we can differentiate between the type of processing, specifically between affect-based and cognitive trust. Cognitive trust is based on users' perception of system reliability, relating to logical reasoning [16, 91]. Affect-based or emotional trust on the other hand is influenced by their perception of the system's aesthetics and subjective beauty, relating to users' feelings [16, 50, 59]. In general, affect-based trust seems to have greater impact on overall trust [91], but may be indirectly influenced by cognitive trust levels [50].

Trust life cycle. Next, we consider the stages in the trust relationship. This can be likened to a life cycle [15], with initial trust transforming into long-term trust with continued use of a system [16]. Trust can be initially fostered, increased, decreased, lost, and regained [16, 68]. In our work, we focus on initial trust, yet it is just as important to maintain continued user trust through the trust life cycle (see Figure 1).

Figure 1: Constantine's feedback model of trust in user-system interaction [15].

Influences on user trust. Finally, we zoom out to recognize external influences on user trust that lie outside our control as system designers. *User criteria* are individual factors that differ between users. These encompass the predisposition or propensity to trust (see Figure 1 and, e.g., [15, 60, 61]), past experiences and familiarity with a software [47, 50], culture [13, 18, 19], and other individual factors such as mental state [41]. Influencing factors extend beyond an individual, also encompassing their *environment.* While these are still somewhat individual, different users may experience them the same way. Examples are the specific task or work environment [41], word of mouth [12], other factors relating to the usage environment [65], or, importantly, the cultural context [13, 14, 18, 19, 23].

4 TRUSTWORTHINESS FACTORS IN INTERFACE DESIGN

To understand how to affect user trust through user interface design, we differentiate between trustworthiness factors, or antecedents. We describe factors that can be influenced directly through interface design changes. In previous works, no such overview exists. Therefore, we summarize and combine different taxonomies and classifications of user trust. For each factor in our list, we reference the taxonomies that include it. We selected only those factors for which we found concrete examples in literature describing how they can be addressed through user interface design. Those examples, comprising exemplary instantiations and empirical evaluations, are referenced as well. The result is a robust taxonomy of trustworthiness factors in interface design (see Table 2).

We differentiate between trustworthiness factors by assigning them to the (perceived) purpose, process, and performance of the system [41, 53, 54].

4.1 Purpose dimension

The *purpose* of the system depends on its intended use [41]. Trustworthiness factors related to this dimension reflect the impression of the designer's intentions that users get from interacting with the system [44, 54].

Benevolence. The trust relationship depends fundamentally on users' belief in benevolence from the trustee, in our case the system [15]. A benevolent system handles user data with care and respects users actions [15].

Credibility. Also referred to as honesty or sincerity, this factor refers to the perceived believability of the system [27]. To design a credible system, its interface should be built in accordance to users' expectations and mental models [27].

Perceived security. While many factors depend on user perceptions, this one is solely determined by it. Through, e.g., a more complex authentication process [90] or visible data security statements in the interface [43], users' sense of security can be improved.

4.2 Process dimension

The *process* dimension describes how the system operates [41]. These factors are defined by users' perception of how appropriate the system design is for its stated purpose [44, 54].

Integrity. Reflects users' impression of the values underlying the system design [81], and their belief that the designers acted ethically and fulfill their promises [38]. For example, certification badges or brand images can convey this [81].

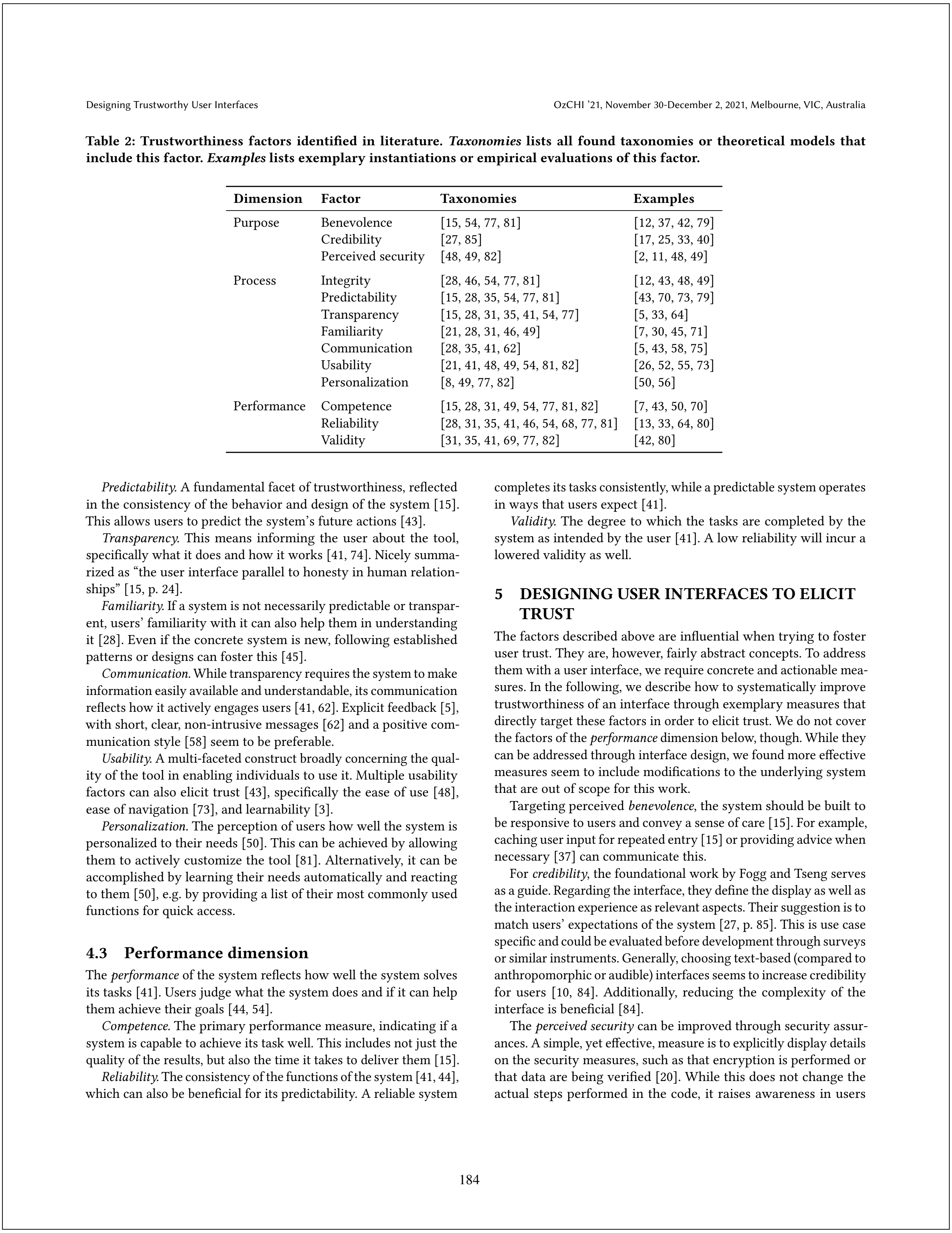


Table 2: Trustworthiness factors identified in literature. *Taxonomies* lists all found taxonomies or theoretical models that include this factor. *Examples* lists exemplary instantiations or empirical evaluations of this factor.

Dimension	Factor	Taxonomies	Examples
Purpose	Benevolence	[15, 54, 77, 81]	[12, 37, 42, 79]
	Credibility	[27, 85]	[17, 25, 33, 40]
	Perceived security	[48, 49, 82]	[2, 11, 48, 49]
Process	Integrity	[28, 46, 54, 77, 81]	[12, 43, 48, 49]
	Predictability	[15, 28, 35, 54, 77, 81]	[43, 70, 73, 79]
	Transparency	[15, 28, 31, 35, 41, 54, 77]	[5, 33, 64]
	Familiarity	[21, 28, 31, 46, 49]	[7, 30, 45, 71]
	Communication	[28, 35, 41, 62]	[5, 43, 58, 75]
	Usability	[21, 41, 48, 49, 54, 81, 82]	[26, 52, 55, 73]
	Personalization	[8, 49, 77, 82]	[50, 56]
Performance	Competence	[15, 28, 31, 49, 54, 77, 81, 82]	[7, 43, 50, 70]
	Reliability	[28, 31, 35, 41, 46, 54, 68, 77, 81]	[13, 33, 64, 80]
	Validity	[31, 35, 41, 69, 77, 82]	[42, 80]

Predictability. A fundamental facet of trustworthiness, reflected in the consistency of the behavior and design of the system [15]. This allows users to predict the system's future actions [43].

Transparency. This means informing the user about the tool, specifically what it does and how it works [41, 74]. Nicely summarized as "the user interface parallel to honesty in human relationships" [15, p. 24].

Familiarity. If a system is not necessarily predictable or transparent, users' familiarity with it can also help them in understanding it [28]. Even if the concrete system is new, following established patterns or designs can foster this [45].

Communication. While transparency requires the system to make information easily available and understandable, its communication reflects how it actively engages users [41, 62]. Explicit feedback [5], with short, clear, non-intrusive messages [62] and a positive communication style [58] seem to be preferable.

Usability. A multi-faceted construct broadly concerning the quality of the tool in enabling individuals to use it. Multiple usability factors can also elicit trust [43], specifically the ease of use [48], ease of navigation [73], and learnability [3].

Personalization. The perception of users how well the system is personalized to their needs [50]. This can be achieved by allowing them to actively customize the tool [81]. Alternatively, it can be accomplished by learning their needs automatically and reacting to them [50], e.g. by providing a list of their most commonly used functions for quick access.

4.3 Performance dimension

The *performance* of the system reflects how well the system solves its tasks [41]. Users judge what the system does and if it can help them achieve their goals [44, 54].

Competence. The primary performance measure, indicating if a system is capable to achieve its task well. This includes not just the quality of the results, but also the time it takes to deliver them [15].

Reliability. The consistency of the functions of the system [41, 44], which can also be beneficial for its predictability. A reliable system completes its tasks consistently, while a predictable system operates in ways that users expect [41].

Validity. The degree to which the tasks are completed by the system as intended by the user [41]. A low reliability will incur a lowered validity as well.

5 DESIGNING USER INTERFACES TO ELICIT TRUST

The factors described above are influential when trying to foster user trust. They are, however, fairly abstract concepts. To address them with a user interface, we require concrete and actionable measures. In the following, we describe how to systematically improve trustworthiness of an interface through exemplary measures that directly target these factors in order to elicit trust. We do not cover the factors of the *performance* dimension below, though. While they can be addressed through interface design, we found more effective measures seem to include modifications to the underlying system that are out of scope for this work.

Targeting perceived *benevolence*, the system should be built to be responsive to users and convey a sense of care [15]. For example, caching user input for repeated entry [15] or providing advice when necessary [37] can communicate this.

For *credibility*, the foundational work by Fogg and Tseng serves as a guide. Regarding the interface, they define the display as well as the interaction experience as relevant aspects. Their suggestion is to match users' expectations of the system [27, p. 85]. This is use case specific and could be evaluated before development through surveys or similar instruments. Generally, choosing text-based (compared to anthropomorphic or audible) interfaces seems to increase credibility for users [10, 84]. Additionally, reducing the complexity of the interface is beneficial [84].

The *perceived security* can be improved through security assurances. A simple, yet effective, measure is to explicitly display details on the security measures, such as that encryption is performed or that data are being verified [20]. While this does not change the actual steps performed in the code, it raises awareness in users which increases their perceived security [20]. Furthermore, forcing users to re-login after a certain period [90] and informing users that unauthorized accesses are blocked [63, 90] can improve this. These findings can be summarized as actively making users aware of the (ostensible) security measures that are implemented.

For *integrity*, the interface ideally conveys the ethics of the designers. This can be achieved by adding e.g. certification badges or brand images [81] that suggest an ethics code or value system.

The *predictability* of the interface can be achieved following the guidelines set forth by Gram. They suggest deterministic design that maps the observable state directly to system events, with the system providing users with information about its state and the actions they can take. Furthermore, they describe completeness and consistency of information display as important [32, p. 296].

Increasing the *transparency* that users experience can be a vital aspect. As a basic measure, explanatory texts should make clear what functionality exists [9]. Especially for safety-relevant scenarios, the system should also be transparent about the risks and limitations of its functions [17, 51, 70].

Evoking *familiarity* is of course very dependent on the users' previous experience. Still, just by following established design patterns and metaphors, users' familiarity with them can be evoked [45].

Regarding *communication*, short and clear notifications about the status of the system serve to inform users [23, 81]. When designing these messages, etiquette are important [54, 67, 89]. These can be defined as being non-interrupting and patient [67], as well as having a positive tone [36].

The *usability* of a user interface is a more complex topic and its own branch of research. Yet, from works addressing usability as an antecedent of trust specifically, we can derive some concrete suggestions. First, ease of navigation and user guidance are beneficial [23, 73, 76]. Similarly, consistency in design and color schemes improves usability and trustworthiness [19, 23, 81]. For non-intuitive interfaces, learnability was found to be effective [6, 73, 76]. This can mean giving users the opportunity to learn about the functions of the system and encouraging them to explore it [6]. If training is required, tools can directly embed tutorials to ease discovery [83]. Additionally, the ease of use and subjective appeal of the interface can be relevant [29, 78, 86]. Beyond generally aiming to reduce the required cognitive effort, this can mean improving the reaction speed [86], reducing clutter and animations [29], and designing a layout with, e.g., high classical aesthetic appeal [78]. Finally, attractive as well as readable typography, covering font choice and text size, are also facets to consider [23, 24].

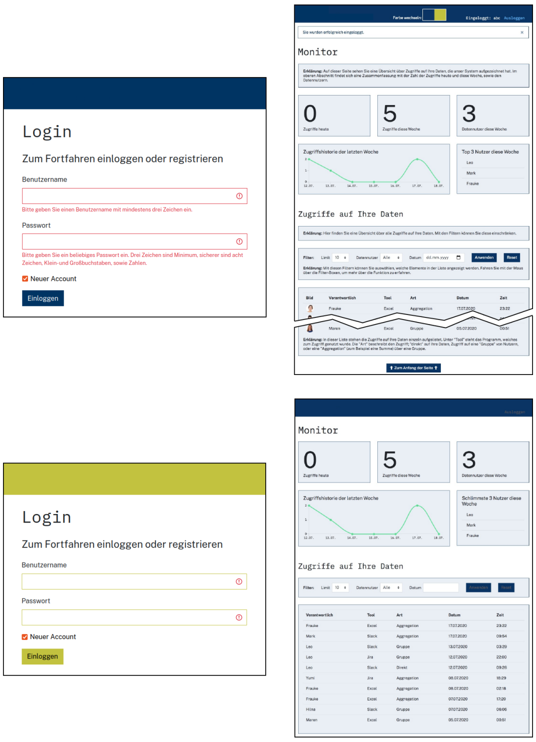
For *personalization*, various measures can be effective. Allowing customization of the interface to match the user needs [81] is a sensible step, while more advanced systems might try to predict user wishes [50].

6 PRELIMINARY EVALUATION

To assess the effectiveness of influencing trust through user interface design, we created a proof of concept interface for a preliminary evaluation. We developed two user interface variants: One variant aimed to follow our recommendations to elicit trust (variant A), the other explicitly disregarded our findings (variant B). Twelve people participated in our study (ten students, two apprentices; eight female, four male). Each participant was asked to assess both variants. They had to register, then they were led to variant A. After exploring it and answering the questionnaire, they were then shown variant B. Finally, they answered the same questionnaire again to allow us to compare the results. Their responses were recorded on a seven-point Likert scale.

6.1 Design variants

We created two interface variants for participants to explore (see Figure 2). Both followed the same basic structure: They started with a login page where users had to authenticate, followed by the main page in form of a dashboard listing various data. The dashboard was designed for the use case of logging data usages [see, e.g., 4, 92], showing how data of the individual were accessed and by whom.



Bitte geben Sie einen Benutzername mit mindestens drei Zeichen ein.

Bitte geben Sie ein beliebiges Passwort ein. Drei Zeichen sind Minimum, sicherer sind acht Zeichen, Klein-und Großbuchstaben, sowie Zahlen.

☑ Neuer Account

Sie wurden erfolgreich eingeloggt. ×

Erklärung: Auf dieser Seite sehen Sie eine Übersicht über Zugriffe auf Ihre Daten, die unser System aufgezeichnet hat. Im oberen Abschnitt findet sich eine Zusammenfassung mit der Zahl der Zugriffe heute und diese Woche, sowie den Datennutzern.

0 Zugriffe heute — 5 Zugriffe diese Woche — 3 Datennutzer diese Woche

Top 3 Nutzer diese Woche: Leo, Mark, Frauke

Erklärung: Hier finden Sie eine Übersicht über alle Zugriffe auf Ihre Daten. Mit den Filtern können Sie diese einschränken.

Filter: Limit 10 Datennutzer Alle Datum dd.mm.yyyy

Erklärung: Mit diesen Filtern können Sie auswählen, welche Elemente in der Liste angezeigt werden. Fahren Sie mit der Maus über die Filter-Boxen, um mehr über die Funktion zu erfahren.

Bild	Verantwortlich	Tool	Art	Datum	Zeit
	Frauke	Excel	Aggregation	17.07.2020	23:22
	Maren	Excel	Gruppe	05.07.2020	00:51

Erklärung: In dieser Liste stehen die Zugriffe auf Ihre Daten einzeln aufgelistet. Unter "Tool" steht das Programm, welches zum Zugriff genutzt wurde. Die "Art" beschreibt den Zugriff; "direkt" auf Ihre Daten, Zugriff auf eine "Gruppe" von Nutzern, oder eine "Aggregation" (zum Beispiel eine Summe) über eine Gruppe.

Schlimmste 3 Nutzer diese Woche: Leo, Mark, Frauke

Filter: Limit 10 Datennutzer Alle Datum

Verantwortlich	Tool	Art	Datum	Zeit
Frauke	Excel	Aggregation	17.07.2020	23:22
Mark	Slack	Aggregation	17.07.2020	09:54
Leo	Slack	Gruppe	13.07.2020	03:29
Leo	Jira	Gruppe	12.07.2020	22:00
Leo	Slack	Direkt	12.07.2020	09:26
Yumi	Jira	Aggregation	08.07.2020	18:29
Frauke	Excel	Aggregation	08.07.2020	02:18
Frauke	Excel	Aggregation	07.07.2020	17:20
Hilna	Slack	Gruppe	07.07.2020	06:06
Maren	Excel	Gruppe	05.07.2020	00:51

Figure 2: Screenshots of the two interface design variants. Variant A (above) was developed to elicit trust, variant B (below) to reduce it.

To target personalization, variant A offered a color scheme picker, while variant B did not. For transparency and communication, a message was displayed after successful login, with a status bar showing the logged-in user. Both were missing from variant B. A more positive communication style for variant A meant that it listed the "best" users, instead of the "worst", when showing who accessed the data most frequently. Improving predictability, a detailed description of available functions was shown only for variant A. Also, the same color scheme was used for the login page and dashboard, whereas for variant B it switched from yellow to blue. Finally, we targeted usability. To improve ease of navigation, variant A had a button at the bottom to return "back to top" that B did not have. Additionally, the "Log out" button on variant A was made easier to notice with brighter text. For ease of use and subjective appeal, the colors of text on buttons for variant B in general were made less legible compared to A, with a reduced contrast to the background color. Furthermore, a semaphore informing users how strong the password they selected was and if it was valid was shown on variant A, but not on B.

6.2 Questionnaire

Participants answered ten questions related to our design changes, with their responses being recorded on a seven-point Likert scale. Lower values on the scale indicate lower levels of agreement. We used questions developed and validated in previous works to evaluate trust.

Covering the induced familiarity and predictability, we asked:

- FP: I am familiar with how the system works [from 50].

For transparency, learnability, and communication, we asked:

- TLC: I find the system easy to learn to use [from 23].

Regarding their perceived security, we asked:

- PS: I think the authentication is very secure, that is, it protects me against attacks [from 93].

Addressing usability directly, we asked:

- U1: I find the system easy to use [from 17].
- U2: I can find easily what I am looking for on the interface [from 73].

To assess their faith in the system generally, we asked:

- F1: I believe advice from the system even when I don't know for certain that it is correct [from 57].
- F2: When I am uncertain about a decision I believe the system rather than myself [from 57].
- F3: If I am not sure about a decision, I have faith that the system will provide the best solution [from 57].

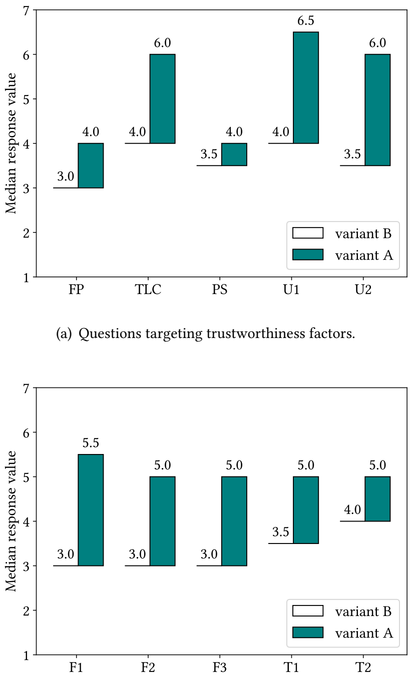
In addition, we asked participants about their trust and the perceived trustworthiness of the system directly:

- T1: I trust the system [from 17, 59].
- T2: I believe the system to be trustworthy [from 17, 22, 33].

6.3 Results

We analyze the results by comparing the median value of all responses for variant A (eliciting trust) to variant B. We use the median instead of averaging the response values, as that can be problematic for ordinal data such as Likert responses. It prevents outliers from affecting the result while still allowing us to summarize the responses in a single value. For visualization, we use hat graphs [88].

For most trustworthiness factors, we find a clear increase in the median level of agreement (see Figure 3(a)). On average, the median increased by 1.7 points, with the lowest increase (0.5) for question PS and the highest (2.5) for questions U1 and U2. This suggests that our changes had the greatest impact on usability. Considering questions FP and TLC, we find an interesting difference: participants seemingly did not understand how the system worked when they first used it, but may have found they could learn to do so. The low difference for question PS seems plausible, as our modifications did not explicitly target perceived security and both variants used the same authentication process.

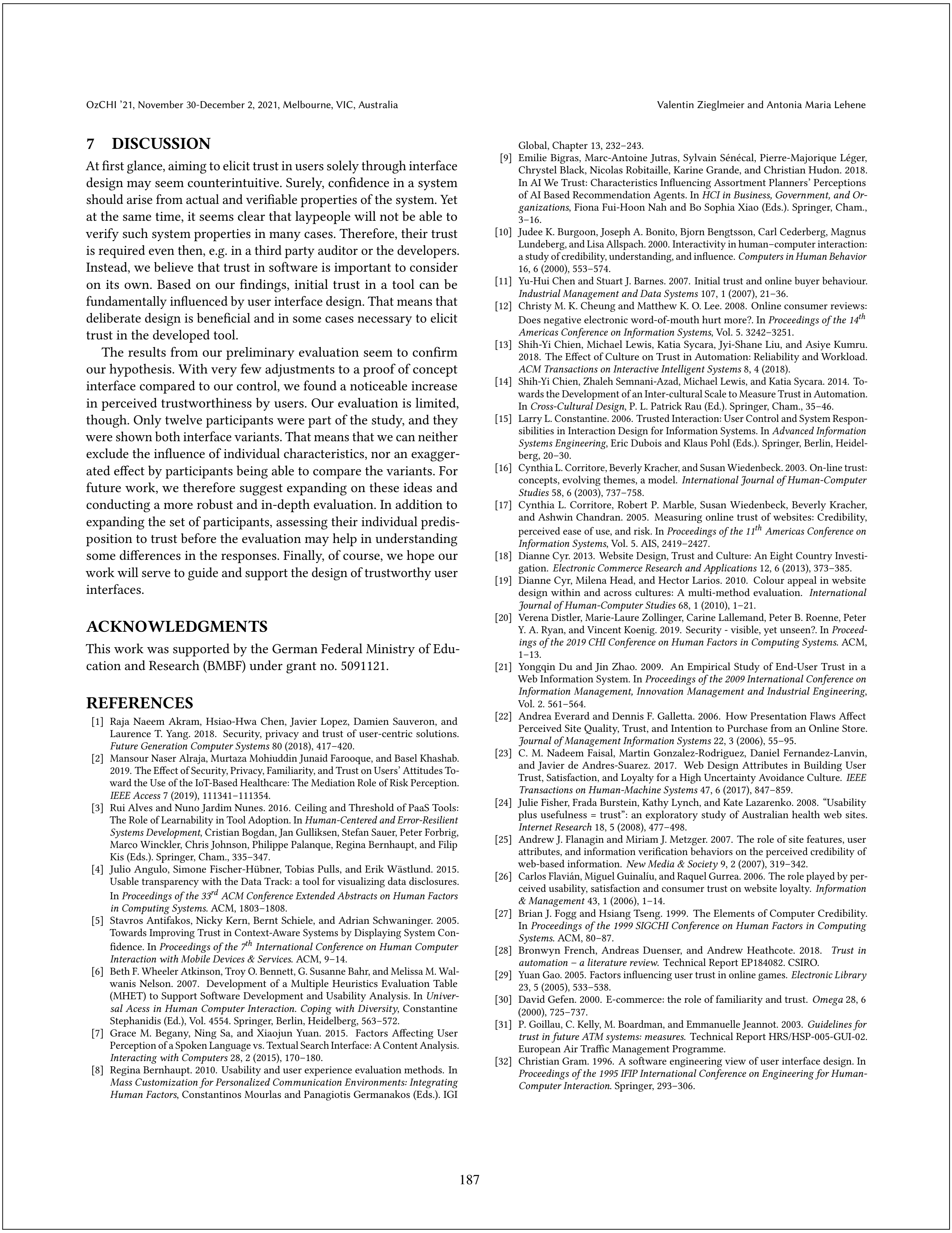


(a) Questions targeting trustworthiness factors.

(b) Questions targeting faith and trust broadly.

Figure 3: Questionnaire results ($n = 12$). The median of all responses for variant A (trust-eliciting) and variant B is compared.

Considering the overarching goal of trust and the related construct of faith, we also find noticeable increases in the median level of agreement (see Figure 3(b)). The average increase is 1.8 points, with the lowest (1) for question T2 and the highest (2.5) for question F1. For variant A, we can see that the median is stable at 5, except for F1 with 5.5. This means that participants did not overwhelmingly agree with the questions, but showed a clear tendency towards agreement. The lower increase for question T2 is not due to variant A being less trustworthy, but due to variant B seemingly also eliciting some trust in participants. In all, this suggests that our changes had the impact we aimed for.



7 DISCUSSION

At first glance, aiming to elicit trust in users solely through interface design may seem counterintuitive. Surely, confidence in a system should arise from actual and verifiable properties of the system. Yet at the same time, it seems clear that laypeople will not be able to verify such system properties in many cases. Therefore, their trust is required even then, e.g. in a third party auditor or the developers. Instead, we believe that trust in software is important to consider on its own. Based on our findings, initial trust in a tool can be fundamentally influenced by user interface design. That means that deliberate design is beneficial and in some cases necessary to elicit trust in the developed tool.

The results from our preliminary evaluation seem to confirm our hypothesis. With very few adjustments to a proof of concept interface compared to our control, we found a noticeable increase in perceived trustworthiness by users. Our evaluation is limited, though. Only twelve participants were part of the study, and they were shown both interface variants. That means that we can neither exclude the influence of individual characteristics, nor an exaggerated effect by participants being able to compare the variants. For future work, we therefore suggest expanding on these ideas and conducting a more robust and in-depth evaluation. In addition to expanding the set of participants, assessing their individual predisposition to trust before the evaluation may help in understanding some differences in the responses. Finally, of course, we hope our work will serve to guide and support the design of trustworthy user interfaces.

ACKNOWLEDGMENTS

This work was supported by the German Federal Ministry of Education and Research (BMBF) under grant no. 5091121.

REFERENCES

[1] Raja Naeem Akram, Hsiao-Hwa Chen, Javier Lopez, Damien Sauveron, and Laurence T. Yang. 2018. Security, privacy and trust of user-centric solutions. *Future Generation Computer Systems* 80 (2018), 417–420.

[2] Mansour Naser Alraja, Murtaza Mohiuddin Junaid Farooque, and Basel Khashab. 2019. The Effect of Security, Privacy, Familiarity, and Trust on Users' Attitudes Toward the Use of the IoT-Based Healthcare: The Mediation Role of Risk Perception. *IEEE Access* 7 (2019), 111341–111354.

[3] Rui Alves and Nuno Jardim Nunes. 2016. Ceiling and Threshold of PaaS Tools: The Role of Learnability in Tool Adoption. In *Human-Centered and Error-Resilient Systems Development*, Cristian Bogdan, Jan Gulliksen, Stefan Sauer, Peter Forbrig, Marco Winckler, Chris Johnson, Philippe Palanque, Regina Bernhaupt, and Filip Kis (Eds.). Springer, Cham., 335–347.

[4] Julio Angulo, Simone Fischer-Hübner, Tobias Pulls, and Erik Wästlund. 2015. Usable transparency with the Data Track: a tool for visualizing data disclosures. In *Proceedings of the 33rd ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 1803–1808.

[5] Stavros Antifakos, Nicky Kern, Bernt Schiele, and Adrian Schwaninger. 2005. Towards Improving Trust in Context-Aware Systems by Displaying System Confidence. In *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services*. ACM, 9–14.

[6] Beth F. Wheeler Atkinson, Troy O. Bennett, G. Susanne Bahr, and Melissa M. Walwanis Nelson. 2007. Development of a Multiple Heuristics Evaluation Table (MHET) to Support Software Development and Usability Analysis. In *Universal Acess in Human Computer Interaction. Coping with Diversity*, Constantine Stephanidis (Ed.), Vol. 4554. Springer, Berlin, Heidelberg, 563–572.

[7] Grace M. Begany, Ning Sa, and Xiaojun Yuan. 2015. Factors Affecting User Perception of a Spoken Language vs. Textual Search Interface: A Content Analysis. *Interacting with Computers* 28, 2 (2015), 170–180.

[8] Regina Bernhaupt. 2010. Usability and user experience evaluation methods. In *Mass Customization for Personalized Communication Environments: Integrating Human Factors*, Constantinos Mourlas and Panagiotis Germanakos (Eds.). IGI Global, Chapter 13, 232–243.

[9] Emilie Bigras, Marc-Antoine Jutras, Sylvain Sénécal, Pierre-Majorique Léger, Chrystel Black, Nicolas Robitaille, Karine Grande, and Christian Hudon. 2018. In AI We Trust: Characteristics Influencing Assortment Planners' Perceptions of AI Based Recommendation Agents. In *HCI in Business, Government, and Organizations*, Fiona Fui-Hoon Nah and Bo Sophia Xiao (Eds.). Springer, Cham., 3–16.

[10] Judee K. Burgoon, Joseph A. Bonito, Bjorn Bengtsson, Carl Cederberg, Magnus Lundeberg, and Lisa Allspach. 2000. Interactivity in human–computer interaction: a study of credibility, understanding, and influence. *Computers in Human Behavior* 16, 6 (2000), 553–574.

[11] Yu-Hui Chen and Stuart J. Barnes. 2007. Initial trust and online buyer behaviour. *Industrial Management and Data Systems* 107, 1 (2007), 21–36.

[12] Christy M. K. Cheung and Matthew K. O. Lee. 2008. Online consumer reviews: Does negative electronic word-of-mouth hurt more?. In *Proceedings of the 14th Americas Conference on Information Systems*, Vol. 5. 3242–3251.

[13] Shih-Yi Chien, Michael Lewis, Katia Sycara, Jyi-Shane Liu, and Asiye Kumru. 2018. The Effect of Culture on Trust in Automation: Reliability and Workload. *ACM Transactions on Interactive Intelligent Systems* 8, 4 (2018).

[14] Shih-Yi Chien, Zhaleh Semnani-Azad, Michael Lewis, and Katia Sycara. 2014. Towards the Development of an Inter-cultural Scale to Measure Trust in Automation. In *Cross-Cultural Design*, P. L. Patrick Rau (Ed.). Springer, Cham., 35–46.

[15] Larry L. Constantine. 2006. Trusted Interaction: User Control and System Responsibilities in Interaction Design for Information Systems. In *Advanced Information Systems Engineering*, Eric Dubois and Klaus Pohl (Eds.). Springer, Berlin, Heidelberg, 20–30.

[16] Cynthia L. Corritore, Beverly Kracher, and Susan Wiedenbeck. 2003. On-line trust: concepts, evolving themes, a model. *International Journal of Human-Computer Studies* 58, 6 (2003), 737–758.

[17] Cynthia L. Corritore, Robert P. Marble, Susan Wiedenbeck, Beverly Kracher, and Ashwin Chandran. 2005. Measuring online trust of websites: Credibility, perceived ease of use, and risk. In *Proceedings of the 11th Americas Conference on Information Systems*, Vol. 5. AIS, 2419–2427.

[18] Dianne Cyr. 2013. Website Design, Trust and Culture: An Eight Country Investigation. *Electronic Commerce Research and Applications* 12, 6 (2013), 373–385.

[19] Dianne Cyr, Milena Head, and Hector Larios. 2010. Colour appeal in website design within and across cultures: A multi-method evaluation. *International Journal of Human-Computer Studies* 68, 1 (2010), 1–21.

[20] Verena Distler, Marie-Laure Zollinger, Carine Lallemand, Peter B. Roenne, Peter Y. A. Ryan, and Vincent Koenig. 2019. Security - visible, yet unseen?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 1–13.

[21] Yongqin Du and Jin Zhao. 2009. An Empirical Study of End-User Trust in a Web Information System. In *Proceedings of the 2009 International Conference on Information Management, Innovation Management and Industrial Engineering*, Vol. 2. 561–564.

[22] Andrea Everard and Dennis F. Galletta. 2006. How Presentation Flaws Affect Perceived Site Quality, Trust, and Intention to Purchase from an Online Store. *Journal of Management Information Systems* 22, 3 (2006), 55–95.

[23] C. M. Nadeem Faisal, Martin Gonzalez-Rodriguez, Daniel Fernandez-Lanvin, and Javier de Andres-Suarez. 2017. Web Design Attributes in Building User Trust, Satisfaction, and Loyalty for a High Uncertainty Avoidance Culture. *IEEE Transactions on Human-Machine Systems* 47, 6 (2017), 847–859.

[24] Julie Fisher, Frada Burstein, Kathy Lynch, and Kate Lazarenko. 2008. "Usability plus usefulness = trust": an exploratory study of Australian health web sites. *Internet Research* 18, 5 (2008), 477–498.

[25] Andrew J. Flanagin and Miriam J. Metzger. 2007. The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media & Society* 9, 2 (2007), 319–342.

[26] Carlos Flavián, Miguel Guinalíu, and Raquel Gurrea. 2006. The role played by perceived usability, satisfaction and consumer trust on website loyalty. *Information & Management* 43, 1 (2006), 1–14.

[27] Brian J. Fogg and Hsiang Tseng. 1999. The Elements of Computer Credibility. In *Proceedings of the 1999 SIGCHI Conference on Human Factors in Computing Systems*. ACM, 80–87.

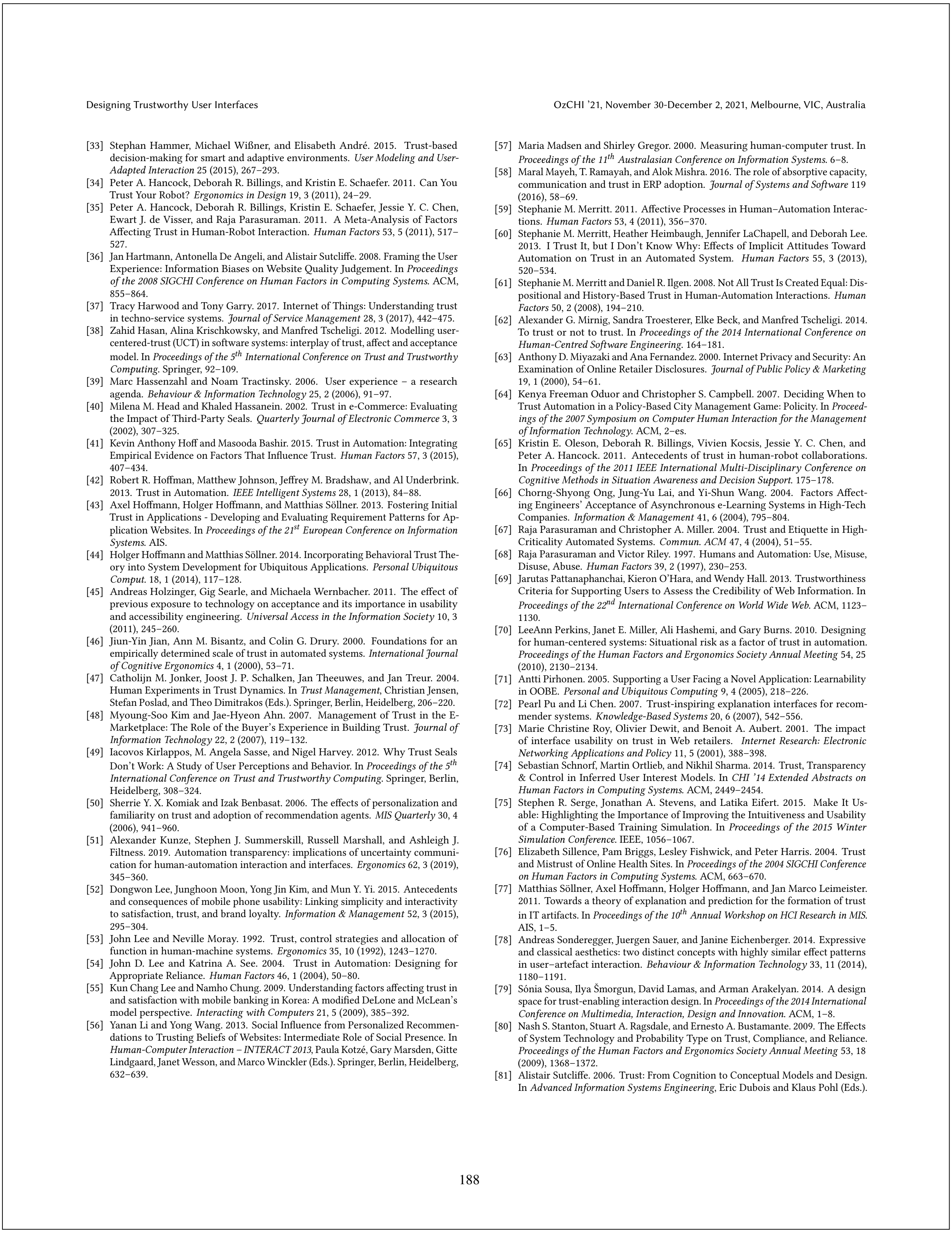
[28] Bronwyn French, Andreas Duenser, and Andrew Heathcote. 2018. *Trust in automation – a literature review.* Technical Report EP184082. CSIRO.

[29] Yuan Gao. 2005. Factors influencing user trust in online games. *Electronic Library* 23, 5 (2005), 533–538.

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

[31] P. Goillau, C. Kelly, M. Boardman, and Emmanuelle Jeannot. 2003. *Guidelines for trust in future ATM systems: measures.* Technical Report HRS/HSP-005-GUI-02. European Air Traffic Management Programme.

[32] Christian Gram. 1996. A software engineering view of user interface design. In *Proceedings of the 1995 IFIP International Conference on Engineering for Human-Computer Interaction*. Springer, 293–306.



[33] Stephan Hammer, Michael Wißner, and Elisabeth André. 2015. Trust-based decision-making for smart and adaptive environments. *User Modeling and User-Adapted Interaction* 25 (2015), 267–293.

[34] Peter A. Hancock, Deborah R. Billings, and Kristin E. Schaefer. 2011. Can You Trust Your Robot? *Ergonomics in Design* 19, 3 (2011), 24–29.

[35] Peter A. Hancock, Deborah R. Billings, Kristin E. Schaefer, Jessie Y. C. Chen, Ewart J. de Visser, and Raja Parasuraman. 2011. A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *Human Factors* 53, 5 (2011), 517–527.

[36] Jan Hartmann, Antonella De Angeli, and Alistair Sutcliffe. 2008. Framing the User Experience: Information Biases on Website Quality Judgement. In *Proceedings of the 2008 SIGCHI Conference on Human Factors in Computing Systems*. ACM, 855–864.

[37] Tracy Harwood and Tony Garry. 2017. Internet of Things: Understanding trust in techno-service systems. *Journal of Service Management* 28, 3 (2017), 442–475.

[38] Zahid Hasan, Alina Krischkowsky, and Manfred Tscheligi. 2012. Modelling user-centered-trust (UCT) in software systems: interplay of trust, affect and acceptance model. In *Proceedings of the 5th International Conference on Trust and Trustworthy Computing*. Springer, 92–109.

[39] Marc Hassenzahl and Noam Tractinsky. 2006. User experience – a research agenda. *Behaviour & Information Technology* 25, 2 (2006), 91–97.

[40] Milena M. Head and Khaled Hassanein. 2002. Trust in e-Commerce: Evaluating the Impact of Third-Party Seals. *Quarterly Journal of Electronic Commerce* 3, 3 (2002), 307–325.

[41] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors* 57, 3 (2015), 407–434.

[42] Robert R. Hoffman, Matthew Johnson, Jeffrey M. Bradshaw, and Al Underbrink. 2013. Trust in Automation. *IEEE Intelligent Systems* 28, 1 (2013), 84–88.

[43] Axel Hoffmann, Holger Hoffmann, and Matthias Söllner. 2013. Fostering Initial Trust in Applications - Developing and Evaluating Requirement Patterns for Application Websites. In *Proceedings of the 21st European Conference on Information Systems*. AIS.

[44] Holger Hoffmann and Matthias Söllner. 2014. Incorporating Behavioral Trust Theory into System Development for Ubiquitous Applications. *Personal Ubiquitous Comput.* 18, 1 (2014), 117–128.

[45] Andreas Holzinger, Gig Searle, and Michaela Wernbacher. 2011. The effect of previous exposure to technology on acceptance and its importance in usability and accessibility engineering. *Universal Access in the Information Society* 10, 3 (2011), 245–260.

[46] Jiun-Yin Jian, Ann M. Bisantz, and Colin G. Drury. 2000. Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics* 4, 1 (2000), 53–71.

[47] Catholijn M. Jonker, Joost J. P. Schalken, Jan Theeuwes, and Jan Treur. 2004. Human Experiments in Trust Dynamics. In *Trust Management*, Christian Jensen, Stefan Poslad, and Theo Dimitrakos (Eds.). Springer, Berlin, Heidelberg, 206–220.

[48] Myoung-Soo Kim and Jae-Hyeon Ahn. 2007. Management of Trust in the E-Marketplace: The Role of the Buyer's Experience in Building Trust. *Journal of Information Technology* 22, 2 (2007), 119–132.

[49] Iacovos Kirlappos, M. Angela Sasse, and Nigel Harvey. 2012. Why Trust Seals Don't Work: A Study of User Perceptions and Behavior. In *Proceedings of the 5th International Conference on Trust and Trustworthy Computing*. Springer, Berlin, Heidelberg, 308–324.

[50] Sherrie Y. X. Komiak and Izak Benbasat. 2006. The effects of personalization and familiarity on trust and adoption of recommendation agents. *MIS Quarterly* 30, 4 (2006), 941–960.

[51] Alexander Kunze, Stephen J. Summerskill, Russell Marshall, and Ashleigh J. Filtness. 2019. Automation transparency: implications of uncertainty communication for human-automation interaction and interfaces. *Ergonomics* 62, 3 (2019), 345–360.

[52] Dongwon Lee, Junghoon Moon, Yong Jin Kim, and Mun Y. Yi. 2015. Antecedents and consequences of mobile phone usability: Linking simplicity and interactivity to satisfaction, trust, and brand loyalty. *Information & Management* 52, 3 (2015), 295–304.

[53] John Lee and Neville Moray. 1992. Trust, control strategies and allocation of function in human-machine systems. *Ergonomics* 35, 10 (1992), 1243–1270.

[54] John D. Lee and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors* 46, 1 (2004), 50–80.

[55] Kun Chang Lee and Namho Chung. 2009. Understanding factors affecting trust in and satisfaction with mobile banking in Korea: A modified DeLone and McLean's model perspective. *Interacting with Computers* 21, 5 (2009), 385–392.

[56] Yanan Li and Yong Wang. 2013. Social Influence from Personalized Recommendations to Trusting Beliefs of Websites: Intermediate Role of Social Presence. In *Human-Computer Interaction – INTERACT 2013*, Paula Kotzé, Gary Marsden, Gitte Lindgaard, Janet Wesson, and Marco Winckler (Eds.). Springer, Berlin, Heidelberg, 632–639.

[57] Maria Madsen and Shirley Gregor. 2000. Measuring human-computer trust. In *Proceedings of the 11th Australasian Conference on Information Systems*. 6–8.

[58] Maral Mayeh, T. Ramayah, and Alok Mishra. 2016. The role of absorptive capacity, communication and trust in ERP adoption. *Journal of Systems and Software* 119 (2016), 58–69.

[59] Stephanie M. Merritt. 2011. Affective Processes in Human–Automation Interactions. *Human Factors* 53, 4 (2011), 356–370.

[60] Stephanie M. Merritt, Heather Heimbaugh, Jennifer LaChapell, and Deborah Lee. 2013. I Trust It, but I Don't Know Why: Effects of Implicit Attitudes Toward Automation on Trust in an Automated System. *Human Factors* 55, 3 (2013), 520–534.

[61] Stephanie M. Merritt and Daniel R. Ilgen. 2008. Not All Trust Is Created Equal: Dispositional and History-Based Trust in Human-Automation Interactions. *Human Factors* 50, 2 (2008), 194–210.

[62] Alexander G. Mirnig, Sandra Troesterer, Elke Beck, and Manfred Tscheligi. 2014. To trust or not to trust. In *Proceedings of the 2014 International Conference on Human-Centred Software Engineering*. 164–181.

[63] Anthony D. Miyazaki and Ana Fernandez. 2000. Internet Privacy and Security: An Examination of Online Retailer Disclosures. *Journal of Public Policy & Marketing* 19, 1 (2000), 54–61.

[64] Kenya Freeman Oduor and Christopher S. Campbell. 2007. Deciding When to Trust Automation in a Policy-Based City Management Game: Policity. In *Proceedings of the 2007 Symposium on Computer Human Interaction for the Management of Information Technology*. ACM, 2–es.

[65] Kristin E. Oleson, Deborah R. Billings, Vivien Kocsis, Jessie Y. C. Chen, and Peter A. Hancock. 2011. Antecedents of trust in human-robot collaborations. In *Proceedings of the 2011 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support*. 175–178.

[66] Chorng-Shyong Ong, Jung-Yu Lai, and Yi-Shun Wang. 2004. Factors Affecting Engineers' Acceptance of Asynchronous e-Learning Systems in High-Tech Companies. *Information & Management* 41, 6 (2004), 795–804.

[67] Raja Parasuraman and Christopher A. Miller. 2004. Trust and Etiquette in High-Criticality Automated Systems. *Commun. ACM* 47, 4 (2004), 51–55.

[68] Raja Parasuraman and Victor Riley. 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors* 39, 2 (1997), 230–253.

[69] Jarutas Pattanaphanchai, Kieron O'Hara, and Wendy Hall. 2013. Trustworthiness Criteria for Supporting Users to Assess the Credibility of Web Information. In *Proceedings of the 22nd International Conference on World Wide Web*. ACM, 1123–1130.

[70] LeeAnn Perkins, Janet E. Miller, Ali Hashemi, and Gary Burns. 2010. Designing for human-centered systems: Situational risk as a factor of trust in automation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 54, 25 (2010), 2130–2134.

[71] Antti Pirhonen. 2005. Supporting a User Facing a Novel Application: Learnability in OOBE. *Personal and Ubiquitous Computing* 9, 4 (2005), 218–226.

[72] Pearl Pu and Li Chen. 2007. Trust-inspiring explanation interfaces for recommender systems. *Knowledge-Based Systems* 20, 6 (2007), 542–556.

[73] Marie Christine Roy, Olivier Dewit, and Benoit A. Aubert. 2001. The impact of interface usability on trust in Web retailers. *Internet Research: Electronic Networking Applications and Policy* 11, 5 (2001), 388–398.

[74] Sebastian Schnorf, Martin Ortlieb, and Nikhil Sharma. 2014. Trust, Transparency & Control in Inferred User Interest Models. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2449–2454.

[75] Stephen R. Serge, Jonathan A. Stevens, and Latika Eifert. 2015. Make It Usable: Highlighting the Importance of Improving the Intuitiveness and Usability of a Computer-Based Training Simulation. In *Proceedings of the 2015 Winter Simulation Conference*. IEEE, 1056–1067.

[76] Elizabeth Sillence, Pam Briggs, Lesley Fishwick, and Peter Harris. 2004. Trust and Mistrust of Online Health Sites. In *Proceedings of the 2004 SIGCHI Conference on Human Factors in Computing Systems*. ACM, 663–670.

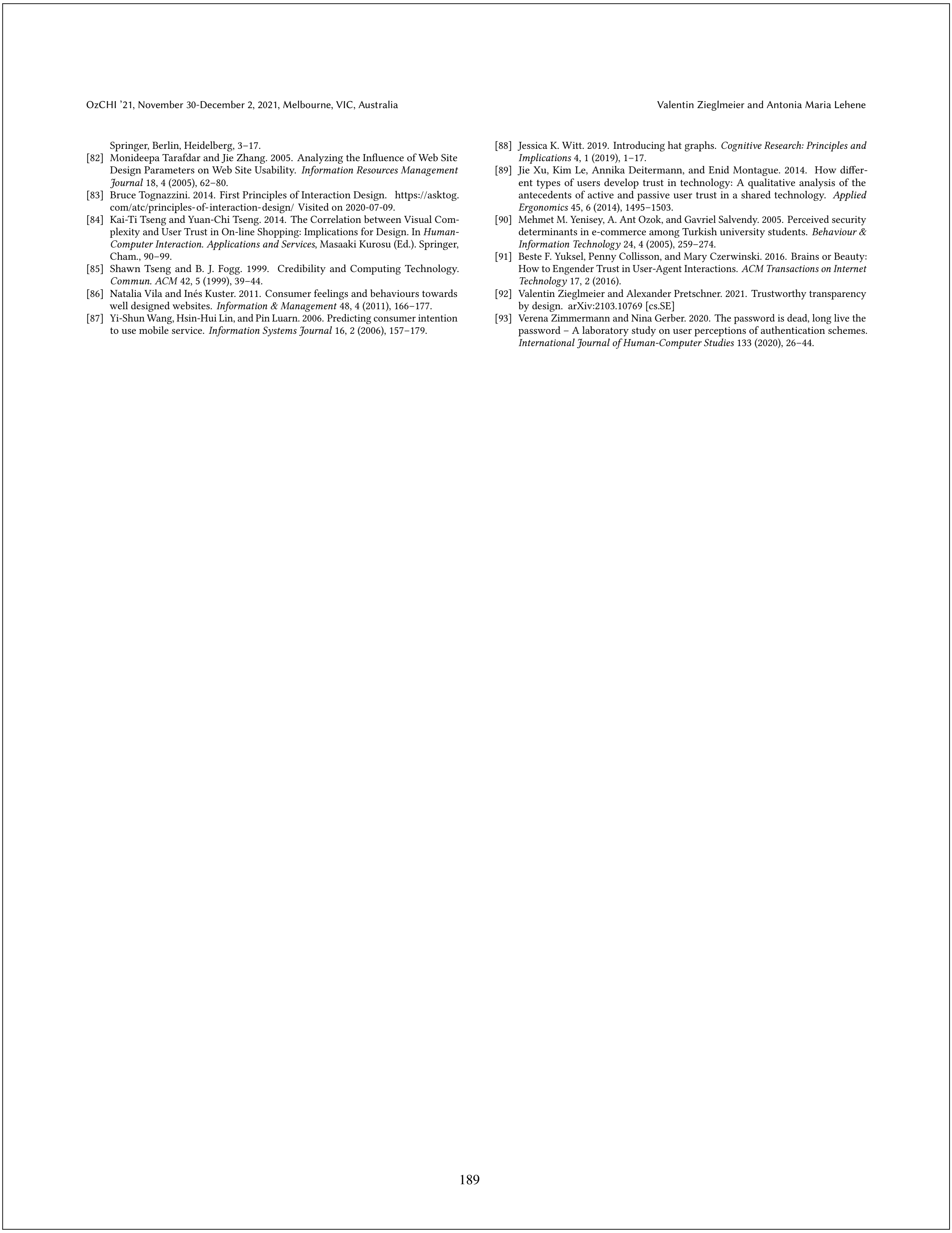
[77] Matthias Söllner, Axel Hoffmann, Holger Hoffmann, and Jan Marco Leimeister. 2011. Towards a theory of explanation and prediction for the formation of trust in IT artifacts. In *Proceedings of the 10th Annual Workshop on HCI Research in MIS*. AIS, 1–5.

[78] Andreas Sonderegger, Juergen Sauer, and Janine Eichenberger. 2014. Expressive and classical aesthetics: two distinct concepts with highly similar effect patterns in user–artefact interaction. *Behaviour & Information Technology* 33, 11 (2014), 1180–1191.

[79] Sónia Sousa, Ilya Šmorgun, David Lamas, and Arman Arakelyan. 2014. A design space for trust-enabling interaction design. In *Proceedings of the 2014 International Conference on Multimedia, Interaction, Design and Innovation*. ACM, 1–8.

[80] Nash S. Stanton, Stuart A. Ragsdale, and Ernesto A. Bustamante. 2009. The Effects of System Technology and Probability Type on Trust, Compliance, and Reliance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 53, 18 (2009), 1368–1372.

[81] Alistair Sutcliffe. 2006. Trust: From Cognition to Conceptual Models and Design. In *Advanced Information Systems Engineering*, Eric Dubois and Klaus Pohl (Eds.).



Springer, Berlin, Heidelberg, 3–17.

[82] Monideepa Tarafdar and Jie Zhang. 2005. Analyzing the Influence of Web Site Design Parameters on Web Site Usability. *Information Resources Management Journal* 18, 4 (2005), 62–80.

[83] Bruce Tognazzini. 2014. First Principles of Interaction Design. https://asktog.com/atc/principles-of-interaction-design/ Visited on 2020-07-09.

[84] Kai-Ti Tseng and Yuan-Chi Tseng. 2014. The Correlation between Visual Complexity and User Trust in On-line Shopping: Implications for Design. In *Human-Computer Interaction. Applications and Services*, Masaaki Kurosu (Ed.). Springer, Cham., 90–99.

[85] Shawn Tseng and B. J. Fogg. 1999. Credibility and Computing Technology. *Commun. ACM* 42, 5 (1999), 39–44.

[86] Natalia Vila and Inés Kuster. 2011. Consumer feelings and behaviours towards well designed websites. *Information & Management* 48, 4 (2011), 166–177.

[87] Yi-Shun Wang, Hsin-Hui Lin, and Pin Luarn. 2006. Predicting consumer intention to use mobile service. *Information Systems Journal* 16, 2 (2006), 157–179.

[88] Jessica K. Witt. 2019. Introducing hat graphs. *Cognitive Research: Principles and Implications* 4, 1 (2019), 1–17.

[89] Jie Xu, Kim Le, Annika Deitermann, and Enid Montague. 2014. How different types of users develop trust in technology: A qualitative analysis of the antecedents of active and passive user trust in a shared technology. *Applied Ergonomics* 45, 6 (2014), 1495–1503.

[90] Mehmet M. Yenisey, A. Ant Ozok, and Gavriel Salvendy. 2005. Perceived security determinants in e-commerce among Turkish university students. *Behaviour & Information Technology* 24, 4 (2005), 259–274.

[91] Beste F. Yuksel, Penny Collisson, and Mary Czerwinski. 2016. Brains or Beauty: How to Engender Trust in User-Agent Interactions. *ACM Transactions on Internet Technology* 17, 2 (2016).

[92] Valentin Zieglmeier and Alexander Pretschner. 2021. Trustworthy transparency by design. arXiv:2103.10769 [cs.SE]

[93] Verena Zimmermann and Nina Gerber. 2020. The password is dead, long live the password – A laboratory study on user perceptions of authentication schemes. *International Journal of Human-Computer Studies* 133 (2020), 26–44.

7. Increasing Employees' Willingness to Share: Introducing Appeal Strategies for People Analytics

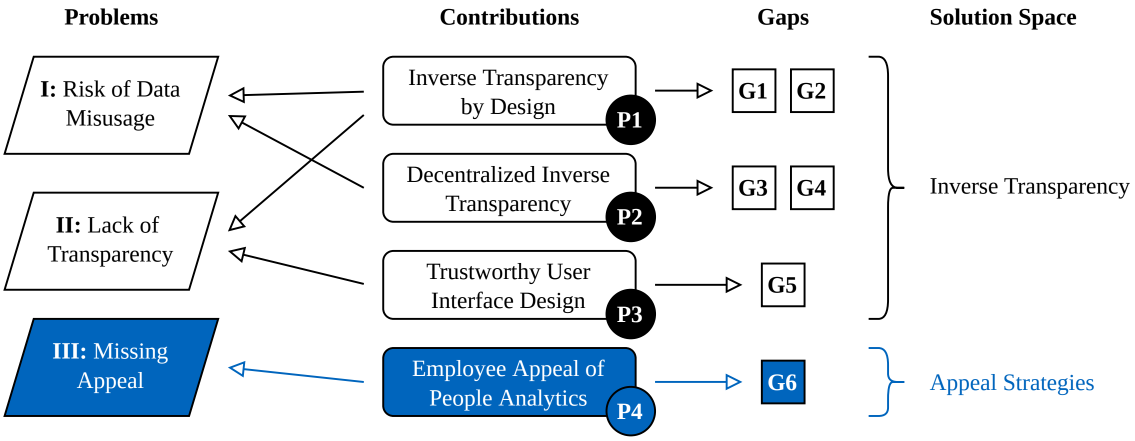


Figure 7.1.: Relationship of our fourth contribution to the big picture.

Summary. The following summary is partially adapted from our paper [see 4].

- *Problem:* By reducing the risks in PA (see Chapters 4 and 5) and providing transparency (see Chapters 4 and 6), employees' concerns—which impact their willingness to share data—can be reduced. Yet, this does not suffice, as the individual's privacy calculus (see Section 2.4.1) furthermore includes the expected benefits. Therefore, we need to alleviate the *missing appeal* (problem **III**) of PA.

- *Solution:* To provide employees with positive motivation to share their data, we can incorporate appeal strategies into PA.

- *Gap:* Most existing research on software appeal focuses on the consumer context, which differs from the workplace. Therefore, the strategies cannot be transferred. No systematic guidance exists on how to increase the appeal of PA given their workplace context (**G6**).

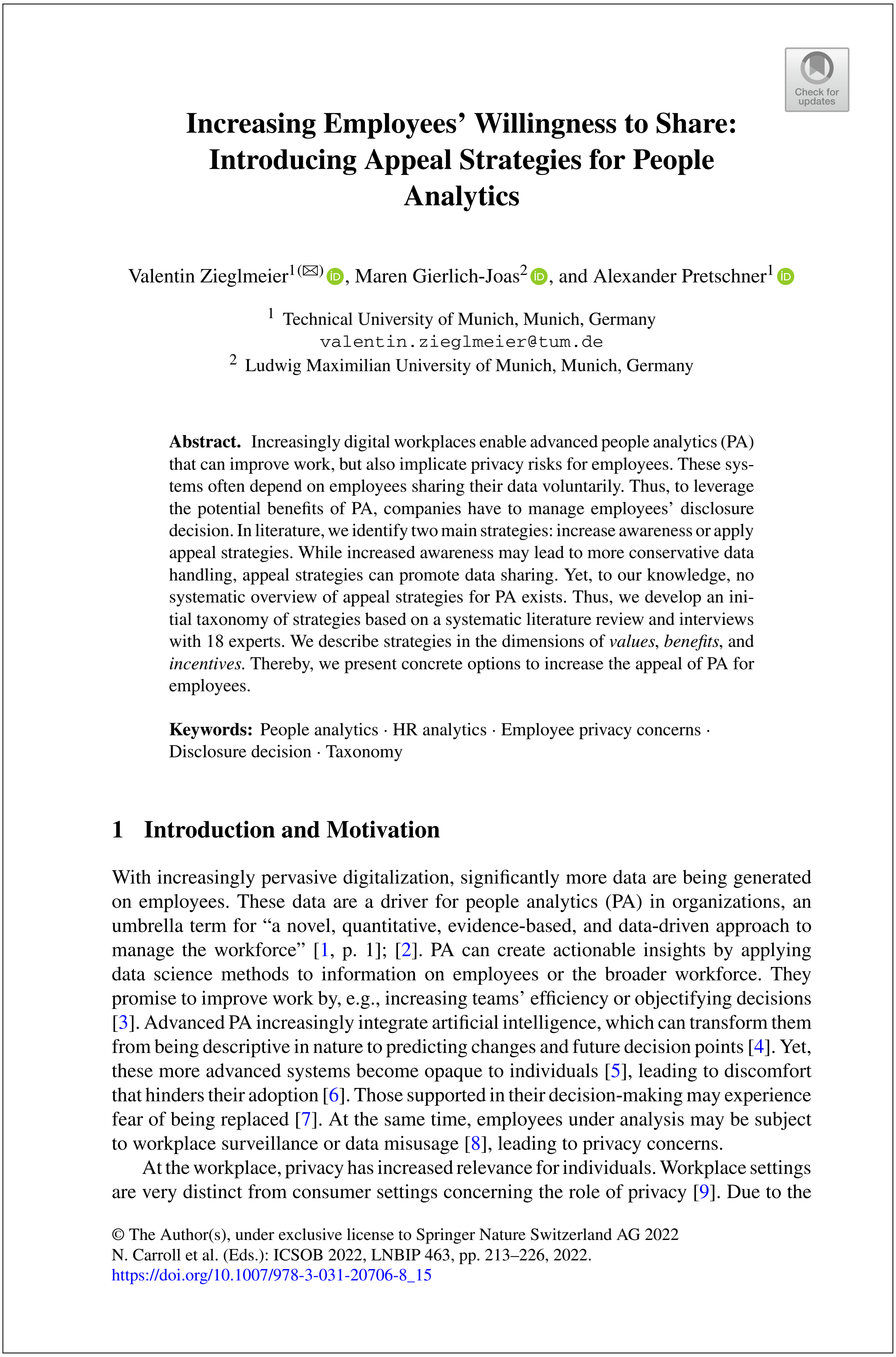
- *Contribution:* Based on a systematic literature review and expert interviews, we present a taxonomy of appeal strategies for PA. From the examples found, we inductively derive three overarching dimensions, namely *values*, *benefits*, and *incentives*.

- *Limitations:* Our taxonomy is based on literature and expert opinions, but it was not empirically tested with employees, who are the main stakeholders in our model. Furthermore, as only limited literature exists, our systematic approach may not suffice to ensure the completeness of the taxonomy.

Author Contributions. VZ developed the initial research idea and theorized about the underlying dimensions of appeal strategies. In discussion with AP, VZ further refined the idea. VZ and MGJ developed the research plan. VZ conducted the systematic literature review with support from MGJ. Together, VZ and MGJ acquired interview participants, performed the interviews, coded them, and distilled the results. VZ and MGJ wrote an initial manuscript in close discussion with AP. Based on this, VZ reframed the paper to strengthen the story and created the final manuscript with feedback from AP.

Copyright Note. © 2022 The authors. Reproduced with permission from Springer Nature. Valentin Zieglmeier, Maren Gierlich-Joas, and Alexander Pretschner. "Increasing Employees' Willingness to Share: Introducing Appeal Strategies for People Analytics." In: *Proceedings of the 13th International Conference on Software Business*. Lecture Notes in Business Information Processing 463. Springer, 2022, pp. 213–226.

On the following pages, the published paper is reprinted in accordance with the license granted to us by Springer Nature. The version of record can be found with the following DOI: 10.1007/978-3-031-20706-8_15.



Increasing Employees' Willingness to Share: Introducing Appeal Strategies for People Analytics

Valentin Zieglmeier[1](✉), Maren Gierlich-Joas[2], and Alexander Pretschner[1]

[1] Technical University of Munich, Munich, Germany
valentin.zieglmeier@tum.de
[2] Ludwig Maximilian University of Munich, Munich, Germany

Abstract. Increasingly digital workplaces enable advanced people analytics (PA) that can improve work, but also implicate privacy risks for employees. These systems often depend on employees sharing their data voluntarily. Thus, to leverage the potential benefits of PA, companies have to manage employees' disclosure decision. In literature, we identify two main strategies: increase awareness or apply appeal strategies. While increased awareness may lead to more conservative data handling, appeal strategies can promote data sharing. Yet, to our knowledge, no systematic overview of appeal strategies for PA exists. Thus, we develop an initial taxonomy of strategies based on a systematic literature review and interviews with 18 experts. We describe strategies in the dimensions of *values*, *benefits*, and *incentives*. Thereby, we present concrete options to increase the appeal of PA for employees.

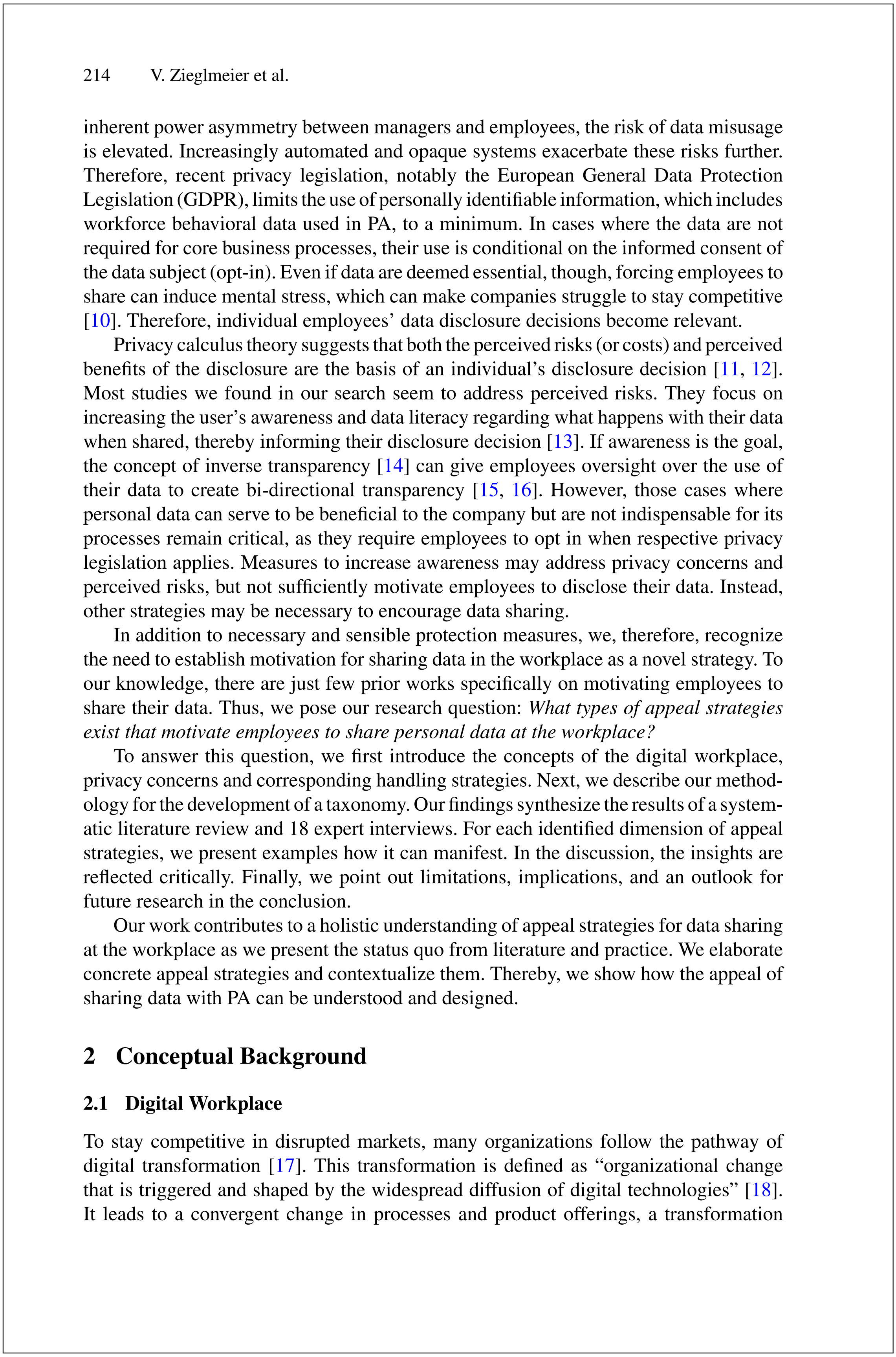
Keywords: People analytics · HR analytics · Employee privacy concerns · Disclosure decision · Taxonomy

1 Introduction and Motivation

With increasingly pervasive digitalization, significantly more data are being generated on employees. These data are a driver for people analytics (PA) in organizations, an umbrella term for "a novel, quantitative, evidence-based, and data-driven approach to manage the workforce" [1, p. 1]; [2]. PA can create actionable insights by applying data science methods to information on employees or the broader workforce. They promise to improve work by, e.g., increasing teams' efficiency or objectifying decisions [3]. Advanced PA increasingly integrate artificial intelligence, which can transform them from being descriptive in nature to predicting changes and future decision points [4]. Yet, these more advanced systems become opaque to individuals [5], leading to discomfort that hinders their adoption [6]. Those supported in their decision-making may experience fear of being replaced [7]. At the same time, employees under analysis may be subject to workplace surveillance or data misusage [8], leading to privacy concerns.

At the workplace, privacy has increased relevance for individuals. Workplace settings are very distinct from consumer settings concerning the role of privacy [9]. Due to the

N. Carroll et al. (Eds.): ICSOB 2022, LNBIP 463, pp. 213–226, 2022.
https://doi.org/10.1007/978-3-031-20706-8_15



inherent power asymmetry between managers and employees, the risk of data misusage is elevated. Increasingly automated and opaque systems exacerbate these risks further. Therefore, recent privacy legislation, notably the European General Data Protection Legislation (GDPR), limits the use of personally identifiable information, which includes workforce behavioral data used in PA, to a minimum. In cases where the data are not required for core business processes, their use is conditional on the informed consent of the data subject (opt-in). Even if data are deemed essential, though, forcing employees to share can induce mental stress, which can make companies struggle to stay competitive [10]. Therefore, individual employees' data disclosure decisions become relevant.

Privacy calculus theory suggests that both the perceived risks (or costs) and perceived benefits of the disclosure are the basis of an individual's disclosure decision [11, 12]. Most studies we found in our search seem to address perceived risks. They focus on increasing the user's awareness and data literacy regarding what happens with their data when shared, thereby informing their disclosure decision [13]. If awareness is the goal, the concept of inverse transparency [14] can give employees oversight over the use of their data to create bi-directional transparency [15, 16]. However, those cases where personal data can serve to be beneficial to the company but are not indispensable for its processes remain critical, as they require employees to opt in when respective privacy legislation applies. Measures to increase awareness may address privacy concerns and perceived risks, but not sufficiently motivate employees to disclose their data. Instead, other strategies may be necessary to encourage data sharing.

In addition to necessary and sensible protection measures, we, therefore, recognize the need to establish motivation for sharing data in the workplace as a novel strategy. To our knowledge, there are just few prior works specifically on motivating employees to share their data. Thus, we pose our research question: *What types of appeal strategies exist that motivate employees to share personal data at the workplace?*

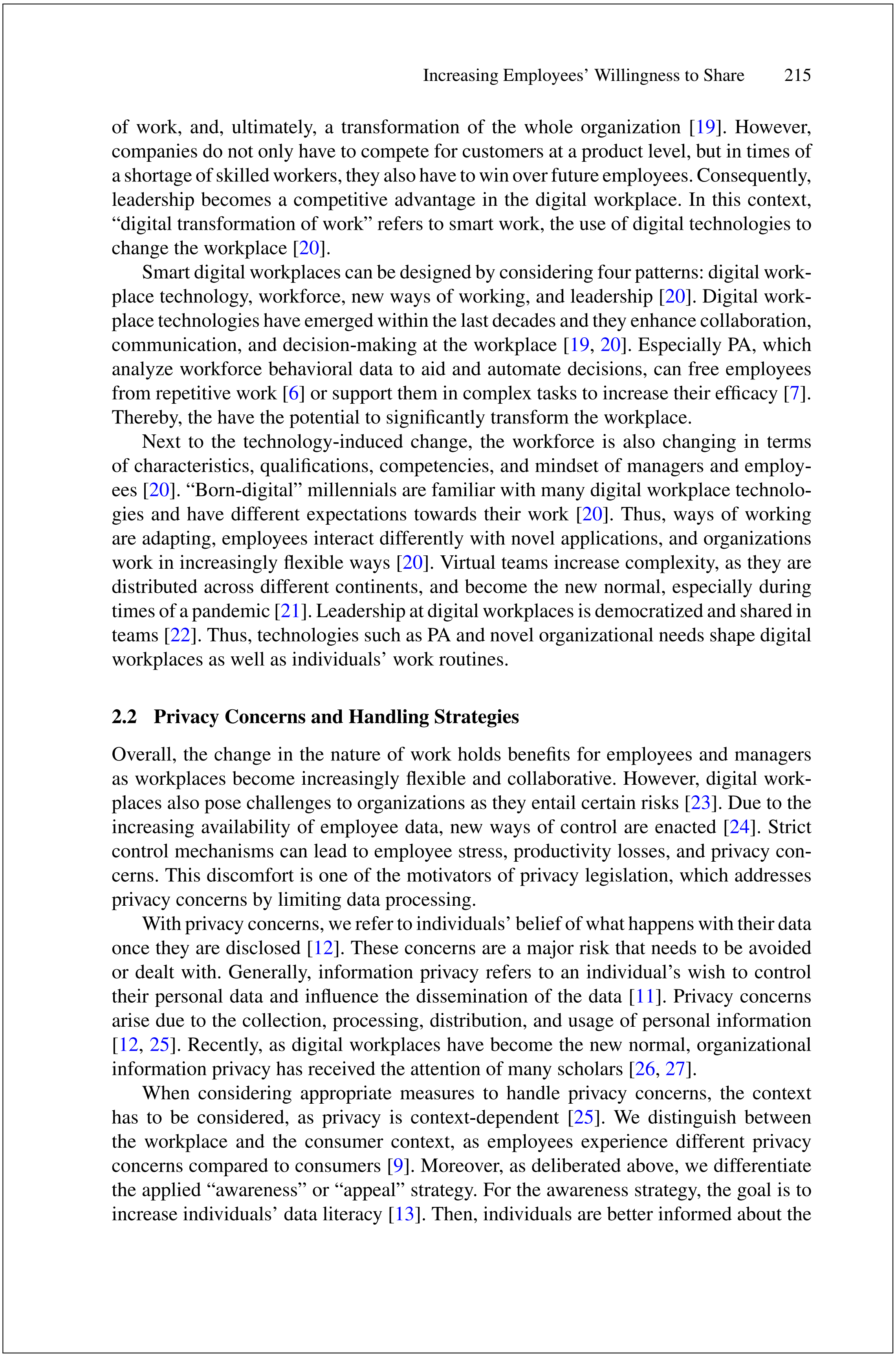
To answer this question, we first introduce the concepts of the digital workplace, privacy concerns and corresponding handling strategies. Next, we describe our methodology for the development of a taxonomy. Our findings synthesize the results of a systematic literature review and 18 expert interviews. For each identified dimension of appeal strategies, we present examples how it can manifest. In the discussion, the insights are reflected critically. Finally, we point out limitations, implications, and an outlook for future research in the conclusion.

Our work contributes to a holistic understanding of appeal strategies for data sharing at the workplace as we present the status quo from literature and practice. We elaborate concrete appeal strategies and contextualize them. Thereby, we show how the appeal of sharing data with PA can be understood and designed.

2 Conceptual Background

2.1 Digital Workplace

To stay competitive in disrupted markets, many organizations follow the pathway of digital transformation [17]. This transformation is defined as "organizational change that is triggered and shaped by the widespread diffusion of digital technologies" [18]. It leads to a convergent change in processes and product offerings, a transformation



of work, and, ultimately, a transformation of the whole organization [19]. However, companies do not only have to compete for customers at a product level, but in times of a shortage of skilled workers, they also have to win over future employees. Consequently, leadership becomes a competitive advantage in the digital workplace. In this context, "digital transformation of work" refers to smart work, the use of digital technologies to change the workplace [20].

Smart digital workplaces can be designed by considering four patterns: digital workplace technology, workforce, new ways of working, and leadership [20]. Digital workplace technologies have emerged within the last decades and they enhance collaboration, communication, and decision-making at the workplace [19, 20]. Especially PA, which analyze workforce behavioral data to aid and automate decisions, can free employees from repetitive work [6] or support them in complex tasks to increase their efficacy [7]. Thereby, the have the potential to significantly transform the workplace.

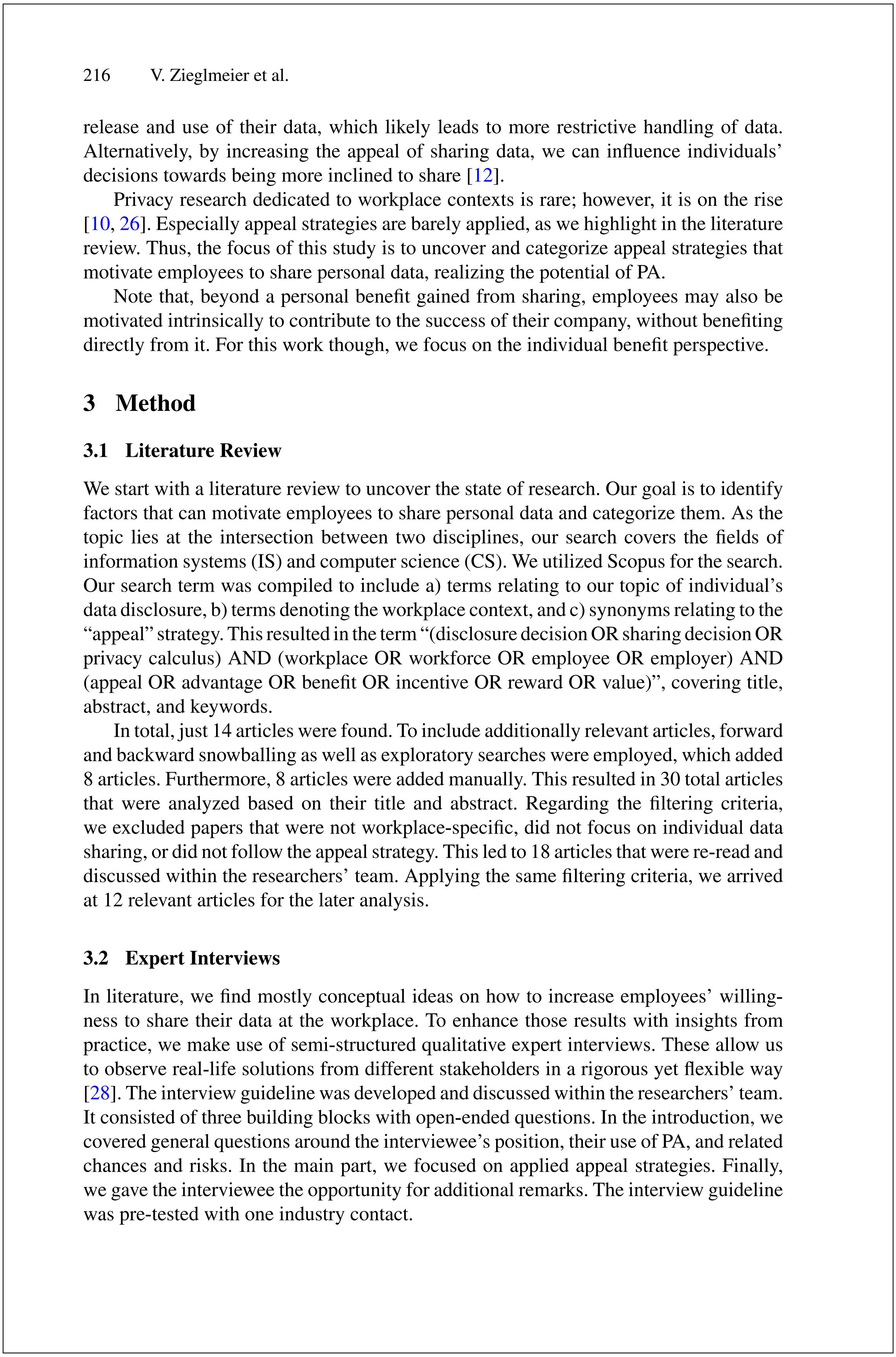
Next to the technology-induced change, the workforce is also changing in terms of characteristics, qualifications, competencies, and mindset of managers and employees [20]. "Born-digital" millennials are familiar with many digital workplace technologies and have different expectations towards their work [20]. Thus, ways of working are adapting, employees interact differently with novel applications, and organizations work in increasingly flexible ways [20]. Virtual teams increase complexity, as they are distributed across different continents, and become the new normal, especially during times of a pandemic [21]. Leadership at digital workplaces is democratized and shared in teams [22]. Thus, technologies such as PA and novel organizational needs shape digital workplaces as well as individuals' work routines.

2.2 Privacy Concerns and Handling Strategies

Overall, the change in the nature of work holds benefits for employees and managers as workplaces become increasingly flexible and collaborative. However, digital workplaces also pose challenges to organizations as they entail certain risks [23]. Due to the increasing availability of employee data, new ways of control are enacted [24]. Strict control mechanisms can lead to employee stress, productivity losses, and privacy concerns. This discomfort is one of the motivators of privacy legislation, which addresses privacy concerns by limiting data processing.

With privacy concerns, we refer to individuals' belief of what happens with their data once they are disclosed [12]. These concerns are a major risk that needs to be avoided or dealt with. Generally, information privacy refers to an individual's wish to control their personal data and influence the dissemination of the data [11]. Privacy concerns arise due to the collection, processing, distribution, and usage of personal information [12, 25]. Recently, as digital workplaces have become the new normal, organizational information privacy has received the attention of many scholars [26, 27].

When considering appropriate measures to handle privacy concerns, the context has to be considered, as privacy is context-dependent [25]. We distinguish between the workplace and the consumer context, as employees experience different privacy concerns compared to consumers [9]. Moreover, as deliberated above, we differentiate the applied "awareness" or "appeal" strategy. For the awareness strategy, the goal is to increase individuals' data literacy [13]. Then, individuals are better informed about the



release and use of their data, which likely leads to more restrictive handling of data. Alternatively, by increasing the appeal of sharing data, we can influence individuals' decisions towards being more inclined to share [12].

Privacy research dedicated to workplace contexts is rare; however, it is on the rise [10, 26]. Especially appeal strategies are barely applied, as we highlight in the literature review. Thus, the focus of this study is to uncover and categorize appeal strategies that motivate employees to share personal data, realizing the potential of PA.

Note that, beyond a personal benefit gained from sharing, employees may also be motivated intrinsically to contribute to the success of their company, without benefiting directly from it. For this work though, we focus on the individual benefit perspective.

3 Method

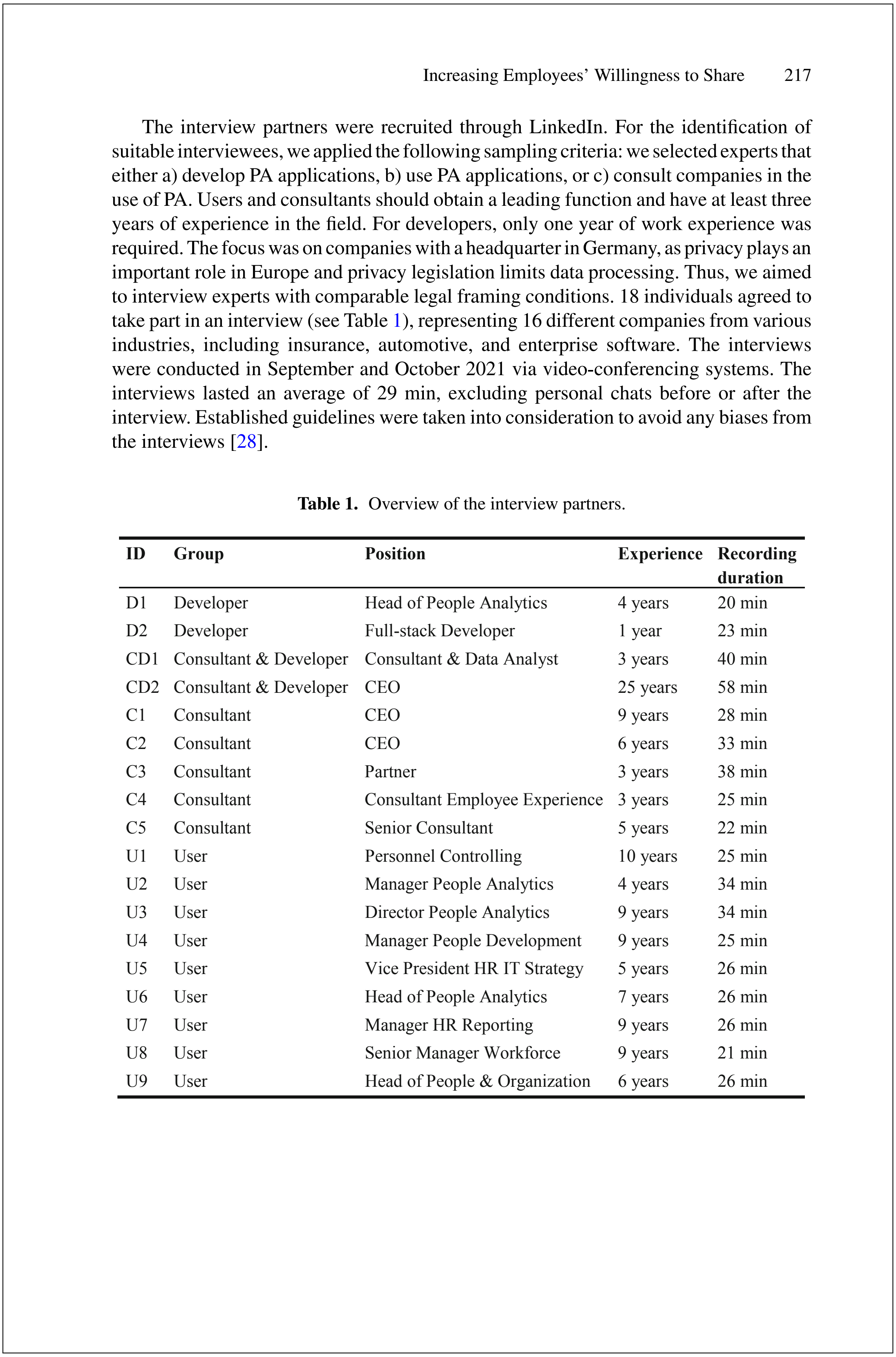
3.1 Literature Review

We start with a literature review to uncover the state of research. Our goal is to identify factors that can motivate employees to share personal data and categorize them. As the topic lies at the intersection between two disciplines, our search covers the fields of information systems (IS) and computer science (CS). We utilized Scopus for the search. Our search term was compiled to include a) terms relating to our topic of individual's data disclosure, b) terms denoting the workplace context, and c) synonyms relating to the "appeal" strategy. This resulted in the term "(disclosure decision OR sharing decision OR privacy calculus) AND (workplace OR workforce OR employee OR employer) AND (appeal OR advantage OR benefit OR incentive OR reward OR value)", covering title, abstract, and keywords.

In total, just 14 articles were found. To include additionally relevant articles, forward and backward snowballing as well as exploratory searches were employed, which added 8 articles. Furthermore, 8 articles were added manually. This resulted in 30 total articles that were analyzed based on their title and abstract. Regarding the filtering criteria, we excluded papers that were not workplace-specific, did not focus on individual data sharing, or did not follow the appeal strategy. This led to 18 articles that were re-read and discussed within the researchers' team. Applying the same filtering criteria, we arrived at 12 relevant articles for the later analysis.

3.2 Expert Interviews

In literature, we find mostly conceptual ideas on how to increase employees' willingness to share their data at the workplace. To enhance those results with insights from practice, we make use of semi-structured qualitative expert interviews. These allow us to observe real-life solutions from different stakeholders in a rigorous yet flexible way [28]. The interview guideline was developed and discussed within the researchers' team. It consisted of three building blocks with open-ended questions. In the introduction, we covered general questions around the interviewee's position, their use of PA, and related chances and risks. In the main part, we focused on applied appeal strategies. Finally, we gave the interviewee the opportunity for additional remarks. The interview guideline was pre-tested with one industry contact.



The interview partners were recruited through LinkedIn. For the identification of suitable interviewees, we applied the following sampling criteria: we selected experts that either a) develop PA applications, b) use PA applications, or c) consult companies in the use of PA. Users and consultants should obtain a leading function and have at least three years of experience in the field. For developers, only one year of work experience was required. The focus was on companies with a headquarter in Germany, as privacy plays an important role in Europe and privacy legislation limits data processing. Thus, we aimed to interview experts with comparable legal framing conditions. 18 individuals agreed to take part in an interview (see Table 1), representing 16 different companies from various industries, including insurance, automotive, and enterprise software. The interviews were conducted in September and October 2021 via video-conferencing systems. The interviews lasted an average of 29 min, excluding personal chats before or after the interview. Established guidelines were taken into consideration to avoid any biases from the interviews [28].

Table 1. Overview of the interview partners.

ID	Group	Position	Experience	Recording duration
D1	Developer	Head of People Analytics	4 years	20 min
D2	Developer	Full-stack Developer	1 year	23 min
CD1	Consultant & Developer	Consultant & Data Analyst	3 years	40 min
CD2	Consultant & Developer	CEO	25 years	58 min
C1	Consultant	CEO	9 years	28 min
C2	Consultant	CEO	6 years	33 min
C3	Consultant	Partner	3 years	38 min
C4	Consultant	Consultant Employee Experience	3 years	25 min
C5	Consultant	Senior Consultant	5 years	22 min
U1	User	Personnel Controlling	10 years	25 min
U2	User	Manager People Analytics	4 years	34 min
U3	User	Director People Analytics	9 years	34 min
U4	User	Manager People Development	9 years	25 min
U5	User	Vice President HR IT Strategy	5 years	26 min
U6	User	Head of People Analytics	7 years	26 min
U7	User	Manager HR Reporting	9 years	26 min
U8	User	Senior Manager Workforce	9 years	21 min
U9	User	Head of People & Organization	6 years	26 min

All interviews were recorded, transcribed verbatim, and anonymized. For the analysis, we used the tool ATLAS.ti. The coding scheme was developed iteratively and discussed within the researchers' team. For the coding categories, we built on the findings from the literature review and adapted if necessary. The quotes below were translated from German to English. An example of the coding scheme can be found in Table 2.

Table 2. Coding scheme.

Dimension	Code	Example quote
Values	Engagement	"Employees are given a chance to actually change something by themselves" (D1)
Benefits	Time savings	"After all, we save a lot of time and money if we implement these systems." (U7)
Incentives	Monetary rewards	"It's hard to set a real incentive, you can say, you'll get a 50 € Amazon voucher if you fill your skills profile, but of course that doesn't work" (U5)

4 Findings

When looking at the results from our literature review and expert interviews, we can inductively derive three dimensions of appeal strategies: *values*, *benefits*, and *incentives*. Each dimension can be mapped to either the design or the usage phase of a tool (see Fig. 1). While the CS discipline mainly focuses on the design phase, the IS discipline rather addresses the usage phase and partly the design phase. Notably, incentives are the only option available in the usage phase to steer employees' behavior. They remain the only choice if the tool's design cannot be influenced.

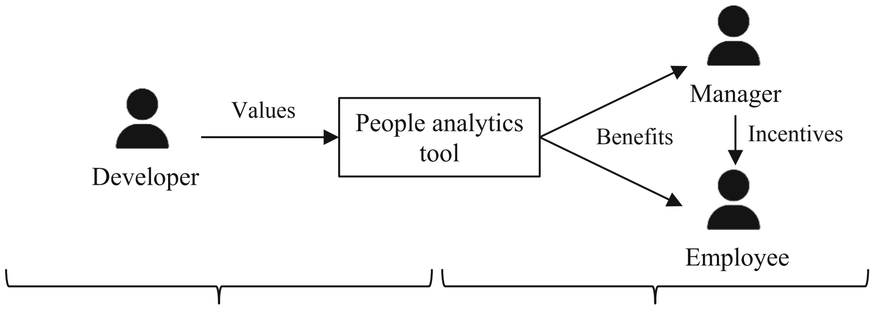
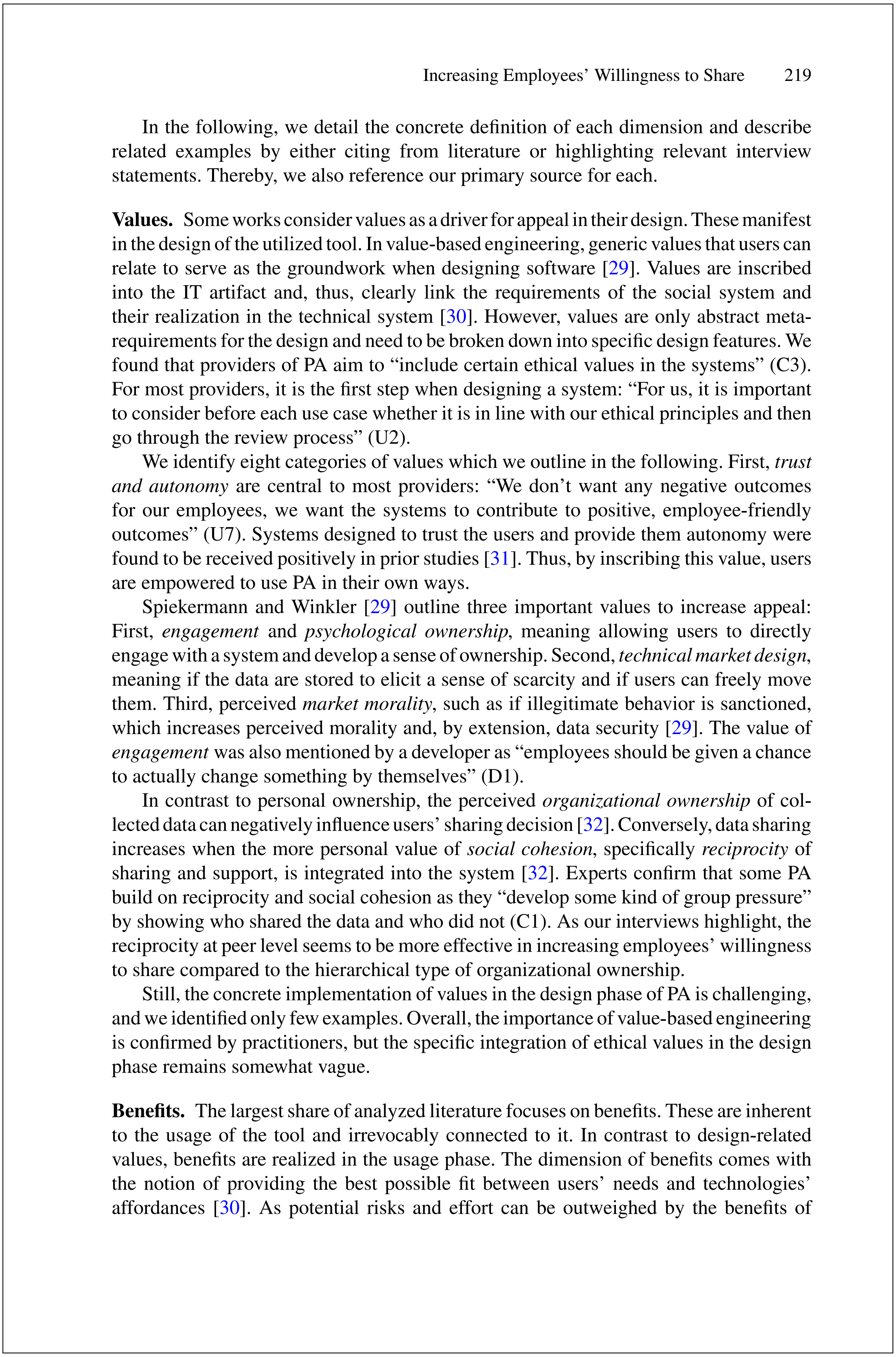


Fig. 1. Interplay between values, benefits, and incentives in the design phase and usage phase.



In the following, we detail the concrete definition of each dimension and describe related examples by either citing from literature or highlighting relevant interview statements. Thereby, we also reference our primary source for each.

Values. Some works consider values as a driver for appeal in their design. These manifest in the design of the utilized tool. In value-based engineering, generic values that users can relate to serve as the groundwork when designing software [29]. Values are inscribed into the IT artifact and, thus, clearly link the requirements of the social system and their realization in the technical system [30]. However, values are only abstract meta-requirements for the design and need to be broken down into specific design features. We found that providers of PA aim to "include certain ethical values in the systems" (C3). For most providers, it is the first step when designing a system: "For us, it is important to consider before each use case whether it is in line with our ethical principles and then go through the review process" (U2).

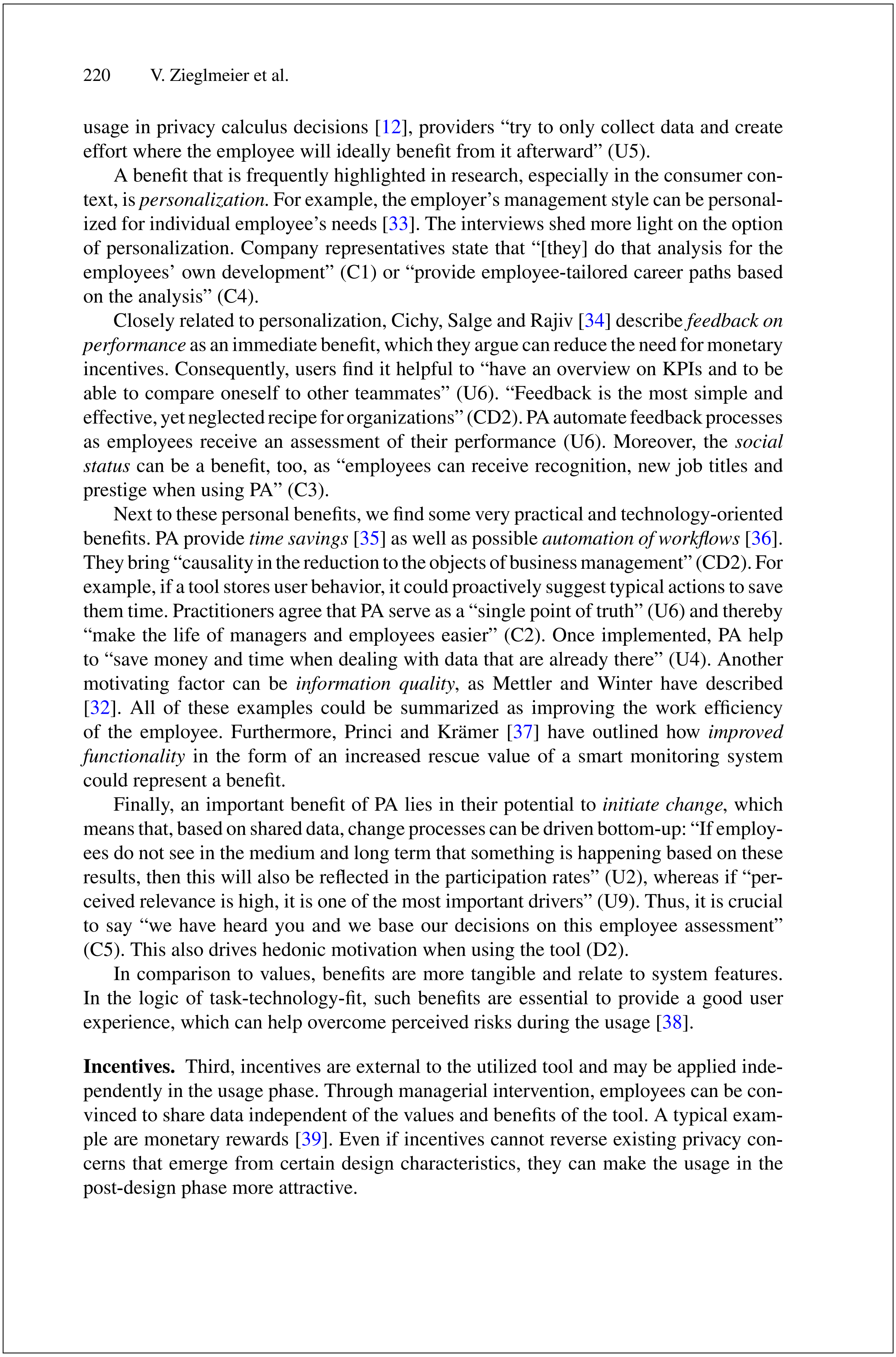
We identify eight categories of values which we outline in the following. First, *trust and autonomy* are central to most providers: "We don't want any negative outcomes for our employees, we want the systems to contribute to positive, employee-friendly outcomes" (U7). Systems designed to trust the users and provide them autonomy were found to be received positively in prior studies [31]. Thus, by inscribing this value, users are empowered to use PA in their own ways.

Spiekermann and Winkler [29] outline three important values to increase appeal: First, *engagement* and *psychological ownership*, meaning allowing users to directly engage with a system and develop a sense of ownership. Second, *technical market design*, meaning if the data are stored to elicit a sense of scarcity and if users can freely move them. Third, perceived *market morality*, such as if illegitimate behavior is sanctioned, which increases perceived morality and, by extension, data security [29]. The value of *engagement* was also mentioned by a developer as "employees should be given a chance to actually change something by themselves" (D1).

In contrast to personal ownership, the perceived *organizational ownership* of collected data can negatively influence users' sharing decision [32]. Conversely, data sharing increases when the more personal value of *social cohesion*, specifically *reciprocity* of sharing and support, is integrated into the system [32]. Experts confirm that some PA build on reciprocity and social cohesion as they "develop some kind of group pressure" by showing who shared the data and who did not (C1). As our interviews highlight, the reciprocity at peer level seems to be more effective in increasing employees' willingness to share compared to the hierarchical type of organizational ownership.

Still, the concrete implementation of values in the design phase of PA is challenging, and we identified only few examples. Overall, the importance of value-based engineering is confirmed by practitioners, but the specific integration of ethical values in the design phase remains somewhat vague.

Benefits. The largest share of analyzed literature focuses on benefits. These are inherent to the usage of the tool and irrevocably connected to it. In contrast to design-related values, benefits are realized in the usage phase. The dimension of benefits comes with the notion of providing the best possible fit between users' needs and technologies' affordances [30]. As potential risks and effort can be outweighed by the benefits of



usage in privacy calculus decisions [12], providers "try to only collect data and create effort where the employee will ideally benefit from it afterward" (U5).

A benefit that is frequently highlighted in research, especially in the consumer context, is *personalization.* For example, the employer's management style can be personalized for individual employee's needs [33]. The interviews shed more light on the option of personalization. Company representatives state that "[they] do that analysis for the employees' own development" (C1) or "provide employee-tailored career paths based on the analysis" (C4).

Closely related to personalization, Cichy, Salge and Rajiv [34] describe *feedback on performance* as an immediate benefit, which they argue can reduce the need for monetary incentives. Consequently, users find it helpful to "have an overview on KPIs and to be able to compare oneself to other teammates" (U6). "Feedback is the most simple and effective, yet neglected recipe for organizations" (CD2). PA automate feedback processes as employees receive an assessment of their performance (U6). Moreover, the *social status* can be a benefit, too, as "employees can receive recognition, new job titles and prestige when using PA" (C3).

Next to these personal benefits, we find some very practical and technology-oriented benefits. PA provide *time savings* [35] as well as possible *automation of workflows* [36]. They bring "causality in the reduction to the objects of business management" (CD2). For example, if a tool stores user behavior, it could proactively suggest typical actions to save them time. Practitioners agree that PA serve as a "single point of truth" (U6) and thereby "make the life of managers and employees easier" (C2). Once implemented, PA help to "save money and time when dealing with data that are already there" (U4). Another motivating factor can be *information quality*, as Mettler and Winter have described [32]. All of these examples could be summarized as improving the work efficiency of the employee. Furthermore, Princi and Krämer [37] have outlined how *improved functionality* in the form of an increased rescue value of a smart monitoring system could represent a benefit.

Finally, an important benefit of PA lies in their potential to *initiate change*, which means that, based on shared data, change processes can be driven bottom-up: "If employees do not see in the medium and long term that something is happening based on these results, then this will also be reflected in the participation rates" (U2), whereas if "perceived relevance is high, it is one of the most important drivers" (U9). Thus, it is crucial to say "we have heard you and we base our decisions on this employee assessment" (C5). This also drives hedonic motivation when using the tool (D2).

In comparison to values, benefits are more tangible and relate to system features. In the logic of task-technology-fit, such benefits are essential to provide a good user experience, which can help overcome perceived risks during the usage [38].

Incentives. Third, incentives are external to the utilized tool and may be applied independently in the usage phase. Through managerial intervention, employees can be convinced to share data independent of the values and benefits of the tool. A typical example are monetary rewards [39]. Even if incentives cannot reverse existing privacy concerns that emerge from certain design characteristics, they can make the usage in the post-design phase more attractive.

Though we consulted many managers in charge of implementing PA, the majority of the interviewees struggled to think about possible incentives to facilitate data sharing. Some ideas touched upon *monetary rewards,* but interviewees "find it difficult to think about anything else, though there might be more [incentives]" (U1). The same is true for the literature, which almost exclusively discusses the characteristic of incentivizing by providing direct compensation or monetary rewards [34, 39–41]. Experts agree that monetary compensations, which are very popular in the consumer context, are perceived as problematical at the workplace: "It's hard to set a real incentive, you can say, you'll get a 50 € Amazon voucher if you fill your skills profile, but of course that doesn't work" (U5). Monetary incentives could actually worsen the situation as "the tool might be perceived even more negatively" (U5). So direct monetary rewards may need to be carefully managed in the workplace.

Alternatively, employees could be incentivized by providing explanations or *mental support by their manager* [42]. If managers act as role models, saying "I also enter my data, there is nothing bad about it" (U4), employees are more open towards sharing their data. One could imagine a skill management system that is used by managers to specifically support employees that rank lower. Next, *gamification* models, such as bonus points that can be collected, are assessed positively by the interviewees (C3). However, gamification remained a hypothetical incentive that has not yet been broadly applied. Lastly, managers handle PA by incentivizing employees to share their data in return for individual *training & coaching* opportunities. With these, the idea is to "give the individual something back for themselves, but not in the form of money" (U8). In one case, shared employee data were used for a so-called "potential conference" where future organizational leaders were promoted based on these insights. Without revealing their data, employees could not be considered in the promotion (U6).

In conclusion, the measures in the usage phase in terms of incentives are scarcely covered. Our empirical findings highlight that current efforts on increasing employees' willingness to share are mostly expressed in the design phase but the daily interaction with the tools has not yet been questioned.

Taxonomy. The resulting taxonomy below (see Table 3) is a synthesis of the findings from literature and interviews. It categorizes all above-described examples in the dimensions of *values*, *benefits*, and *incentives*.

Overall, it is our aim to provide an initial taxonomy for the field of handling employee privacy concerns at the workplace with appeal strategies, as, to the best of our knowledge, no prior works exist. Therefore, the taxonomy is meant to be a stepping stone for future research and needs to be evaluated further.

We created it by first compiling the results from literature and building an initial taxonomy. Then, we supplemented it with the results from the interviews. Mostly, the findings from the interviews overlapped with those from literature. Three examples were added after the interviews, namely the benefit *change initiation*, as well as the incentives *gamification* and *training & coaching*. To assign them to a dimension, we deductively matched them based on our underlying model.

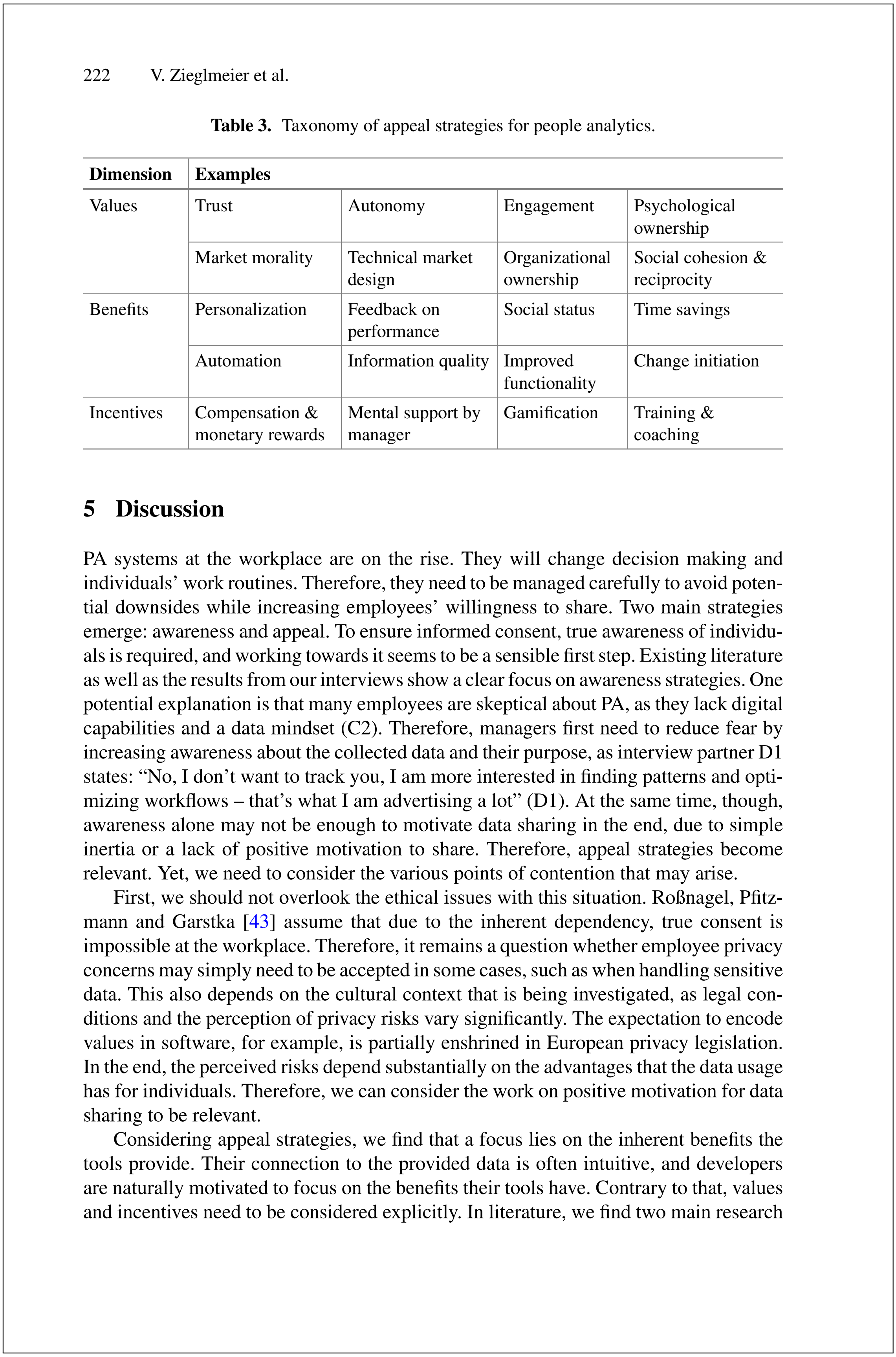


Table 3. Taxonomy of appeal strategies for people analytics.

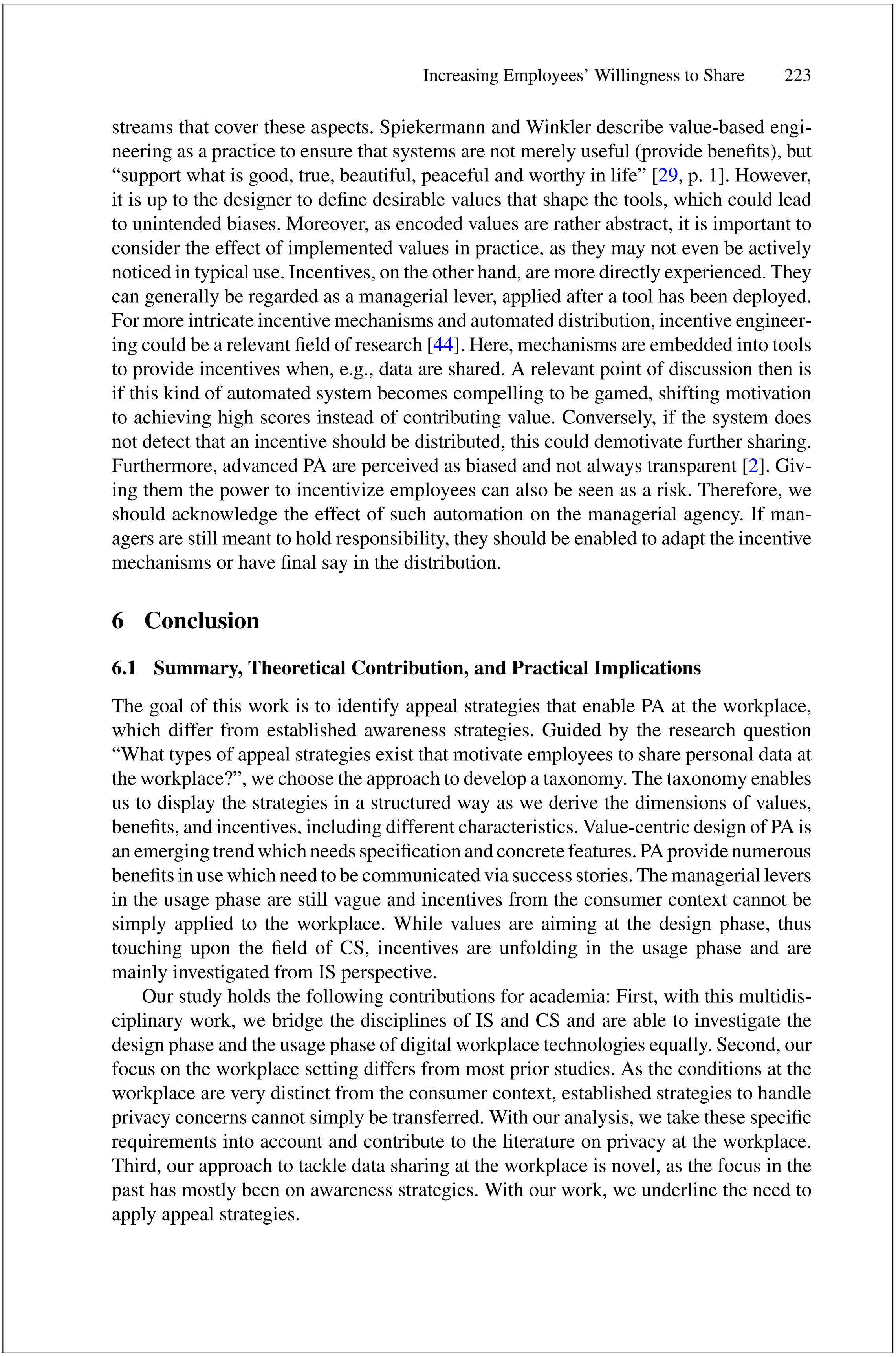
Dimension	Examples			
Values	Trust	Autonomy	Engagement	Psychological ownership
	Market morality	Technical market design	Organizational ownership	Social cohesion & reciprocity
Benefits	Personalization	Feedback on performance	Social status	Time savings
	Automation	Information quality	Improved functionality	Change initiation
Incentives	Compensation & monetary rewards	Mental support by manager	Gamification	Training & coaching

5 Discussion

PA systems at the workplace are on the rise. They will change decision making and individuals' work routines. Therefore, they need to be managed carefully to avoid potential downsides while increasing employees' willingness to share. Two main strategies emerge: awareness and appeal. To ensure informed consent, true awareness of individuals is required, and working towards it seems to be a sensible first step. Existing literature as well as the results from our interviews show a clear focus on awareness strategies. One potential explanation is that many employees are skeptical about PA, as they lack digital capabilities and a data mindset (C2). Therefore, managers first need to reduce fear by increasing awareness about the collected data and their purpose, as interview partner D1 states: "No, I don't want to track you, I am more interested in finding patterns and optimizing workflows – that's what I am advertising a lot" (D1). At the same time, though, awareness alone may not be enough to motivate data sharing in the end, due to simple inertia or a lack of positive motivation to share. Therefore, appeal strategies become relevant. Yet, we need to consider the various points of contention that may arise.

First, we should not overlook the ethical issues with this situation. Roßnagel, Pfitzmann and Garstka [43] assume that due to the inherent dependency, true consent is impossible at the workplace. Therefore, it remains a question whether employee privacy concerns may simply need to be accepted in some cases, such as when handling sensitive data. This also depends on the cultural context that is being investigated, as legal conditions and the perception of privacy risks vary significantly. The expectation to encode values in software, for example, is partially enshrined in European privacy legislation. In the end, the perceived risks depend substantially on the advantages that the data usage has for individuals. Therefore, we can consider the work on positive motivation for data sharing to be relevant.

Considering appeal strategies, we find that a focus lies on the inherent benefits the tools provide. Their connection to the provided data is often intuitive, and developers are naturally motivated to focus on the benefits their tools have. Contrary to that, values and incentives need to be considered explicitly. In literature, we find two main research



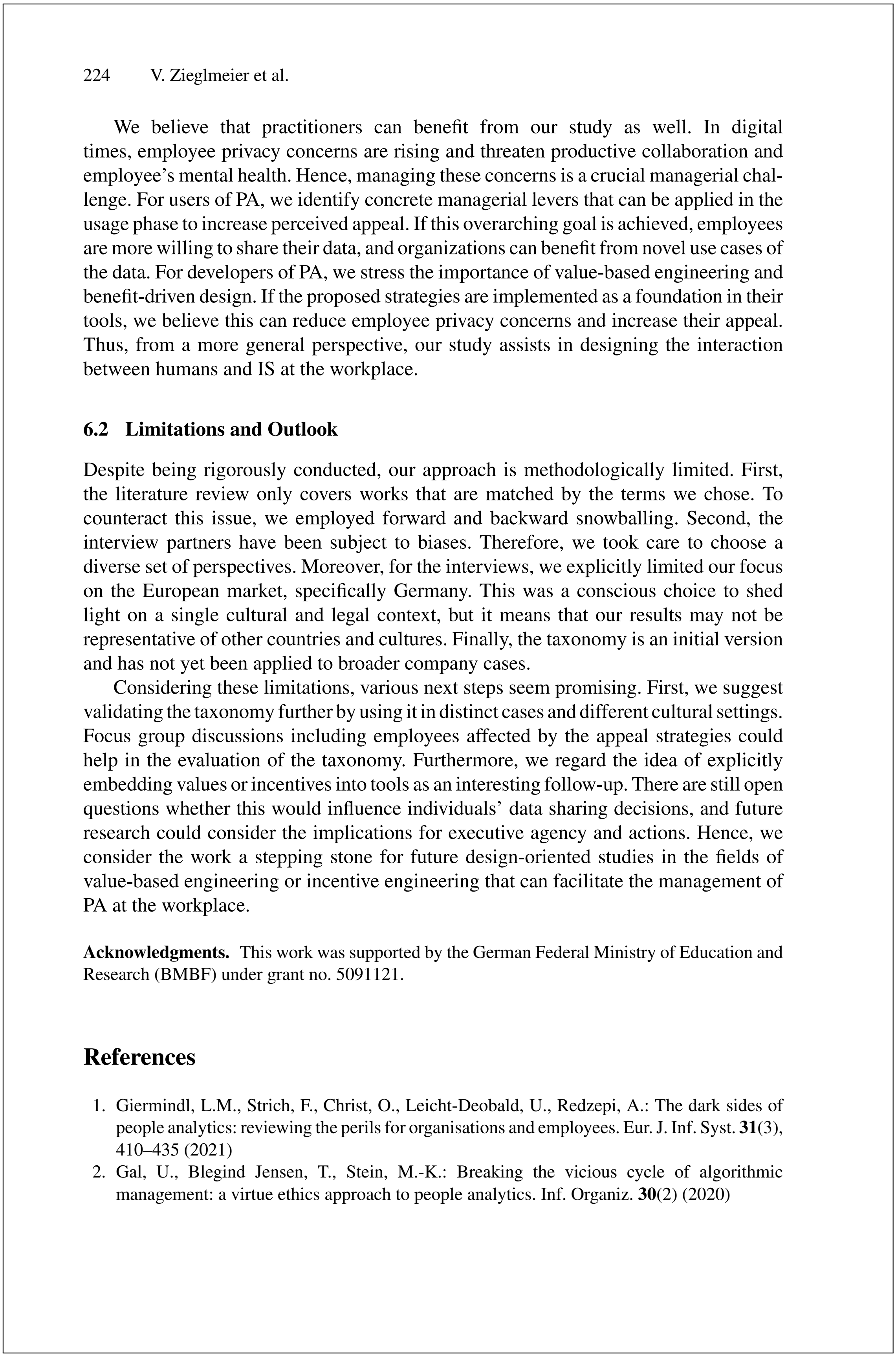
streams that cover these aspects. Spiekermann and Winkler describe value-based engineering as a practice to ensure that systems are not merely useful (provide benefits), but "support what is good, true, beautiful, peaceful and worthy in life" [29, p. 1]. However, it is up to the designer to define desirable values that shape the tools, which could lead to unintended biases. Moreover, as encoded values are rather abstract, it is important to consider the effect of implemented values in practice, as they may not even be actively noticed in typical use. Incentives, on the other hand, are more directly experienced. They can generally be regarded as a managerial lever, applied after a tool has been deployed. For more intricate incentive mechanisms and automated distribution, incentive engineering could be a relevant field of research [44]. Here, mechanisms are embedded into tools to provide incentives when, e.g., data are shared. A relevant point of discussion then is if this kind of automated system becomes compelling to be gamed, shifting motivation to achieving high scores instead of contributing value. Conversely, if the system does not detect that an incentive should be distributed, this could demotivate further sharing. Furthermore, advanced PA are perceived as biased and not always transparent [2]. Giving them the power to incentivize employees can also be seen as a risk. Therefore, we should acknowledge the effect of such automation on the managerial agency. If managers are still meant to hold responsibility, they should be enabled to adapt the incentive mechanisms or have final say in the distribution.

6 Conclusion

6.1 Summary, Theoretical Contribution, and Practical Implications

The goal of this work is to identify appeal strategies that enable PA at the workplace, which differ from established awareness strategies. Guided by the research question "What types of appeal strategies exist that motivate employees to share personal data at the workplace?", we choose the approach to develop a taxonomy. The taxonomy enables us to display the strategies in a structured way as we derive the dimensions of values, benefits, and incentives, including different characteristics. Value-centric design of PA is an emerging trend which needs specification and concrete features. PA provide numerous benefits in use which need to be communicated via success stories. The managerial levers in the usage phase are still vague and incentives from the consumer context cannot be simply applied to the workplace. While values are aiming at the design phase, thus touching upon the field of CS, incentives are unfolding in the usage phase and are mainly investigated from IS perspective.

Our study holds the following contributions for academia: First, with this multidisciplinary work, we bridge the disciplines of IS and CS and are able to investigate the design phase and the usage phase of digital workplace technologies equally. Second, our focus on the workplace setting differs from most prior studies. As the conditions at the workplace are very distinct from the consumer context, established strategies to handle privacy concerns cannot simply be transferred. With our analysis, we take these specific requirements into account and contribute to the literature on privacy at the workplace. Third, our approach to tackle data sharing at the workplace is novel, as the focus in the past has mostly been on awareness strategies. With our work, we underline the need to apply appeal strategies.



We believe that practitioners can benefit from our study as well. In digital times, employee privacy concerns are rising and threaten productive collaboration and employee's mental health. Hence, managing these concerns is a crucial managerial challenge. For users of PA, we identify concrete managerial levers that can be applied in the usage phase to increase perceived appeal. If this overarching goal is achieved, employees are more willing to share their data, and organizations can benefit from novel use cases of the data. For developers of PA, we stress the importance of value-based engineering and benefit-driven design. If the proposed strategies are implemented as a foundation in their tools, we believe this can reduce employee privacy concerns and increase their appeal. Thus, from a more general perspective, our study assists in designing the interaction between humans and IS at the workplace.

6.2 Limitations and Outlook

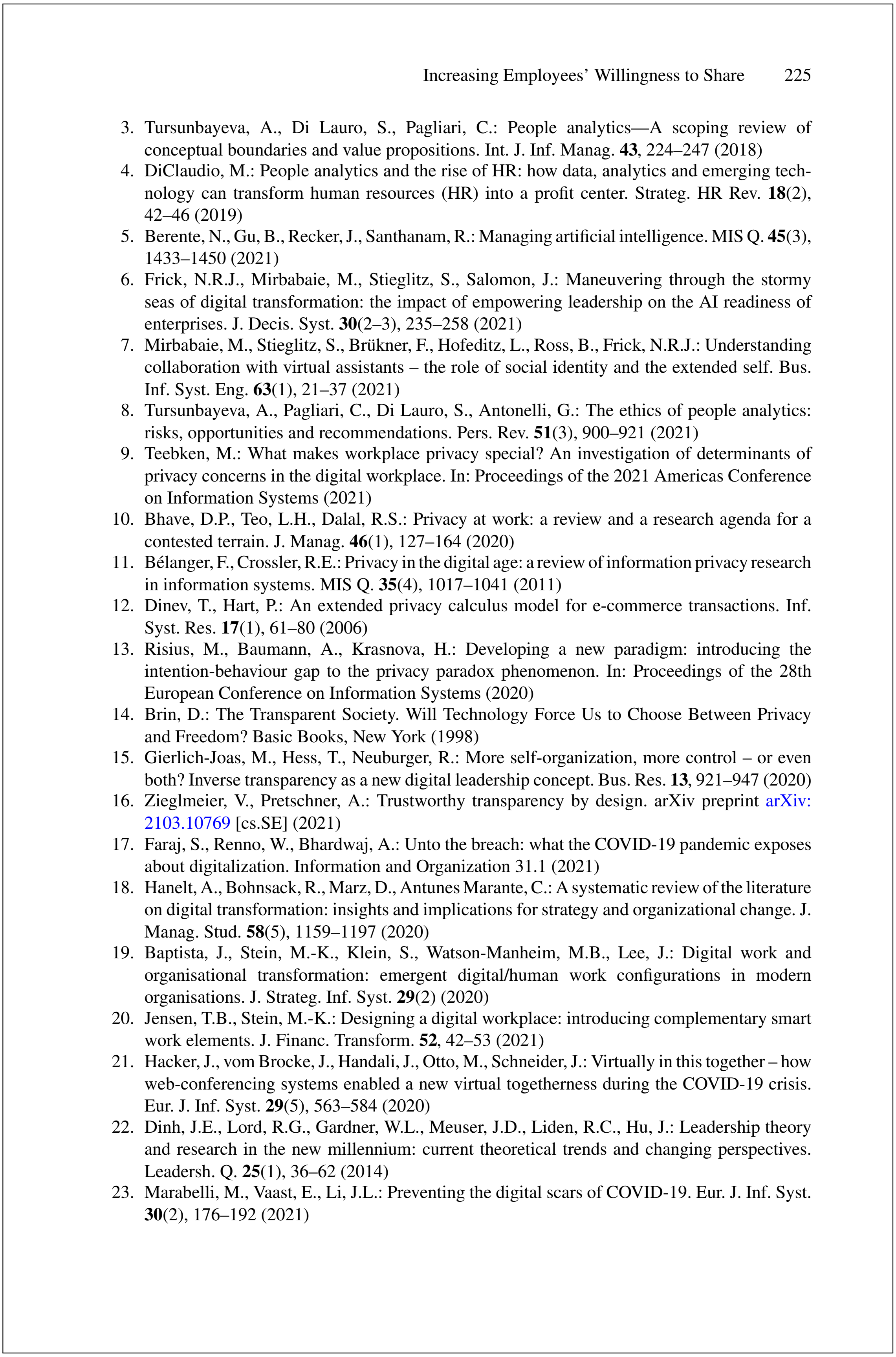
Despite being rigorously conducted, our approach is methodologically limited. First, the literature review only covers works that are matched by the terms we chose. To counteract this issue, we employed forward and backward snowballing. Second, the interview partners have been subject to biases. Therefore, we took care to choose a diverse set of perspectives. Moreover, for the interviews, we explicitly limited our focus on the European market, specifically Germany. This was a conscious choice to shed light on a single cultural and legal context, but it means that our results may not be representative of other countries and cultures. Finally, the taxonomy is an initial version and has not yet been applied to broader company cases.

Considering these limitations, various next steps seem promising. First, we suggest validating the taxonomy further by using it in distinct cases and different cultural settings. Focus group discussions including employees affected by the appeal strategies could help in the evaluation of the taxonomy. Furthermore, we regard the idea of explicitly embedding values or incentives into tools as an interesting follow-up. There are still open questions whether this would influence individuals' data sharing decisions, and future research could consider the implications for executive agency and actions. Hence, we consider the work a stepping stone for future design-oriented studies in the fields of value-based engineering or incentive engineering that can facilitate the management of PA at the workplace.

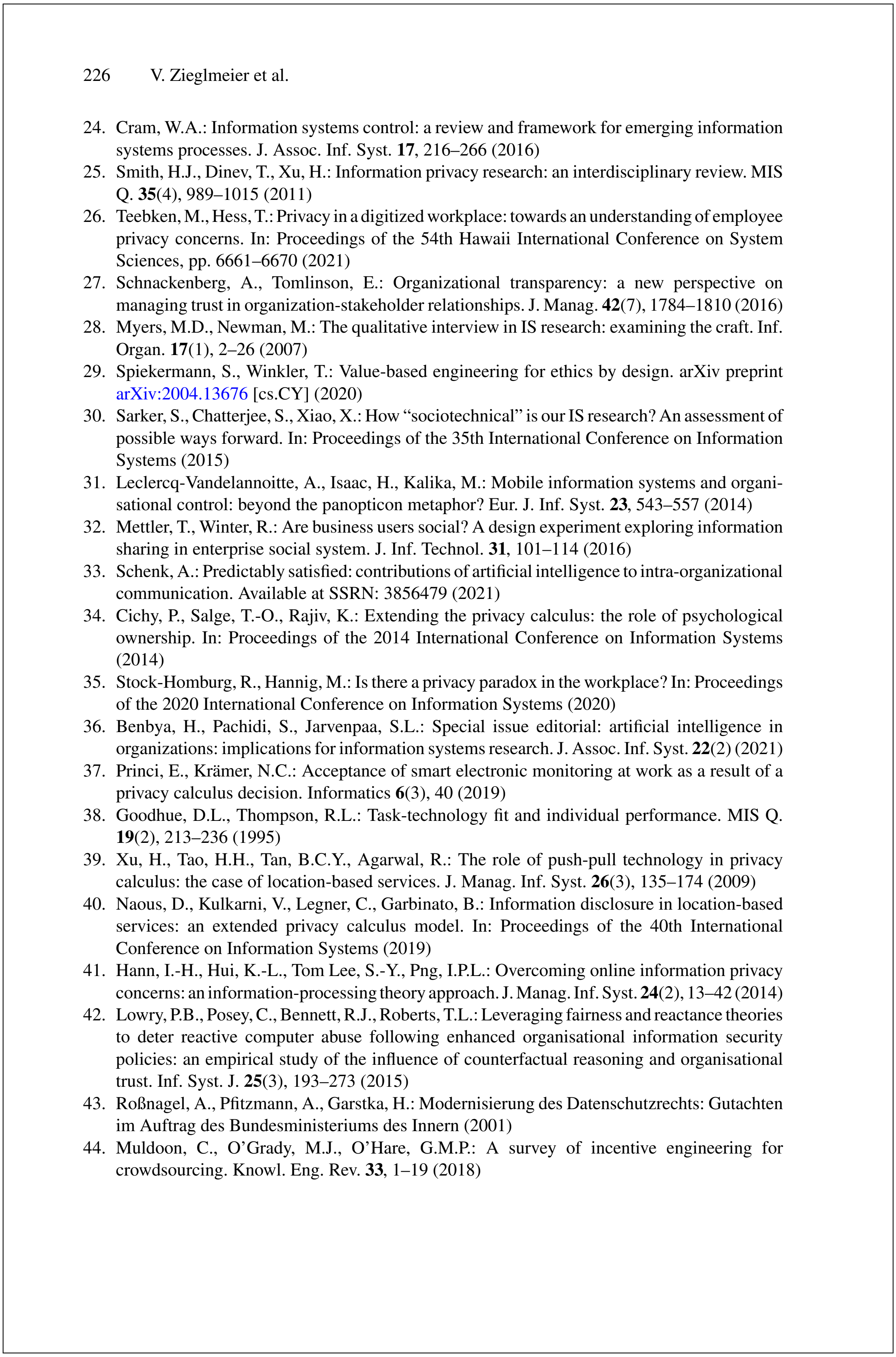
Acknowledgments. This work was supported by the German Federal Ministry of Education and Research (BMBF) under grant no. 5091121.

References

1. Giermindl, L.M., Strich, F., Christ, O., Leicht-Deobald, U., Redzepi, A.: The dark sides of people analytics: reviewing the perils for organisations and employees. Eur. J. Inf. Syst. **31**(3), 410–435 (2021)
2. Gal, U., Blegind Jensen, T., Stein, M.-K.: Breaking the vicious cycle of algorithmic management: a virtue ethics approach to people analytics. Inf. Organiz. **30**(2) (2020)



3. Tursunbayeva, A., Di Lauro, S., Pagliari, C.: People analytics—A scoping review of conceptual boundaries and value propositions. Int. J. Inf. Manag. **43**, 224–247 (2018)
4. DiClaudio, M.: People analytics and the rise of HR: how data, analytics and emerging technology can transform human resources (HR) into a profit center. Strateg. HR Rev. **18**(2), 42–46 (2019)
5. Berente, N., Gu, B., Recker, J., Santhanam, R.: Managing artificial intelligence. MIS Q. **45**(3), 1433–1450 (2021)
6. Frick, N.R.J., Mirbabaie, M., Stieglitz, S., Salomon, J.: Maneuvering through the stormy seas of digital transformation: the impact of empowering leadership on the AI readiness of enterprises. J. Decis. Syst. **30**(2–3), 235–258 (2021)
7. Mirbabaie, M., Stieglitz, S., Brükner, F., Hofeditz, L., Ross, B., Frick, N.R.J.: Understanding collaboration with virtual assistants – the role of social identity and the extended self. Bus. Inf. Syst. Eng. **63**(1), 21–37 (2021)
8. Tursunbayeva, A., Pagliari, C., Di Lauro, S., Antonelli, G.: The ethics of people analytics: risks, opportunities and recommendations. Pers. Rev. **51**(3), 900–921 (2021)
9. Teebken, M.: What makes workplace privacy special? An investigation of determinants of privacy concerns in the digital workplace. In: Proceedings of the 2021 Americas Conference on Information Systems (2021)
10. Bhave, D.P., Teo, L.H., Dalal, R.S.: Privacy at work: a review and a research agenda for a contested terrain. J. Manag. **46**(1), 127–164 (2020)
11. Bélanger, F., Crossler, R.E.: Privacy in the digital age: a review of information privacy research in information systems. MIS Q. **35**(4), 1017–1041 (2011)
12. Dinev, T., Hart, P.: An extended privacy calculus model for e-commerce transactions. Inf. Syst. Res. **17**(1), 61–80 (2006)
13. Risius, M., Baumann, A., Krasnova, H.: Developing a new paradigm: introducing the intention-behaviour gap to the privacy paradox phenomenon. In: Proceedings of the 28th European Conference on Information Systems (2020)
14. Brin, D.: The Transparent Society. Will Technology Force Us to Choose Between Privacy and Freedom? Basic Books, New York (1998)
15. Gierlich-Joas, M., Hess, T., Neuburger, R.: More self-organization, more control – or even both? Inverse transparency as a new digital leadership concept. Bus. Res. **13**, 921–947 (2020)
16. Zieglmeier, V., Pretschner, A.: Trustworthy transparency by design. arXiv preprint arXiv: 2103.10769 [cs.SE] (2021)
17. Faraj, S., Renno, W., Bhardwaj, A.: Unto the breach: what the COVID-19 pandemic exposes about digitalization. Information and Organization 31.1 (2021)
18. Hanelt, A., Bohnsack, R., Marz, D., Antunes Marante, C.: A systematic review of the literature on digital transformation: insights and implications for strategy and organizational change. J. Manag. Stud. **58**(5), 1159–1197 (2020)
19. Baptista, J., Stein, M.-K., Klein, S., Watson-Manheim, M.B., Lee, J.: Digital work and organisational transformation: emergent digital/human work configurations in modern organisations. J. Strateg. Inf. Syst. **29**(2) (2020)
20. Jensen, T.B., Stein, M.-K.: Designing a digital workplace: introducing complementary smart work elements. J. Financ. Transform. **52**, 42–53 (2021)
21. Hacker, J., vom Brocke, J., Handali, J., Otto, M., Schneider, J.: Virtually in this together – how web-conferencing systems enabled a new virtual togetherness during the COVID-19 crisis. Eur. J. Inf. Syst. **29**(5), 563–584 (2020)
22. Dinh, J.E., Lord, R.G., Gardner, W.L., Meuser, J.D., Liden, R.C., Hu, J.: Leadership theory and research in the new millennium: current theoretical trends and changing perspectives. Leadersh. Q. **25**(1), 36–62 (2014)
23. Marabelli, M., Vaast, E., Li, J.L.: Preventing the digital scars of COVID-19. Eur. J. Inf. Syst. **30**(2), 176–192 (2021)



24. Cram, W.A.: Information systems control: a review and framework for emerging information systems processes. J. Assoc. Inf. Syst. **17**, 216–266 (2016)
25. Smith, H.J., Dinev, T., Xu, H.: Information privacy research: an interdisciplinary review. MIS Q. **35**(4), 989–1015 (2011)
26. Teebken, M., Hess, T.: Privacy in a digitized workplace: towards an understanding of employee privacy concerns. In: Proceedings of the 54th Hawaii International Conference on System Sciences, pp. 6661–6670 (2021)
27. Schnackenberg, A., Tomlinson, E.: Organizational transparency: a new perspective on managing trust in organization-stakeholder relationships. J. Manag. **42**(7), 1784–1810 (2016)
28. Myers, M.D., Newman, M.: The qualitative interview in IS research: examining the craft. Inf. Organ. **17**(1), 2–26 (2007)
29. Spiekermann, S., Winkler, T.: Value-based engineering for ethics by design. arXiv preprint arXiv:2004.13676 [cs.CY] (2020)
30. Sarker, S., Chatterjee, S., Xiao, X.: How "sociotechnical" is our IS research? An assessment of possible ways forward. In: Proceedings of the 35th International Conference on Information Systems (2015)
31. Leclercq-Vandelannoitte, A., Isaac, H., Kalika, M.: Mobile information systems and organisational control: beyond the panopticon metaphor? Eur. J. Inf. Syst. **23**, 543–557 (2014)
32. Mettler, T., Winter, R.: Are business users social? A design experiment exploring information sharing in enterprise social system. J. Inf. Technol. **31**, 101–114 (2016)
33. Schenk, A.: Predictably satisfied: contributions of artificial intelligence to intra-organizational communication. Available at SSRN: 3856479 (2021)
34. Cichy, P., Salge, T.-O., Rajiv, K.: Extending the privacy calculus: the role of psychological ownership. In: Proceedings of the 2014 International Conference on Information Systems (2014)
35. Stock-Homburg, R., Hannig, M.: Is there a privacy paradox in the workplace? In: Proceedings of the 2020 International Conference on Information Systems (2020)
36. Benbya, H., Pachidi, S., Jarvenpaa, S.L.: Special issue editorial: artificial intelligence in organizations: implications for information systems research. J. Assoc. Inf. Syst. **22**(2) (2021)
37. Princi, E., Krämer, N.C.: Acceptance of smart electronic monitoring at work as a result of a privacy calculus decision. Informatics **6**(3), 40 (2019)
38. Goodhue, D.L., Thompson, R.L.: Task-technology fit and individual performance. MIS Q. **19**(2), 213–236 (1995)
39. Xu, H., Tao, H.H., Tan, B.C.Y., Agarwal, R.: The role of push-pull technology in privacy calculus: the case of location-based services. J. Manag. Inf. Syst. **26**(3), 135–174 (2009)
40. Naous, D., Kulkarni, V., Legner, C., Garbinato, B.: Information disclosure in location-based services: an extended privacy calculus model. In: Proceedings of the 40th International Conference on Information Systems (2019)
41. Hann, I.-H., Hui, K.-L., Tom Lee, S.-Y., Png, I.P.L.: Overcoming online information privacy concerns: an information-processing theory approach. J. Manag. Inf. Syst. **24**(2), 13–42 (2014)
42. Lowry, P.B., Posey, C., Bennett, R.J., Roberts, T.L.: Leveraging fairness and reactance theories to deter reactive computer abuse following enhanced organisational information security policies: an empirical study of the influence of counterfactual reasoning and organisational trust. Inf. Syst. J. **25**(3), 193–273 (2015)
43. Roßnagel, A., Pfitzmann, A., Garstka, H.: Modernisierung des Datenschutzrechts: Gutachten im Auftrag des Bundesministeriums des Innern (2001)
44. Muldoon, C., O'Grady, M.J., O'Hare, G.M.P.: A survey of incentive engineering for crowdsourcing. Knowl. Eng. Rev. **33**, 1–19 (2018)

Part III.

Conclusion

8. Limitations

This chapter discusses overarching limitations of our thesis, including of our conceptual goal, technical approaches, and empirical studies. Limitations that only affect individual publications are discussed in the respective chapter.

8.1. Limitations of Data Sovereignty

To start with, all our contributions aim towards more data sovereignty for employees. This assumes that the individual, given sufficient transparency and control, can make sovereign decisions about sharing their data. Yet, external social pressures can impact their decision. To start with, in the workplace, an inherent power asymmetry exists between managers and employees. Even if the individual considers a data usage to be problematic, they may feel forced to share due to a fear of negative consequences [see, e.g., 42]. Similarly, social pressure from colleagues can push the individual to conform and, e.g., contribute data [178, Sec. 4.4]. In such situations, even seemingly empowering mechanisms, such as transparency or control tools, are not effective. Seberger et al. find that these mechanisms may, conversely, even lead to resignation and acceptance of privacy violations [204, Sec. 4.5]. For some use cases, this issue could be mitigated by providing the individual with plausible deniability regarding their sharing of data [57, Sec. 4.5]; (see also Section 2.7). This is not always possible, though.

8.2. Limitations of Inverse Transparency

We furthermore present contributions towards providing inverse transparency in PA. That means, giving data owners direct oversight of how their data are used. We assume that those data usage processes that directly depend on an individual's own data are most relevant to them. This includes derived data if they can be linked back to the individual. Yet, this transparency may not suffice. Bertino [28] notes that even if an individual's own data are not used, data processing and decision-making may affect them. For example, a recommender system could base its decision on a population classification that matches the individual, without ever using their specific data sample [28, p. 68]. If the used data are generic enough, though, they would be considered anonymous [176, p. 2]. Therefore, depending on the aims of the system, even anonymous data may be considered relevant for the individual if they match their profile. Furthermore, transparency should also be provided about data collection, provisioning, policies, and processing algorithms [28, p. 67].

Conversely, transparency may not be a means to an end, as there can be too much transparency as well [see 126]. One of our study participants highlighted this fact, explaining that the transparency made them more aware of being analyzed [1, p. 19]. It may therefore be necessary to consider how to filter the logged data usages before displaying them to the individual. For example, techniques from anomaly detection [e.g., 206] could be applied to determine the most critical events.

Finally, by its nature, inverse transparency makes data consumers' actions transparent. This is necessary to a degree but, depending on how fine-grained the provided data are, can also lead to privacy issues for them, such as surveillance [178, Sec. 4.2]. As one example, Polst and Feth note that the data owner could observe data consumers' working hours through the timestamps associated with the usage logs [178, p. 5]. This factor and its implications are an important second-order effect of inverse transparency that may only emerge in long-term studies.

Therefore, we recognize that inverse transparency as an ideal has limitations in practice. If possible, data analysis approaches that do not require the use of personal data may be preferable [e.g., 55, 156].

8.3. Undetectable Data Usage

Even with every PA tool being developed with inverse transparency by design, some data usages cannot be tracked. For example, data consumers can always take a picture of their screen to make an undetectable "copy" [186, pp. 46 f.]. In addition, not all data usage happens digitally. Information may be shared in person or through physical means, such as whiteboards. Even if this happens without malicious intent, these data usages are undetectable with a digital tracking infrastructure. We recognize this issue but do not consider it severe. Theoretically, potentially sensitive information about an individual could be "shared" this way. But the media breaks make large-scale copying or exporting of data infeasible [1, p. 9]. And more advanced analyses rely on large datasets and up-to-date information, which limits the misusage potential of data gathered this way.

8.4. Solution Practicability

Beyond conceptual issues, we furthermore need to critically examine the practicability of our solutions.

First, is it realistic to expect all PA to be built with inverse transparency by design (Chapter 4)? In practice, this ideal may be unattainable. When considering individual companies, though, we find that they use only few PA [see, e.g., 67]. In the tool acquisition process, larger companies already request custom changes and features from tool vendors. Therefore, companies buying PA may be able to enforce the inclusion of inverse transparency for their individual case.

Second, is a fully decentralized approach to inverse transparency (Chapter 5) tangible for employees? The concept of decentralization is itself hard to grasp and can be understood differently [see 39]. Furthermore, most employees probably trust their employer, calling into question the necessity to bear the technical disadvantages of a fully decentralized approach. Still, we consider it important to propose a solution for sensitive data, as they warrant heightened information security.

Third, how relevant is the trustworthiness of a user interface (Chapter 6) in light of other requirements? Design implicates trade-offs. For usability, the measures we identify to improve trustworthiness are often compatible or identical, as usability is a trustworthiness factor [3, Tab. 2]. Yet, other non-functional requirements can conflict in practice [147]. Therefore, working towards maximizing trustworthiness may reduce the practicality of a tool. This goal should therefore be considered in the context of the concrete usage scenario to ensure its feasibility.

Fourth, are appeal strategies for PA (Chapter 7) viable in practice? In our own follow-up research, we find mixed results for the inclusion of data owner benefits in PA [see 9]. The appreciation for the provided benefits varies noticeably between study participants [9, Sec. 4.3.5]. This is in line with research on incentives, which highlights the importance of considering individual personality traits to determine effective incentives [e.g., 131]. Thus, appeal strategies need to be tailored to the individual. This may be difficult to realize, though. Additionally, personalized incentives could be perceived as unfair by employees. Some appeal strategies may therefore simply be impractical.

8.5. Preliminary Evaluations

Finally, while we designed and conducted our empirical studies to the best of our abilities, they are preliminary with limited practical significance. To start with, our studies on inverse transparency by design (Chapter 4) were conducted in a laboratory setting with students, to support internal validity and establish causality. This may limit external validity, though. Students often lack knowledge of the workplace context due to limited personal experience. While we tried to model our studies as realistically as possible, the inherent complexity of the workplace cannot be perfectly reconstructed in an artificial setting. Similarly, all our empirical studies with human participants (Chapters 4 and 6) were small with between 12–15 people. This allowed us to perform in-depth qualitative analyses. Yet, it means that the samples are not representative, which limits the significance of quantitative analyses. Follow-up research could build on our theories to conduct broader field studies. Ideally, these should be situated in a real workplace, with employees working with inverse transparent PA in their daily work. While such a study design is costly and time-consuming, it can further improve our understanding of the suitability of the developed solutions.

9. Summary and Outlook

PA become increasingly pervasive in the workplace. More available data enable advanced use cases that promise to improve, e.g., decision-making or employees' job satisfaction. Currently available PA exhibit three problems for individual employees in their design, though: an elevated *risk of data misusage* (problem **I**), a *lack of transparency* (problem **II**), and a *missing appeal* (problem **III**). We suggest addressing these problems by designing PA for data sovereigns, which means providing inverse transparency and increasing the appeal for employees.

In this thesis, we present four contributions towards this goal. *First*, addressing problems **I** and **II**, we propose a rethink of PA by incorporating inverse transparency by design. This idea represents the foundation of our approach to improve PA design. We give a comprehensive overview of the theoretical requirements and potential implications. In two empirical studies, we find that the approach is judged as practical by developers and positively received by users. *Second*, also addressing problem **I**, we present a solution for fully decentralized inverse transparency. Thereby, we provide an approach for the analysis of sensitive data that are stored on individuals' personal devices. Our developed tool enables inverse transparency when these data are shared directly with data consumers. The tool is secure against expected attacks and GDPR-compliant. Its performance is reasonable and it scales linearly. *Third*, addressing problem **II**, we provide concrete measures to improve the trustworthiness of a transparency dashboard. This facet is important to improve the acceptance and usage of such a tool. When applying the identified measures to a proof of concept dashboard, we find that they can foster trust in users. *Fourth*, addressing problem **III**, we present a taxonomy of appeal strategies for PA. These guide the design of PA towards incorporating data owners as stakeholders, which can improve employees' willingness to share their data. We present concrete examples for both the design and usage phases, and furthermore derive overarching strategy dimensions.

Our contributions are subject to limitations. The power asymmetry and social pressures in the workplace affect an individual, which limits their sovereignty. Additionally, our theoretical goal of inverse transparency may be conceptually limited and needs to be implemented carefully. Even then, some data usages may never be detectable. Beyond such conceptual issues, our solutions are also limited in their practicability. While we can imagine paths to realize them in practice, compromises need to be made. Finally, our evaluation results are promising, showing unanimous support for the idea of inverse transparency, but our empirical studies are preliminary. Field studies could help to assess the external validity of our results.

As next steps, we suggest: (1) Build a complete PA tool with inverse transparency by design, ideally with an industrial partner. Include all relevant stakeholders, notably data owners, in the design process. (2) Integrate appeal strategies into the PA tool that can be tailored to the individual. Consider specificities of the cultural and company context.

(3) Iteratively develop a trustworthy and usable privacy dashboard that integrates inverse transparency. This should be a unified view of the data the company stores on an individual, the permissions they gave to use their data, as well as a prioritized and filtered list of tracked data usages. (4) Deploy the PA tool with inverse transparency and incorporated appeal strategies (→ 1, 2) in the field, linked to the trustworthy privacy dashboard (→ 3). Then, perform a long-term study on the effects of inverse transparency, appeal strategies, and a privacy dashboard in practice.

To conclude, our contributions provide concrete recommendations for designing PA for data sovereigns. Our studies support their suitability for our limited researched context and indicate relevant directions for further research. We hope our work serves as inspiration not only for further research, but importantly also for the design of PA in practice.

Bibliography

First, our own publications in order of their relationship to this thesis (core publications, related publications). Then, the remaining references in alphabetical order.

Own Publications

[1] Valentin Zieglmeier and Alexander Pretschner. "Rethinking People Analytics With Inverse Transparency by Design." *Proceedings of the ACM on Human-Computer Interaction* 7.CSCW2, Article 292 (2023), pp. 1–29. DOI: 10.1145/3610083.

[2] Valentin Zieglmeier, Gabriel Loyola Daiqui, and Alexander Pretschner. "Decentralized Inverse Transparency With Blockchain." *Distributed Ledger Technologies: Research and Practice* 2.3, Article 17 (2023), pp. 1–28. DOI: 10.1145/3592624.

[3] Valentin Zieglmeier and Antonia Maria Lehene. "Designing Trustworthy User Interfaces." In: *Proceedings of the 33rd Australian Conference on Human-Computer Interaction*. ACM. 2021, pp. 182–189. DOI: 10.1145/3520495.3520525.

[4] Valentin Zieglmeier, Maren Gierlich-Joas, and Alexander Pretschner. "Increasing Employees' Willingness to Share: Introducing Appeal Strategies for People Analytics." In: *Proceedings of the 13th International Conference on Software Business*. Lecture Notes in Business Information Processing 463. Springer, 2022, pp. 213–226. DOI: 10.1007/978-3-031-20706-8_15.

[5] Valentin Zieglmeier. "The Inverse Transparency Toolchain: A Fully Integrated and Quickly Deployable Data Usage Logging Infrastructure." *Software Impacts* 17, Article 100554 (September 2023), pp. 1–4. DOI: 10.1016/j.simpa.2023.100554.

[6] Valentin Zieglmeier and Gabriel Loyola Daiqui. "GDPR-Compliant Use of Blockchain for Secure Usage Logs." In: *Proceedings of the 25th International Conference on Evaluation and Assessment in Software Engineering*. ACM. 2021, pp. 313–320. DOI: 10.1145/3463274.3463349.

[7] Valentin Zieglmeier. *Appending Data to Blockchain Is Not Sufficient for Non-repudiation of Receipt*. 2023. arXiv: 2308.04781 [cs.CR].

[8] Maren Gierlich-Joas, Valentin Zieglmeier, Rahild Neuburger, and Thomas Hess. "Leading Agents or Stewards? Exploring Design Principles for Empowerment Through Workplace Technologies." In: *Proceedings of the 42nd International Conference on Information Systems*. AIS. 2021, Article 1519. URL: https://aisel.aisnet.org/icis2021/is_future_work/is_future_work/7.

[9] Patrik Zander and Valentin Zieglmeier. "Data Owner Benefit-Driven Design of People Analytics." *Proceedings of the ACM on Human-Computer Interaction* 7.EICS, Article 173 (2023), pp. 1–38. DOI: 10.1145/3593225.

Other Publications

[10] Rafael Accorsi. "BBox: A Distributed Secure Log Architecture." In: *Proceedings of the 2010 European Public Key Infrastructure Workshop*. Lecture Notes in Computer Science 6711. Springer, 2010, pp. 109–124. DOI: 10.1007/978-3-642-22633-5_8.

[11] Sam Adler-Bell and Michelle Miller. *The Datafication of Employment*. Report. The Century Foundation, 2018-12-19. URL: https://tcf.org/content/report/datafication-employment-surveillance-capitalism-shaping-workers-futures-without-knowledge/.

[12] Rakesh Agrawal, Jerry Kiernan, Ramakrishnan Srikant, and Yirong Xu. "Hippocratic Databases." In: *Proceedings of the 28th International Conference on Very Large Databases*. Elsevier, 2002, pp. 143–154. DOI: 10.1016/b978-155860869-6/50021-4.

[13] Ifeoma Ajunwa. "The 'Black Box' at Work." *Big Data & Society* 7.2 (2020), pp. 1–6. DOI: 10.1177/2053951720938093.

[14] Ifeoma Ajunwa, Kate Crawford, and Jason Schultz. "Limitless Worker Surveillance." *California Law Review* 105.3 (2017), pp. 735–776. DOI: 10.15779/Z38BR8MF94.

[15] Nasim Al Goni, Sherif Saad Ahmed, and Ahmed Ibrahim. "A P2P Optimistic Fair-Exchange (OFE) Scheme for Personal Health Records Using Blockchain Technology." In: *Proceedings of the 3rd International Conference on Wireless, Intelligent and Distributed Environment for Communication*. Lecture Notes on Data Engineering and Communications Technologies 51. Springer, 2020, pp. 1–21. DOI: 10.1007/978-3-030-44372-6_1.

[16] Alessandro Aldini and Roberto Gorrieri. "Security Analysis of a Probabilistic Non-repudiation Protocol." In: *Proceedings of the 2nd Joint International Workshop on Process Algebra and Probabilistic Methods: Performance Modeling and Verification*. Lecture Notes in Computer Science 2399. Springer, 2002, pp. 17–36. DOI: 10.1007/3-540-45605-8_3.

[17] Mike Ananny and Kate Crawford. "Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability." *New Media & Society* 20.3 (2018), pp. 973–989. DOI: 10.1177/1461444816676645.

[18] Julio Angulo, Simone Fischer-Hübner, Tobias Pulls, and Erik Wästlund. "Usable Transparency With the Data Track: A Tool for Visualizing Data Disclosures." In: *Extended Abstracts of the 2015 CHI Conference on Human Factors in Computing Systems*. ACM. 2015, pp. 1803–1808. DOI: 10.1145/2702613.2732701.

[19] Giuseppe Ateniese, Bernardo Magri, Daniele Venturi, and Ewerton Andrade. "Redactable Blockchain – or – Rewriting History in Bitcoin and Friends." In: *Proceedings of the 2nd IEEE European Symposium on Security and Privacy*. IEEE. 2017, pp. 111–126. DOI: 10.1109/eurosp.2017.37.

[20] Naveen Farag Awad and Mayuram S. Krishnan. "The Personalization Privacy Paradox: An Empirical Evaluation of Information Transparency and the Willingness to Be Profiled Online for Personalization." *MIS Quarterly* 30.1 (2006), pp. 13–28. DOI: 10.2307/25148715.

[21] Kenny Awuson-David, Tawfik Al-Hadhrami, Mamoun Alazab, Nazaraf Shah, and Andrii Shalaginov. "BCFL Logging: An Approach to Acquire and Preserve Admissible Digital Forensics Evidence in Cloud Ecosystem." *Future Generation Computer Systems* 122 (September 2021), pp. 1–13. DOI: 10.1016/j.future.2021.03.001.

[22] Hebert Azevedo-Sa, Suresh Kumaar Jayaraman, Connor T. Esterwood, X. Jessie Yang, Lionel P. Robert Jr., and Dawn M. Tilbury. "Real-Time Estimation of Drivers' Trust in Automated Driving Systems." *International Journal of Social Robotics* 13.8 (2021), pp. 1911–1927. DOI: 10.1007/s12369-020-00694-1.

[23] Eugene Bagdasaryan, Griffin Berlstein, Jason Waterman, Eleanor Birrell, Nate Foster, Fred B. Schneider, and Deborah Estrin. "Ancile: Enhancing Privacy for Ubiquitous Computing With Use-Based Privacy." In: *Proceedings of the 18th ACM Workshop on Privacy in the Electronic Society*. ACM. 2019, pp. 111–124. DOI: 10.1145/3338498.3358642.

[24] Jan Bartsch, Tobias Dehling, Florian Lauf, Sven Meister, and Ali Sunyaev. "Let the Computer Say NO! The Neglected Potential of Policy Definition Languages for Data Sovereignty." In: *Selbstbestimmung, Privatheit und Datenschutz: Gestaltungsoptionen für einen europäischen Weg*. Ed. by Michael Friedewald, Michael Kreutzer, and Marit Hansen. DuD-Fachbeiträge. Springer Vieweg, 2022, pp. 449–468. DOI: 10.1007/978-3-658-33306-5_22.

[25] Martijn H. van Beek. "Comparison of Enterprise Digital Rights Management Systems." Master's Thesis. Radboud University Nijmegen, 2007. URL: http://www.cs.ru.nl/mtl/scripties/2007/MartijnVanBeekScriptie.pdf.

[26] Mihir Bellare and Bennet Yee. *Forward Integrity for Secure Audit Logs*. Tech. rep. University of California at San Diego, 1997.

[27] Florian Bemmann, Maximiliane Windl, Jonas Erbe, Sven Mayer, and Heinrich Hussmann. "The Influence of Transparency and Control on the Willingness of Data Sharing in Adaptive Mobile Apps." *Proceedings of the ACM on Human-Computer Interaction* 6.MHCI, Article 189 (2022), pp. 1–26. DOI: 10.1145/3546724.

[28] Elisa Bertino. "The Quest for Data Transparency." *IEEE Security & Privacy* 18.3 (2020), pp. 67–68. DOI: 10.1109/msec.2020.2980593.

[29] Devasheesh P. Bhave, Laurel H. Teo, and Reeshad S. Dalal. "Privacy at Work: A Review and a Research Agenda for a Contested Terrain." *Journal of Management* 46.1 (2020), pp. 127–164. DOI: 10.1177/0149206319878254.

[30] Christoph Bier, Kay Kühne, and Jürgen Beyerer. "PrivacyInsight: The Next Generation Privacy Dashboard." In: *Proceedings of the 4th Annual Privacy Forum*. Lecture Notes in Computer Science 9857. Springer, 2016, pp. 135–152. DOI: 10.1007/978-3-319-44760-5_9.

[31] Eleanor Birrell, Anders Gjerdrum, Robbert van Renesse, Håvard Johansen, Dag Johansen, and Fred B. Schneider. "SGX Enforcement of Use-Based Privacy." In: *Proceedings of the 17th ACM Workshop on Privacy in the Electronic Society*. ACM. 2018, pp. 155–167. DOI: 10.1145/3267323.3268954.

[32] Jeremiah Blocki, Nicolas Christin, Anupam Datta, and Arunesh Sinha. "Audit Mechanisms for Provable Risk Management and Accountable Data Governance." In: *Proceedings of the 3rd International Conference on Decision and Game Theory for Security*. Lecture Notes in Computer Science 7638. Springer, 2012, pp. 38–59. DOI: 10.1007/978-3-642-34266-0_3.

[33] Benjamin Blum. *People Analytics. Eine datenschutzrechtliche Betrachtung moderner Einsatzszenarien für automatisierte, datenbasierte Entscheidungen*. Studien zum deutschen und europäischen Arbeitsrecht 95. Nomos, 2021. DOI: 10.5771/9783748926368.

[34] Matthew T. Bodie, Miriam A. Cherry, Marcia L. McCormick, and Jintong Tang. "The Law and Policy of People Analytics." *University of Colorado Law Review* 88.4 (2017), pp. 961–1042.

[35] Andreas Boes et al. *Inverse Transparenz: Beteiligungsorientierte Ansätze für Datensouveränität in der digitalen Arbeitswelt gestalten*. URL: https://www.inversetransparenz.de/uber-das-projekt (visited on 2023-04-24).

[36] Andreas Boes, Thomas Hess, Alexander Pretschner, Tobias Kämpf, and Elisabeth Vogl. *Daten – Innovation – Privatheit – Mit Inverser Transparenz das Gestaltungsdilemma der digitalen Arbeitswelt lösen*. 2022. URL: https://www.inversetransparenz.de/neuerscheinung-forschungsreport-daten-innovation-privatheit.

[37] Céline Brassart Olsen. "To Track or Not to Track? Employees' Data Privacy in the Age of Corporate Wellness, Mobile Health, and GDPR." *International Data Privacy Law* 10.3 (2020), pp. 236–252. DOI: 10.1093/idpl/ipaa004.

[38] David Brin. *The Transparent Society*. Basic Books, 1998.

[39] Vitalik Buterin. *The Meaning of Decentralization*. 2017-02-06. URL: https://medium.com/@VitalikButerin/the-meaning-of-decentralization-a0c92b76a274 (visited on 2023-05-08).

[40] Fred H. Cate. "Principles for Protecting Privacy." *Cato Journal* 22.1 (2002), pp. 33–58.

[41] Jan Cederquist, Ricardo Corin, Marnix A. C. Dekker, Sandro Etalle, Jerry I. den Hartog, and Gabriele Lenzini. "Audit-Based Compliance Control." *International Journal of Information Security* 6.2-3 (2007), pp. 133–151. DOI: 10.1007/s10207-007-0017-y.

[42] Tomas Chamorro-Premuzic. "Can Surveillance AI Make the Workplace Safe?" *MIT Sloan Management Review* 62.1 (2020), pp. 13–15. URL: https://sloanreview.mit.edu/article/can-surveillance-ai-make-the-workplace-safe/.

[43] Sheshadri Chatterjee, Ranjan Chaudhuri, Demetris Vrontis, and Evangelia Siachou. "Examining the Dark Side of Human Resource Analytics: An Empirical Investigation Using the Privacy Calculus Approach." *International Journal of Manpower* 43.1 (2022), pp. 52–74. DOI: 10.1108/IJM-02-2021-0087.

[44] Ramnath K. Chellappa and Raymond G. Sin. "Personalization Versus Privacy: An Empirical Examination of the Online Consumer's Dilemma." *Information Technology and Management* 6 (2005), pp. 181–202. DOI: 10.1007/s10799-005-5879-y.

[45] Fei Chen, Jiahao Wang, Changkun Jiang, Tao Xiang, and Yuanyuan Yang. "Blockchain Based Non-Repudiable IoT Data Trading: Simpler, Faster, and Cheaper." In: *Proceedings of the 2022 IEEE Conference on Computer Communications*. IEEE. 2022, pp. 1958–1967. DOI: 10.1109/infocom48880.2022.9796857.

[46] Huashan Chen, Marcus Pendleton, Laurent Njilla, and Shouhuai Xu. "A Survey on Ethereum Systems Security: Vulnerabilities, Attacks, and Defenses." *ACM Computing Surveys* 53.3 (2020), pp. 1–43. DOI: 10.1145/3391195.

[47] Patrick Cichy, Torsten-Oliver Salge, and Rajiv Kohli. "Extending the Privacy Calculus: The Role of Psychological Ownership." In: *Proceedings of the 35th International Conference on Information Systems*. AIS. 2014, pp. 1–20. URL: https://aisel.aisnet.org/icis2014/proceedings/ISSecurity/30.

[48] Larry L. Constantine. "Trusted Interaction: User Control and System Responsibilities in Interaction Design for Information Systems." In: *Proceedings of the 2006 International Conference on Advanced Information Systems Engineering*. Lecture Notes in Computer Science 4001. Springer, 2006, pp. 20–30. DOI: 10.1007/11767138_3.

[49] Clotilde Coron. "Quantifying Human Resource Management: A Literature Review." *Personnel Review* 51.4 (2021), pp. 1386–1409. DOI: 10.1108/pr-05-2020-0322.

[50] John Correia and Deborah Compeau. "Information Privacy Awareness (IPA): A Review of the Use, Definition and Measurement of IPA." In: *Proceedings of the 50th Hawaii International Conference on System Sciences*. University of Hawaii at Manoa. 2017, pp. 4021–4030. DOI: 10.24251/hicss.2017.486.

[51] Cynthia L. Corritore, Robert P. Marble, Susan Wiedenbeck, Beverly Kracher, and Ashwin Chandran. "Measuring Online Trust of Websites: Credibility, Perceived Ease of Use, and Risk." In: *Proceedings of the 11th Americas Conference on Information Systems*. AIS. 2005, Article 370. URL: https://aisel.aisnet.org/amcis2005/370/.

[52] Victor Costan and Srinivas Devadas. *Intel SGX Explained*. 2016. Cryptology ePrint Archive: 2016/086.

[53] Lorrie Faith Cranor. "P3P: Making Privacy Policies More Useful." *IEEE Security & Privacy* 1.6 (2003), pp. 50–55. DOI: 10.1109/msecp.2003.1253568.

[54] Mary J. Culnan. "'How Did They Get My Name?': An Exploratory Investigation of Consumer Attitudes Toward Secondary Information Use." *MIS Quarterly* 17.3 (1993), pp. 341–363. DOI: 10.2307/249775.

[55] Giuseppe D'Acquisto, Josep Domingo-Ferrer, Panayiotis Kikiras, Vicenç Torra, Yves-Alexandre de Montjoye, and Athena Bourka. *Privacy by Design in Big Data: An Overview of Privacy Enhancing Technologies in the Era of Big Data Analytics*. Tech. rep. European Union Agency for Cybersecurity (ENISA), 2015. URL: https://www.enisa.europa.eu/publications/big-data-protection.

[56] Peter Dabrock. "From Data Protection to Data Sovereignty. A Multidimensional Governance Approach for Shaping Informational Freedom in the 'onlife'-Era." *Cursor_ Zeitschrift für explorative Theologie* (2019-11-11). DOI: 10.21428/fb61f6aa.f0bf0cc2.

[57] Mina Deng, Kim Wuyts, Riccardo Scandariato, Bart Preneel, and Wouter Joosen. "A Privacy Threat Analysis Framework: Supporting the Elicitation and Fulfillment of Privacy Requirements." *Requirements Engineering* 16.1 (2011), pp. 3–32. DOI: 10.1007/s00766-010-0115-7.

[58] Dominic Deuber, Bernardo Magri, and Sri Aravinda Krishnan Thyagarajan. "Redactable Blockchain in the Permissionless Setting." In: *Proceedings of the 40th IEEE Symposium on Security and Privacy*. IEEE. 2019, pp. 124–138. DOI: 10.1109/sp.2019.00039.

[59] André Deuker. "Addressing the Privacy Paradox by Expanded Privacy Awareness–The Example of Context-Aware Services." In: *Proceedings of the 5th International Summer School on Privacy and Identity Management for Life*. IFIP Advances in Information and Communication Technology 320. IFIP WG 9.2, 9.6/11.4, 11.6, 11.7. Springer, 2010, pp. 275–283. DOI: 10.1007/978-3-642-14282-6_23.

[60] Michael DiClaudio. "People Analytics and the Rise of HR: How Data, Analytics and Emerging Technology Can Transform Human Resources (HR) Into a Profit Center." *Strategic HR Review* 18.2 (2019), pp. 42–46. DOI: 10.1108/shr-11-2018-0096.

[61] Friso van Dijk, Marco Spruit, Chaïm van Toledo, and Matthieu J. S. Brinkhuis. "Pillars of Privacy: Identifying Core Theory in a Network Analysis of Privacy Literature." In: *Proceedings of the 29th European Conference on Information Systems*. AIS. 2021, Article 1420. URL: https://aisel.aisnet.org/ecis2021_rp/84.

[62] Tamara Dinev and Paul Hart. "An Extended Privacy Calculus Model for E-Commerce Transactions." *Information Systems Research* 17.1 (2006), pp. 61–80. DOI: 10.1287/isre.1060.0080.

[63] Stefan Dziembowski, Lisa Eckey, and Sebastian Faust. "FairSwap: How to Fairly Exchange Digital Goods." In: *Proceedings of the 25th ACM SIGSAC Conference on Computer and Communications Security*. ACM. 2018, pp. 967–984. DOI: 10.1145/3243734.3243857.

[64] Mahdi Ebnali, Kevin Hulme, Aliakbar Ebnali-Heidari, and Adel Mazloumi. "How Does Training Effect Users' Attitudes and Skills Needed for Highly Automated Driving?" *Transportation Research Part F: Traffic Psychology and Behaviour* 66 (October 2019), pp. 184–195. DOI: 10.1016/j.trf.2019.09.001.

[65] Lisa Eckey, Sebastian Faust, and Benjamin Schlosser. "OptiSwap: Fast Optimistic Fair Exchange." In: *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security*. ACM. 2020, pp. 543–557. DOI: 10.1145/3320269.3384749.

[66] Bhaskara S. Egala, Ashok K. Pradhan, Shubham Gupta, Kshira Sagar Sahoo, Muhammad Bilal, and Kyung-Sup Kwak. "CoviBlock: A Secure Blockchain-Based Smart Healthcare Assisting System." *Sustainability* 14.24, Article 16844 (2022), pp. 1–21. DOI: 10.3390/su142416844.

[67] Fabian Engl, Philipp Trubjansky, and Frank Herrmann. "Ein funktionaler Vergleich der SAP Analytics Cloud und Microsoft Power BI zur Verwendung im Bereich People Analytics bei Vitesco Technologies." *Anwendungen und Konzepte der Wirtschaftsinformatik* 15 (2022). DOI: 10.26034/lu.akwi.2022.3337.

[68] Sandro Etalle and William H. Winsborough. "A Posteriori Compliance Control." In: *Proceedings of the 12th ACM Symposium on Access Control Models and Technologies*. ACM. 2007, pp. 11–20. DOI: 10.1145/1266840.1266843.

[69] Andrea Everard and Dennis F. Galletta. "How Presentation Flaws Affect Perceived Site Quality, Trust, and Intention to Purchase From an Online Store." *Journal of Management Information Systems* 22.3 (2005), pp. 56–95. DOI: 10.2753/mis0742-1222220303.

[70] C. M. Nadeem Faisal, Martin Gonzalez-Rodriguez, Daniel Fernandez-Lanvin, and Javier de Andres-Suarez. "Web Design Attributes in Building User Trust, Satisfaction, and Loyalty for a High Uncertainty Avoidance Culture." *IEEE Transactions on Human-Machine Systems* 47.6 (2016), pp. 847–859. DOI: 10.1109/thms.2016.2620901.

[71] Salvatore V. Falletta and Wendy L. Combs. "The HR Analytics Cycle: A Seven-Step Process for Building Evidence-Based and Ethical HR Analytics Capabilities." *Journal of Work-Applied Management* 13.1 (2020), pp. 51–68. DOI: 10.1108/jwam-03-2020-0020.

[72] Samer Faraj, Stella Pachidi, and Karla Sayegh. "Working and Organizing in the Age of the Learning Algorithm." *Information and Organization* 28.1 (2018), pp. 62–70. DOI: 10.1016/j.infoandorg.2018.02.005.

[73] Florian M. Farke, David G. Balash, Maximilian Golla, Markus Dürmuth, and Adam J. Aviv. "Are Privacy Dashboards Good for End Users? Evaluating User Perceptions and Reactions to Google's My Activity." In: *Proceedings of the 30th USENIX Security Symposium*. USENIX. 2021, pp. 483–500. URL: https://www.usenix.org/conference/usenixsecurity21/presentation/farke.

[74] Simon Farshid, Andreas Reitz, and Peter Roßbach. "Design of a Forgetting Blockchain: A Possible Way to Accomplish GDPR Compatibility." In: *Proceedings of the 52nd Hawaii International Conference on System Sciences*. University of Hawaii at Manoa. 2019, pp. 7087–7095. DOI: 10.24251/hicss.2019.850.

[75] Joan Feigenbaum, James A. Hendler, Aaron D. Jaggard, Daniel J. Weitzner, and Rebecca N. Wright. "Accountability and Deterrence in Online Life." In: *Proceedings of the 3rd International Web Science Conference*. ACM. 2011, pp. 1–7. DOI: 10.1145/2527031.2527043.

[76] Alex Fleck. *Microsoft Workers' Council Partnerships Boost the Company's Product and Service Rollouts*. Microsoft, 2022-01-03. URL: https://www.microsoft.com/insidetrack/blog/microsoft-workers-council-partnerships-boost-the-companys-product-and-service-rollouts/ (visited on 2022-07-07).

[77] Brian J. Fogg and Hsiang Tseng. “The Elements of Computer Credibility.” In: *Proceedings of the 1999 CHI Conference on Human Factors in Computing Systems*. ACM. 1999, pp. 80–87. DOI: 10.1145/302979.303001.

[78] Felix Freiling and Edita Bajramovic. “Principles of Secure Logging for Safekeeping Digital Evidence.” In: *Proceedings of the 11th International Conference on IT Security Incident Management & IT Forensics*. IEEE. 2018, pp. 65–75. DOI: 10.1109/imf.2018.00012.

[79] Bronwyn French, Andreas Duenser, and Andrew Heathcote. *Trust in Automation – A Literature Review*. Tech. rep. EP184082. CSIRO, 2018.

[80] Anna-Katharina Frison, Philipp Wintersberger, Andreas Riener, Clemens Schartmüller, Linda Ng Boyle, Erika Miller, and Klemens Weigl. “In UX We Trust: Investigation of Aesthetics and Usability of Driver-Vehicle Interfaces and Their Impact on the Perception of Automated Driving.” In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM. 2019, pp. 1–13. DOI: 10.1145/3290605.3300374.

[81] Alexander Fromm and Vladislav Stepa. “HDFT++ Hybrid Data Flow Tracking for SaaS Cloud Services.” In: *Proceedings of the 4th International Conference on Cyber Security and Cloud Computing*. IEEE. 2017, pp. 333–338. DOI: 10.1109/cscloud.2017.9.

[82] Uri Gal, Tina Blegind Jensen, and Mari-Klara Stein. “Breaking the Vicious Cycle of Algorithmic Management: A Virtue Ethics Approach to People Analytics.” *Information and Organization* 30.2, Article 100301 (2020), pp. 1–15. DOI: 10.1016/j.infoandorg.2020.100301.

[83] Andrew Gambino, Jinyoung Kim, S. Shyam Sundar, Jun Ge, and Mary Beth Rosson. “User Disbelief in Privacy Paradox: Heuristics That Determine Disclosure.” In: *Extended Abstracts of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM. 2016, pp. 2837–2843. DOI: 10.1145/2851581.2892413.

[84] Benoît Garbinato and Ian Rickebusch. “Impossibility Results on Fair Exchange.” In: *Proceedings of the 10th International Conference on Innovative Internet Community Systems*. Lecture Notes in Informatics 165. Gesellschaft für Informatik, 2010. URL: https://dl.gi.de/items/eebbbc38-cd71-4334-8ed5-f0b112840854.

[85] Carrie Gates and Jacob Slonim. “Owner-Controlled Information.” In: *Proceedings of the 2003 Workshop on New Security Paradigms*. ACM. 2003, pp. 103–111. DOI: 10.1145/986655.986670.

[86] Chunpeng Ge, Siwei Sun, and Pawel Szalachowski. “Permissionless Blockchains and Secure Logging.” In: *Proceedings of the 2019 IEEE International Conference on Blockchain and Cryptocurrency*. IEEE. 2019, pp. 56–60. DOI: 10.1109/bloc.2019.8751306.

[87] General Data Protection Regulation. “Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons With Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation).”

Official Journal of the European Union 59 (2016), pp. 1–88. URL: http://data.europa.eu/eli/reg/2016/679/oj.

[88] Maren Gierlich-Joas, Thomas Hess, and Rahild Neuburger. "More Self-Organization, More Control—or Even Both? Inverse Transparency As a Digital Leadership Concept." *Business Research* 13.3 (2020), pp. 921–947. DOI: 10.1007/s40685-020-00130-0.

[89] Maren Gierlich-Joas, Mena Teebken, and Thomas Hess. "A Synthesized Perspective on Privacy and Transparency in the Digital Workplace." In: *Proceedings of the 55th Hawaii International Conference on System Sciences*. University of Hawaii at Manoa. 2022, pp. 5191–5200. DOI: 10.24251/hicss.2022.633.

[90] Lisa Marie Giermindl, Franz Strich, Oliver Christ, Ulrich Leicht-Deobald, and Abdullah Redzepi. "The Dark Sides of People Analytics: Reviewing the Perils for Organisations and Employees." *European Journal of Information Systems* 31.3 (2022), pp. 410–435. DOI: 10.1080/0960085X.2021.1927213.

[91] Harald Gjermundrød, Ioanna Dionysiou, and Kyriakos Costa. "privacyTracker: A Privacy-By-Design GDPR-Compliant Framework With Verifiable Data Traceability Controls." In: *Proceedings of the 2nd International Workshop on Technical and Legal Aspects of Data Privacy and Security*. Lecture Notes in Computer Science 9881. Springer, 2016, pp. 3–15. DOI: 10.1007/978-3-319-46963-8_1.

[92] Thomas Götz. *Big Data im Personalmanagement. Datenschutzrecht und betriebliche Mitbestimmung*. Theorie und Praxis des Arbeitsrechts 17. Nomos, 2020. DOI: 10.5771/9783748909958.

[93] Adam Grant. "The Surprising Value of Obvious Insights." *MIT Sloan Management Review* 60.3 (2019), pp. 8–10. URL: https://sloanreview.mit.edu/article/the-surprising-value-of-obvious-insights/.

[94] Elias Grünewald and Frank Pallas. "TILT: A GDPR-Aligned Transparency Information Language and Toolkit for Practical Privacy Engineering." In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. ACM. 2021, pp. 636–646. DOI: 10.1145/3442188.3445925.

[95] Elias Grünewald, Paul Wille, Frank Pallas, Maria C. Borges, and Max-Robert Ulbricht. "TIRA: An OpenAPI Extension and Toolbox for GDPR Transparency in RESTful Architectures." In: *Proceedings of the 2021 IEEE European Symposium on Security and Privacy Workshops*. IEEE. 2021, pp. 312–319. DOI: 10.1109/eurospw54576.2021.00039.

[96] Gorka Guardiola-Múzquiz and Enrique Soriano-Salvador. "SealFSv2: Combining Storage-Based and Ratcheting for Tamper-Evident Logging." *International Journal of Information Security* 22.2 (2022), pp. 1–20. DOI: 10.1007/s10207-022-00643-1.

[97] Jochen Günther. "Digital Workplace – Herausforderungen und Implikationen für die Gestaltung." *HMD Praxis der Wirtschaftsinformatik* 54.6 (2017), pp. 859–873. DOI: 10.1365/s40702-017-0364-8.

[98] Angela T. Hall, Dwight D. Frink, and M. Ronald Buckley. "An Accountability Account: A Review and Synthesis of the Theoretical and Empirical Research on Felt Accountability." *Journal of Organizational Behavior* 38.2 (2017), pp. 204–224. DOI: 10.1002/job.2052.

[99] Il-Horn Hann, Kai-Lung Hui, Sang-Yong Tom Lee, and Ivan P. L. Png. "Overcoming Online Information Privacy Concerns: An Information-Processing Theory Approach." *Journal of Management Information Systems* 24.2 (2007), pp. 13–42. DOI: 10.2753/mis0742-1222240202.

[100] Milena M. Head and Khaled Hassanein. "Trust in E-Commerce: Evaluating the Impact of Third-Party Seals." *Quarterly Journal of Electronic Commerce* 3.3 (2002), pp. 307–325.

[101] Hans Hedbom. "A Survey on Transparency Tools for Enhancing Privacy." In: *Proceedings of the 4th International Summer School on The Future of Identity in the Information Society*. IFIP Advances in Information and Communication Technology 298. IFIP WG 9.2, 9.6, 11.6, 11.7/FIDIS. Springer, 2009, pp. 67–82. DOI: 10.1007/978-3-642-03315-5_5.

[102] Alexander Hicks, Vasilios Mavroudis, Mustafa Al-Bassam, Sarah Meiklejohn, and Steven J. Murdoch. *VAMS: Verifiable Auditing of Access to Confidential Data*. 2018. arXiv: 1805.04772 [cs.CR].

[103] Mike Hintze and Khaled El Emam. "Comparing the Benefits of Pseudonymisation and Anonymisation under the GDPR." *Journal of Data Protection & Privacy* 2.2 (2018), pp. 145–158. URL: https://www.henrystewartpublications.com/sites/default/files/JDPP2.2Comparingthebenefitsofpseudonymisationandanonymisationunderthe GDPR.pdf.

[104] Kevin Anthony Hoff and Masooda Bashir. "Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust." *Human Factors* 57.3 (2015), pp. 407–434. DOI: 10.1177/0018720814547570.

[105] Jason E. Holt. "Logcrypt: Forward Security and Public Verification for Secure Audit Logs." In: *Proceedings of the 4th Australasian Information Security Workshop*. Australian Computer Society. ACM, 2006, pp. 203–211. URL: https://dl.acm.org/doi/abs/10.5555/1151828.1151852.

[106] Christian Holthaus, Young-kul Park, and Ruth Stock-Homburg. "People Analytics und Datenschutz – ein Widerspruch?" *Datenschutz und Datensicherheit – DuD* 39.10 (2015), pp. 676–681. DOI: 10.1007/s11623-015-0497-2.

[107] Laurens van Hoye, Pieter-Jan Maenhaut, Tim Wauters, Bruno Volckaert, and Filip de Turck. "Logging Mechanism for Cross-Organizational Collaborations Using Hyperledger Fabric." In: *Proceedings of the 2019 IEEE International Conference on Blockchain and Cryptocurrency*. IEEE. 2019, pp. 352–359. DOI: 10.1109/bloc.2019.8751380.

[108] Joschka A. Hüllmann, Simone Krebber, and Patrick Troglauer. “The IT Artifact in People Analytics: Reviewing Tools to Understand a Nascent Field.” In: *Proceedings of the 16th International Conference on Wirtschaftsinformatik*. Lecture Notes in Information Systems and Organisation 48. Springer, 2021, pp. 238–254. DOI: 10.1007/978-3-030-86800-0_18.

[109] Patrik Hummel, Matthias Braun, Steffen Augsberg, and Peter Dabrock. “Sovereignty and Data Sharing.” *ICT Discoveries* 1.2 (2018). URL: https://www.itu.int/en/journal/002/Pages/11.aspx.

[110] Patrik Hummel, Matthias Braun, Steffen Augsberg, Ulrich von Ulmenstein, and Peter Dabrock. *Datensouveränität: Governance-Ansätze für den Gesundheitsbereich*. Springer VS, 2021. DOI: 10.1007/978-3-658-33755-1.

[111] ISO/IEC. *Information Security — Non-repudiation*. Standard 13888-1:2020. International Organization for Standardization and International Electrotechnical Commission, 2020. URL: https://www.iso.org/standard/76153.html.

[112] Ibrahim M. Al-Jabri, Mustafa I. Eid, and Amer Abed. “The Willingness to Disclose Personal Information: Trade-Off between Privacy Concerns and Benefits.” *Information & Computer Security* 28.2 (2019), pp. 161–181. DOI: 10.1108/ics-01-2018-0012.

[113] Katie Jacobs. “The Ethics of Gathering Employee Data.” *HR Magazine* (2017-03-21). URL: https://www.hrmagazine.co.uk/content/features/the-ethics-of-gathering-employee-data.

[114] Milena Janic, Jan Pieter Wijbenga, and Thijs Veugen. “Transparency Enhancing Tools (TETs): An Overview.” In: *Proceedings of the 3rd Workshop on Socio-Technical Aspects in Security and Trust*. IEEE. 2013, pp. 18–25. DOI: 10.1109/stast.2013.11.

[115] Matthias Jarke, Boris Otto, and Sudha Ram. “Data Sovereignty and Data Space Ecosystems.” *Business & Information Systems Engineering* 61.5 (2019), pp. 549–550. DOI: 10.1007/s12599-019-00614-2.

[116] Nesrine Kaaniche and Maryline Laurent. “BDUA: Blockchain-Based Data Usage Auditing.” In: *Proceedings of the 11th International Conference on Cloud Computing*. IEEE. 2018, pp. 630–637. DOI: 10.1109/cloud.2018.00087.

[117] Severin Kacianka, Kristian Beckers, Florian Kelbert, and Prachi Kumari. “How Accountability Is Implemented and Understood in Research Tools.” In: *Proceedings of the 18th International Conference on Product-Focused Software Process Improvement*. Lecture Notes in Computer Science 10611. Springer, 2017, pp. 199–218. DOI: 10.1007/978-3-319-69926-4_15.

[118] Severin Kacianka and Alexander Pretschner. “Designing Accountable Systems.” In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. ACM. 2021, pp. 424–437. DOI: 10.1145/3442188.3445905.

[119] Friedemann Kainer and Christian Weber. “Datenschutzrechtliche Aspekte des ’Talentmanagements’.” *Betriebs-Berater* 72.46 (2017), pp. 2740–2747.

[120] Yuan J. Kang, Allan M. Schiffman, and Jeff Shrager. "RAPPD: A Language and Prototype for Recipient-Accountable Private Personal Data." In: *Proceedings of the 2014 IEEE S&P International Workshop on Data Usage Management*. IEEE. 2014, pp. 49–56. DOI: 10.1109/spw.2014.16.

[121] Jodi Kantor and Arya Sundaram. "The Rise of the Worker Productivity Score." *The New York Times* (2022-08-14). URL: https://www.nytimes.com/interactive/2022/08/14/business/worker-productivity-tracking.html.

[122] Esther Kaplan. "The Spy Who Fired Me." *Harper's Magazine* (March 2015). URL: https://harpers.org/archive/2015/03/the-spy-who-fired-me/.

[123] Vishal Karande, Erick Bauman, Zhiqiang Lin, and Latifur Khan. "SGX-Log: Securing System Logs With SGX." In: *Proceedings of the 12th ACM Asia Conference on Computer and Communications Security*. ACM. 2017, pp. 19–30. DOI: 10.1145/3052973.3053034.

[124] Florian Kelbert and Alexander Pretschner. "Data Usage Control for Distributed Systems." *ACM Transactions on Privacy and Security* 21.3, Article 12 (2018), pp. 1–32. DOI: 10.1145/3183342.

[125] Shaji A. Khan and Jintong Tang. "The Paradox of Human Resource Analytics: Being Mindful of Employees." *Journal of General Management* 42.2 (2016), pp. 57–66. DOI: 10.1109/emr.2017.8233300.

[126] René F. Kizilcec. "How Much Information? Effects of Transparency on Trust in an Algorithmic Interface." In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM. 2016, pp. 2390–2395. DOI: 10.1145/2858036.2858402.

[127] Peter H. Klopfer and Daniel I. Rubenstein. "The Concept *Privacy* and Its Biological Basis." *Journal of Social Issues* 33.3 (1977), pp. 52–65. DOI: 10.1111/j.1540-4560.1977.tb01882.x.

[128] Stefan Köpsell and Petr Švenda. "Secure Logging of Retained Data for an Anonymity Service." In: *Proceedings of the 5th International Summer School on Privacy and Identity Management for Life*. IFIP Advances in Information and Communication Technology 320. IFIP WG 9.2, 9.6/11.4, 11.6, 11.7. Springer, 2010, pp. 284–298. DOI: 10.1007/978-3-642-14282-6_24.

[129] Tim Kraska, Michael Stonebraker, Michael Brodie, Sacha Servan-Schreiber, and Daniel Weitzner. "SchengenDB: A Data Protection Database Proposal." In: *Proceedings of the 2019 VLDB Workshop on Polystore Systems for Heterogeneous Data in Multiple Databases with Privacy and Security Assurances*. Lecture Notes in Computer Science 11721. VLDB Endowment. Springer, 2019, pp. 24–38. DOI: 10.1007/978-3-030-33752-0_2.

[130] Steve Kremer, Olivier Markowitch, and Jianying Zhou. "An Intensive Survey of Fair Non-repudiation Protocols." *Computer Communications* 25.17 (2002), pp. 1606–1621. DOI: 10.1016/s0140-3664(02)00049-x.

[131] Sebastian Kube, Michel André Maréchal, and Clemens Puppe. "The Currency of Reciprocity: Gift Exchange in the Workplace." *American Economic Review* 102.4 (2012), pp. 1644–1662. DOI: 10.1257/aer.102.4.1644.

[132] Christiane Kuhn, Maximilian Noppel, Christian Wressnegger, and Thorsten Strufe. “Plausible Deniability for Anonymous Communication.” In: *Proceedings of the 20th Workshop on Workshop on Privacy in the Electronic Society*. ACM, 2021, pp. 17–32. DOI: 10.1145/3463676.3485605.

[133] Prachi Kumari. “Requirements Analysis for Privacy in Social Networks.” In: *Proceedings of the 8th International Workshop for Technical, Economic and Legal Aspects of Business Models for Virtual Goods*. 2010, pp. 28–40. URL: http://www.virtualgoods.org/2010/VirtualGoodsBook2010_38.pdf.

[134] Alptekin Küpçü and Anna Lysyanskaya. “Usable Optimistic Fair Exchange.” *Computer Networks* 56.1 (2012), pp. 50–63. DOI: 10.1016/j.comnet.2011.08.005.

[135] Marc Langheinrich. “A Privacy Awareness System for Ubiquitous Computing Environments.” In: *Proceedings of the 4th International Conference on Ubiquitous Computing*. Springer. 2002, pp. 237–245. DOI: 10.1007/3-540-45809-3_19.

[136] Robert S. Laufer and Maxine Wolfe. “Privacy As a Concept and a Social Issue: A Multidimensional Developmental Theory.” *Journal of Social Issues* 33.3 (1977), pp. 22–42. DOI: 10.1111/j.1540-4560.1977.tb01880.x.

[137] Ji-Hwan Lee and Chi-Hoon Song. “Effects of Trust and Perceived Risk on User Acceptance of a New Technology Service.” *Social Behavior and Personality: An International Journal* 41.4 (2013), pp. 587–597. DOI: 10.2224/sbp.2013.41.4.587.

[138] John D. Lee and Katrina A. See. “Trust in Automation: Designing for Appropriate Reliance.” *Human Factors* 46.1 (2004), pp. 50–80. DOI: 10.1518/hfes.46.1.50.30392.

[139] Seungho Lee, Wonsuk Choi, Hyo Jin Jo, and Dong Hoon Lee. “Poster: Secure Logging Infrastructure Employing Heterogeneous Trusted Execution Environments.” In: *Posters of the 27th Network and Distributed System Security Symposium*. 2020, pp. 1–2. URL: https://www.ndss-symposium.org/wp-content/uploads/2020/02/NDSS2020posters_paper_8.pdf.

[140] Ulrich Leicht-Deobald, Thorsten Busch, Christoph Schank, Antoinette Weibel, Simon Schafheitle, Isabelle Wildhaber, and Gabriel Kasper. “The Challenges of Algorithm-Based HR Decision-Making for Personal Integrity.” *Journal of Business Ethics* 160.2 (2019), pp. 377–392. DOI: 10.1007/978-3-031-18794-0_5.

[141] Paul Leonardi and Noshir Contractor. “Better People Analytics.” *Harvard Business Review* (November–December 2018), pp. 70–81. URL: https://hbr.org/2018/11/better-people-analytics.

[142] Yuanchun Li, Fanglin Chen, Toby Jia-Jun Li, Yao Guo, Gang Huang, Matthew Fredrikson, Yuvraj Agarwal, and Jason I. Hong. “PrivacyStreams: Enabling Transparency in Personal Data Processing for Mobile Apps.” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1.3, Article 76 (2017), pp. 1–26. DOI: 10.1145/3130941.

[143] Reyner Aranta Lika, Daksha A/P. V. Ramasamy, Danushyaa A/P. Murugiah, and Sarfraz Nawaz Brohi. "A Data Tracking and Monitoring Mechanism." In: *Proceedings of the 2019 Conference on Intelligent Computing and Innovation on Data Science*. Lecture Notes in Networks and Systems 118. Springer, 2021, pp. 773–781. DOI: 10.1007/978-981-15-3284-9_83.

[144] Konstantinos Limniotis and Marit Hansen. *Recommendations on Shaping Technology According to GDPR Provisions – An Overview on Data Pseudonymisation*. Tech. rep. European Union Agency for Cybersecurity (ENISA), 2018. URL: https://www.enisa.europa.eu/publications/recommendations-on-shaping-technology-according-to-gdpr-provisions.

[145] Jing Liu and Laurent Vigneron. "Design and Verification of a Non-repudiation Protocol Based on Receiver-Side Smart Card." *IET Information Security* 4.1 (2010), pp. 15–29. DOI: 10.1049/iet-ifs.2009.0086.

[146] Di Ma and Gene Tsudik. "A New Approach to Secure Logging." *ACM Transactions on Storage* 5.1 (2009), pp. 1–21. DOI: 10.1145/1502777.1502779.

[147] Dewi Mairiza, Didar Zowghi, and Nurie Nurmuliani. "Managing Conflicts Among Non-functional Requirements." In: *Proceedings of the 12th Australian Workshop on Requirements Engineering*. University of Technology, Sydney, 2009, pp. 11–19. URL: https://opus.lib.uts.edu.au/handle/10453/10861.

[148] Naresh K. Malhotra, Sung S. Kim, and James Agarwal. "Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model." *Information Systems Research* 15.4 (2004), pp. 336–355. DOI: 10.1287/isre.1040.0032.

[149] Ivan Manokha. "The Implications of Digital Employee Monitoring and People Analytics for Power Relations in the Workplace." *Surveillance and Society* 18.4 (2020), pp. 540–554. DOI: 10.24908/ss.v18i4.13776.

[150] Olivier Markowitch and Yves Roggeman. "Probabilistic Non-repudiation Without Trusted Third Party." In: *Proceedings of the 2nd Conference on Security in Communication Networks*. 1999, pp. 25–36.

[151] Steven McCartney and Na Fu. "Promise Versus Reality: A Systematic Review of the Ongoing Debates in People Analytics." *Journal of Organizational Effectiveness: People and Performance* 9.2 (2022), pp. 281–311. DOI: 10.1108/joepp-01-2021-0013.

[152] Tobias Mettler and Robert Winter. "Are Business Users Social? A Design Experiment Exploring Information Sharing in Enterprise Social Systems." *Journal of Information Technology* 31.2 (2016), pp. 101–114. DOI: 10.1057/jit.2015.28.

[153] Tobias Mettler and Jochen Wulf. "Physiolytics at the Workplace: Affordances and Constraints of Wearables Use From an Employee's Perspective." *Information Systems Journal* 29.1 (2019), pp. 245–273. DOI: 10.1111/isj.12205.

[154] Microsoft. *Microsoft Viva Insights*. URL: https://www.microsoft.com/en-US/microsoft-viva/insights (visited on 2023-04-05).

[155] John Mitsianis. "A New Approach to Enforcing Non-repudiation of Receipt." Manuscript inaccessible. 2001.

[156] Yves-Alexandre de Montjoye, Erez Shmueli, Samuel S. Wang, and Alex Sandy Pentland. "OpenPDS: Protecting the Privacy of Metadata Through SafeAnswers." *PLOS ONE* 9.7 (2014), pp. 1–9. DOI: 10.1371/journal.pone.0098790.

[157] Richard T. Mowday, Richard M. Steers, and Lyman W. Porter. "The Measurement of Organizational Commitment." *Journal of Vocational Behavior* 14.2 (1979), pp. 224–247. DOI: 10.1016/0001-8791(79)90072-1.

[158] Patrick Murmann and Simone Fischer-Hübner. "Tools for Achieving Usable Ex Post Transparency: A Survey." *IEEE Access* 5 (2017), pp. 22965–22991. DOI: 10.1109/access.2017.2765539.

[159] Dana Naous, Vaibhav Kulkarni, Christine Legner, and Benoit Garbinato. "Information Disclosure in Location-Based Services: An Extended Privacy Calculus Model." In: *Proceedings of the 40th International Conference on Information Systems*. AIS. 2019, Article 3118, pp. 1–17. URL: https://aisel.aisnet.org/icis2019/cyber_security_privacy_ethics_IS/cyber_security_privacy/40.

[160] Patricia A. Norberg, Daniel R. Horne, and David A. Horne. "The Privacy Paradox: Personal Information Disclosure Intentions Versus Behaviors." *Journal of Consumer Affairs* 41.1 (2007), pp. 100–126. DOI: 10.1111/j.1745-6606.2006.00070.x.

[161] Hassan N. Noura, Ola Salman, Ali Chehab, and Raphaël Couturier. "DistLog: A Distributed Logging Scheme for IoT Forensics." *Ad Hoc Networks* 98, Article 102061 (March 2020). DOI: 10.1016/j.adhoc.2019.102061.

[162] Chorng-Shyong Ong, Jung-Yu Lai, and Yi-Shun Wang. "Factors Affecting Engineers' Acceptance of Asynchronous E-Learning Systems in High-Tech Companies." *Information & Management* 41.6 (2004), pp. 795–804. DOI: 10.1016/j.im.2003.08.012.

[163] OrgVue. *Making People Count: From Workforce Analytics to Organizational Planning*. URL: https://www.orgvue.com/resources/infographic/infographic-making-people-count-2019-study-on-workforce-analytics/ (visited on 2023-04-05).

[164] Riccardo Paccagnella, Pubali Datta, Wajih Ul Hassan, Adam Bates, Christopher Fletcher, Andrew Miller, and Dave Tian. "Custos: Practical Tamper-Evident Auditing of Operating Systems Using Trusted Execution." In: *Proceedings of the 27th Network and Distributed System Security Symposium*. 2020, pp. 1–18. DOI: 10.14722/ndss.2020.24065.

[165] Ugo Pagallo, Eleonora Bassi, Marco Crepaldi, and Massimo Durante. "Chronicle of a Clash Foretold: Blockchains and the GDPR's Right to Erasure." In: *Proceedings of the 31st Annual Conference on Legal Knowledge and Information Systems*. 2018, pp. 81–90. DOI: 10.3233/978-1-61499-935-5-81.

[166] Henning Pagnia and Felix C. Gärtner. *On the Impossibility of Fair Exchange Without a Trusted Third Party*. Tech. rep. TUD-BS-1999-02. Darmstadt University of Technology, 1999.

[167] Raja Parasuraman and Victor Riley. “Humans and Automation: Use, Misuse, Disuse, Abuse.” *Human Factors* 39.2 (1997), pp. 230–253. DOI: 10.1518/001872097778543886.

[168] Jaehong Park and Ravi Sandhu. “Towards Usage Control Models: Beyond Traditional Access Control.” In: *Proceedings of the 7th ACM Symposium on Access Control Models and Technologies*. ACM. 2002, pp. 57–64. DOI: 10.1145/507711.507722.

[169] Thomas F. J.-M. Pasquier and David Eyers. “Information Flow Audit for Transparency and Compliance in the Handling of Personal Data.” In: *Proceedings of the 2nd IC2E Workshop on Legal and Technical Issues in Cloud Computing and Cloud-Supported Internet of Things*. IEEE. 2016, pp. 112–117. DOI: 10.1109/ic2ew.2016.29.

[170] Sameer Patil and Jennifer Lai. “Who Gets to Know What When: Configuring Privacy Permissions in an Awareness Application.” In: *Proceedings of the 2005 CHI Conference on Human Factors in Computing Systems*. ACM. 2005, pp. 101–110. DOI: 10.1145/1054972.1054987.

[171] Siani Pearson and Marco Casassa-Mont. “Sticky Policies: An Approach for Managing Privacy across Multiple Parties.” *Computer* 44.9 (2011), pp. 60–68. DOI: 10.1109/mc.2011.225.

[172] Gene Pease. *People Analytics – Privacy vs. Transparency*. 2018-03-14. URL: https://genepease.com/people-analytics-privacy-vs-transparency/ (visited on 2022-07-25).

[173] Erica Pedersen. “People Analytics and Individual Autonomy: Employing Predictive Algorithms As Omniscient Gatekeepers in the Digital Age Workplace.” *Columbia Business Law Review* 2020.3 (2020), pp. 1122–1164. DOI: 10.52214/cblr.v2020i3.7814.

[174] Roel Peeters and Tobias Pulls. “Insynd: Improved Privacy-Preserving Transparency Logging.” In: *Proceedings of the 21st European Symposium on Research in Computer Security*. Lecture Notes in Computer Science 9879. Springer, 2016, pp. 121–139. DOI: 10.1007/978-3-319-45741-3_7.

[175] Dirk Petersen. *Data Ethics: 6 Steps for Ethically Sound People Analytics*. Visier. URL: https://www.visier.com/clarity/six-steps-ethically-sound-people-analytics/ (visited on 2022-07-25).

[176] Andreas Pfitzmann and Marit Köhntopp. “Anonymity, Unobservability, and Pseudonymity – A Proposal for Terminology.” In: *Proceedings of the International Workshop on Design Issues in Anonymity and Unobservability*. Lecture Notes in Computer Science 2009. Springer, 2001, pp. 1–9. DOI: 10.1007/3-540-44702-4_1.

[177] Polinode. *Polinode – Powerful Network Analysis in the Cloud*. URL: https://polinode.com/ (visited on 2023-04-06).

[178] Svenja Polst and Denis Feth. “Privacy Ad Absurdum–How Workplace Privacy Dashboards Compromise Privacy.” In: *Tagungsband des 6. Mensch und Computer Workshop zu Usable Security und Privacy*. Gesellschaft für Informatik. 2020, pp. 1–7. DOI: 10.18420/muc2020-ws119-004.

[179] Svenja Polst, Patricia Kelbert, and Denis Feth. "Company Privacy Dashboards: Employee Needs and Requirements." In: *Proceedings of the 1st International Conference on HCI for Cybersecurity, Privacy and Trust*. Lecture Notes in Computer Science 11594. Springer, 2019, pp. 429–440. DOI: 10.1007/978-3-030-22351-9_29.

[180] Stefanie Pötzsch. "Privacy Awareness: A Means to Solve the Privacy Paradox?" In: *Proceedings of the 4th International Summer School on The Future of Identity in the Information Society*. IFIP Advances in Information and Communication Technology 298. IFIP WG 9.2, 9.6/11.6, 11.7/FIDIS. Springer, 2009, pp. 226–236. DOI: 10.1007/978-3-642-03315-5_17.

[181] Dean Povey. "Optimistic Security: A New Access Control Paradigm." In: *Proceedings of the 1999 Workshop on New Security Paradigms*. ACM. 1999, pp. 40–45. DOI: 10.1145/335169.335188.

[182] Giulia Pozzi, Federico Pigni, and Claudio Vitari. "Affordance Theory in the IS Discipline: A Review and Synthesis of the Literature." In: *Proceedings of the 20th Americas Conference on Information Systems*. AIS. 2014, Article 2, pp. 1–12. URL: https://aisel.aisnet.org/amcis2014/SocioTechnicalIssues/GeneralPresentations/2.

[183] Alexander Pretschner. "Achieving Accountability With Distributed Data Usage Control Technology." In: *Proceedings of the 2nd International Workshop on Accountability: Science, Technology, and Policy*. MIT. 2014, pp. 1–4. URL: http://dig.csail.mit.edu/2014/AccountableSystems2014/abs/pretschner-AccountabilityViaUsageControl.pdf.

[184] Alexander Pretschner. "An Overview of Distributed Usage Control." In: *Proceedings of the 2nd International Conference on Knowledge Engineering, Pinciples and Techniques*. 2009, pp. 25–33.

[185] Alexander Pretschner, Manuel Hilty, and David Basin. "Distributed Usage Control." *Communications of the ACM* 49.9 (2006), pp. 39–44. DOI: 10.1145/1151030.1151053.

[186] Alexander Pretschner, Manuel Hilty, Florian Schütz, Christian Schaefer, and Thomas Walter. "Usage Control Enforcement: Present and Future." *IEEE Security & Privacy* 6.4 (2008), pp. 44–53. DOI: 10.1109/msp.2008.101.

[187] Christian Priebe, Kapil Vaswani, and Manuel Costa. "EnclaveDB: A Secure Database Using SGX." In: *Proceedings of the 39th IEEE Symposium on Security and Privacy*. IEEE. 2018, pp. 264–278. DOI: 10.1109/sp.2018.00025.

[188] Pearl Pu and Li Chen. "Trust-Inspiring Explanation Interfaces for Recommender Systems." *Knowledge-Based Systems* 20.6 (2007), pp. 542–556. DOI: 10.1016/j.knosys.2007.04.004.

[189] Tobias Pulls and Roel Peeters. "Balloon: A Forward-Secure Append-Only Persistent Authenticated Data Structure." In: *Proceedings of the 20th European Symposium on Research in Computer Security*. Lecture Notes in Computer Science 9327. Springer, 2015, pp. 622–641. DOI: 10.1007/978-3-319-24177-7_31.

[190] Tobias Pulls, Roel Peeters, and Karel Wouters. “Distributed Privacy-Preserving Transparency Logging.” In: *Proceedings of the 12th ACM Workshop on Privacy in the Electronic Society*. ACM. 2013, pp. 83–94. DOI: 10.1145/2517840.2517847.

[191] Benedikt Putz, Florian Menges, and Günther Pernul. “A Secure and Auditable Logging Infrastructure Based on a Permissioned Blockchain.” *Computers & Security* 87, Article 101602 (November 2019), pp. 1–10. DOI: 10.1016/j.cose.2019.101602.

[192] Aravind Ramachandran and Murat Kantarcioglu. *Using Blockchain and Smart Contracts for Secure Data Provenance Management*. 2017. arXiv: 1709.10000 [cs.CR].

[193] Till Remmers, Utz Schäffer, and Daniel Schaupp. *Disentangling the Bright and Dark Sides of Transparency – An Integrated Analysis of Psychological Consequences*. 2020. SSRN: 3708074.

[194] Marten Risius, Annika Baumann, and Hanna Krasnova. “Developing a New Paradigm: Introducing the Intention-Behaviour Gap to the Privacy Paradox Phenomenon.” In: *Proceedings of the 28th European Conference on Information Systems*. AIS. 2020, Article 150, pp. 1–15. URL: https://aisel.aisnet.org/ecis2020_rp/150.

[195] Alexander Roßnagel, Andreas Pfitzmann, and Hans-Jürgen Garstka. *Modernisierung Des Datenschutzrechts*. Gutachten (Expertise). Bundesministerium des Inneren, 2001. URL: https://d-nb.info/1261262492/34.

[196] Richard M. Ryan and Edward L. Deci. “Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being.” *American Psychologist* 55.1 (2000), pp. 68–78. DOI: 10.1037/0003-066x.55.1.68.

[197] Javier Sánchez-Monedero and Lina Dencik. *The Datafication of the Workplace*. Working Paper. Cardiff University, 2019. URL: https://orca.cardiff.ac.uk/id/eprint/125552/.

[198] SAP. *SAP SuccessFactors Workforce Analytics: Optimize Performance and Results With Data-Driven Insights*. 2020. URL: https://www.sap.com/products/hcm/workforce-planning-hr-analytics.html?pdf-asset=12e85371-c37c-0010-82c7-eda71af511fa (visited on 2022-12-28).

[199] Sapience. *Workforce Efficiency*. URL: https://sapienceanalytics.com/workforce-efficiency/ (visited on 2023-04-06).

[200] Rathindra Sarathy and Han Li. “Understanding Online Information Disclosure As a Privacy Calculus Adjusted by Exchange Fairness.” In: *Proceedings of the 28th International Conference on Information Systems*. AIS. 2007, Article 21, pp. 1–14. URL: https://aisel.aisnet.org/icis2007/21.

[201] Chian Chyi Saw and Anushia Inthiran. “Designing for Trust on E-Commerce Websites Using Two of the Big Five Personality Traits.” *Journal of Theoretical and Applied Electronic Commerce Research* 17.2 (2022), pp. 375–393. DOI: 10.3390/jtaer17020020.

[202] Christian Schaefer and Christine Edman. “Transparent Logging With Hyperledger Fabric.” In: *Proceedings of the 2019 IEEE International Conference on Blockchain and Cryptocurrency*. IEEE. 2019, pp. 65–69. DOI: 10.1109/bloc.2019.8751339.

[203] Stefan Schorradt, Edita Bajramovic, and Felix Freiling. “On the Feasibility of Secure Logging for Industrial Control Systems Using Blockchain.” In: *Proceedings of the 3rd Central European Cybersecurity Conference*. ACM. 2019, pp. 1–6. DOI: 10.1145/3360664.3360668.

[204] John S. Seberger, Marissel Llavore, Nicholas Nye Wyant, Irina Shklovski, and Sameer Patil. “Empowering Resignation: There’s an App for That.” In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM. 2021, Article 552, pp. 1–18. DOI: 10.1145/3411764.3445293.

[205] Oshani Seneviratne and Lalana Kagal. “Enabling Privacy Through Transparency.” In: *Proceedings of the 12th International Conference on Privacy, Security and Trust*. IEEE. 2014, pp. 121–128. DOI: 10.1109/pst.2014.6890931.

[206] Asaf Shabtai, Maya Bercovitch, Lior Rokach, and Yuval Elovici. “Optimizing Data Misuse Detection.” *ACM Transactions on Knowledge Discovery from Data* 8.3, Article 16 (2014), pp. 1–23. DOI: 10.1145/2611520.

[207] Louis Shekhtman and Erez Waisbard. “EngraveChain: A Blockchain-Based Tamper-Proof Distributed Log System.” *Future Internet* 13.6, Article 143 (2021), pp. 1–16. DOI: 10.3390/fi13060143.

[208] H. Jeff Smith, Tamara Dinev, and Heng Xu. “Information Privacy Research: An Interdisciplinary Review.” *MIS Quarterly* 35.4 (2011), pp. 989–1015. DOI: 10.2307/41409970.

[209] H. Jeff Smith, Sandra J. Milberg, and Sandra J. Burke. “Information Privacy: Measuring Individuals’ Concerns about Organizational Practices.” *MIS Quarterly* 20.2 (1996), pp. 167–196. DOI: 10.2307/249477.

[210] Matthias Söllner, Axel Hoffmann, Holger Hoffmann, and Jan Marco Leimeister. “Towards a Theory of Explanation and Prediction for the Formation of Trust in IT Artifacts.” In: *Proceedings of the 10th Annual Workshop on HCI Research in MIS*. AIS. 2011, Article 6, pp. 1–5. URL: https://aisel.aisnet.org/sighci2011/6.

[211] Matthias Söllner, Axel Hoffmann, Holger Hoffmann, Arno Wacker, and Jan Marco Leimeister. “Understanding the Formation of Trust in IT Artifacts.” In: *Proceedings of the 33rd International Conference on Information Systems*. AIS. 2012, Article 11, pp. 1–18. URL: https://aisel.aisnet.org/icis2012/proceedings/HumanBehavior/11.

[212] Ruth Stock-Homburg and Martin Hannig. “Is There a Privacy Paradox in the Workplace?” In: *Proceedings of the 41st International Conference on Information Systems*. AIS. 2020, Article 1586, pp. 1–17. URL: https://aisel.aisnet.org/icis2020/cyber_security_privacy/cyber_security_privacy/4.

[213] Eugene F. Stone and Dianna L. Stone. “Privacy in Organizations: Theoretical Issues, Research Findings, and Protection Mechanisms.” *Research in Personnel and Human Resources Management* 8.3 (1990), pp. 349–411.

[214] Chun Hui Suen, Ryan K. L. Ko, Yu Shyang Tan, Peter Jagadpramana, and Bu Sung Lee. "S2Logger: End-To-End Data Tracking Mechanism for Cloud Data Provenance." In: *Proceedings of the 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*. IEEE. 2013, pp. 594–602. DOI: 10.1109/trustcom.2013.73.

[215] Smitha Sundareswaran, Anna Squicciarini, and Dan Lin. "Ensuring Distributed Accountability for Data Sharing in the Cloud." *IEEE Transactions on Dependable and Secure Computing* 9.4 (2012), pp. 556–568. DOI: 10.1109/tdsc.2012.26.

[216] Alistair Sutcliffe. "Trust: From Cognition to Conceptual Models and Design." In: *Proceedings of the 18th International Conference on Advanced Information Systems Engineering*. Springer. 2006, pp. 3–17. DOI: 10.1007/11767138_1.

[217] Mena Angela Teebken and Thomas Hess. "Privacy in a Digitized Workplace: Towards an Understanding of Employee Privacy Concerns." In: *Proceedings of the 54th Hawaii International Conference on System Sciences*. University of Hawaii at Manoa. 2021, pp. 6661–6670. DOI: 10.24251/hicss.2021.800.

[218] Teramind. *Insider Threat Prevention & Detection Powered by Behavior Analytics*. URL: https://www.teramind.co/solutions/insider-threat-detection (visited on 2023-04-06).

[219] Danan Thilakanathan, Shiping Chen, Surya Nepal, and Rafael Calvo. "SafeProtect: Controlled Data Sharing With User-Defined Policies in Cloud-Based Collaborative Environment." *IEEE Transactions on Emerging Topics in Computing* 4.2 (2015), pp. 301–315. DOI: 10.1109/tetc.2015.2502429.

[220] Jan Tolsdorf, Delphine Reinhardt, and Luigi Lo Iacono. "Employees' Privacy Perceptions: Exploring the Dimensionality and Antecedents of Personal Data Sensitivity and Willingness to Disclose." *Proceedings on Privacy Enhancing Technologies* 2022.2 (2022), pp. 68–94. DOI: 10.2478/popets-2022-0036.

[221] Alin Tomescu, Vivek Bhupatiraju, Dimitrios Papadopoulos, Charalampos Papamanthou, Nikos Triandopoulos, and Srinivas Devadas. "Transparency Logs Via Append-Only Authenticated Dictionaries." In: *Proceedings of the 26th ACM SIGSAC Conference on Computer and Communications Security*. ACM. 2019, pp. 1299–1316. DOI: 10.1145/3319535.3345652.

[222] Nguyen Binh Truong, Kai Sun, Gyu Myoung Lee, and Yike Guo. "GDPR-Compliant Personal Data Management: A Blockchain-Based Solution." *IEEE Transactions on Information Forensics and Security* 15 (2019), pp. 1746–1761. DOI: 10.1109/tifs.2019.2948287.

[223] Aizhan Tursunbayeva, Stefano Di Lauro, and Claudia Pagliari. "People Analytics—A Scoping Review of Conceptual Boundaries and Value Propositions." *International Journal of Information Management* 43 (December 2018), pp. 224–247. DOI: 10.1016/j.ijinfomgt.2018.08.002.

[224] Aizhan Tursunbayeva, Claudia Pagliari, Stefano Di Lauro, and Gilda Antonelli. "The Ethics of People Analytics: Risks, Opportunities and Recommendations." *Personnel Review* 51.3 (2021), pp. 900–921. DOI: 10.1108/pr-12-2019-0680.

[225] Visier. *Talent Retention Analytics*. URL: https://www.visier.com/products/talent-retention/ (visited on 2023-04-06).

[226] Paul Georg Wagner, Pascal Birnstill, and Jürgen Beyerer. "Distributed Usage Control Enforcement Through Trusted Platform Modules and SGX Enclaves." In: *Proceedings of the 23rd ACM on Symposium on Access Control Models and Technologies*. ACM. 2018, pp. 85–91. DOI: 10.1145/3205977.3205990.

[227] Guilin Wang. "Generic Non-repudiation Protocols Supporting Transparent Off-Line TTP." *Journal of Computer Security* 14.5 (2006), pp. 441–467. DOI: 10.3233/jcs-2006-14504.

[228] Liang Wang, Jiayan Liu, and Wenyuan Liu. "Staged Data Delivery Protocol: A Blockchain-Based Two-Stage Protocol for Non-repudiation Data Delivery." *Concurrency and Computation: Practice and Experience* 33.13, Article e6240 (2021). DOI: 10.1002/cpe.6240.

[229] Alexander Weinhard, Matthias Hauser, and Frédéric Thiesse. "Explaining Adoption of Pervasive Retail Systems With a Model Based on UTAUT2 and the Extended Privacy Calculus." In: *Proceedings of the 2017 Pacific Asia Conference on Information Systems*. AIS. 2017, Article 217, pp. 1–12. URL: https://aisel.aisnet.org/pacis2017/217.

[230] Daniel J. Weitzner, Harold Abelson, Tim Berners-Lee, Joan Feigenbaum, James Hendler, and Gerald Jay Sussman. *Information Accountability*. Tech. rep. MIT-CSAIL-TR-2007-034. Massachussetts Institute of Technology, 2007. DOI: 10.1145/1349026.1349043.

[231] Daniel J. Weitzner, Harold Abelson, Tim Berners-Lee, Joan Feigenbaum, James Hendler, and Gerald Jay Sussman. "Information Accountability." *Communications of the ACM* 51.6 (2008), pp. 82–87. DOI: 10.1145/1349026.1349043.

[232] Daniel J. Weitzner, Harold Abelson, Tim Berners-Lee, Chris Hanson, James Hendler, Lalana Kagal, Deborah L. McGuinness, Gerald Jay Sussman, and K. Krasnow Waterman. *Transparent Accountable Data Mining: New Strategies for Privacy Protection*. Tech. rep. MIT-CSAIL-TR-2006-007. Massachussetts Institute of Technology, 2006. URL: https://dspace.mit.edu/handle/1721.1/30972.

[233] Patrick Westphal, Javier David Fernandez Garcia, Sabrina Kirrane, and Jens Lehmann. "SPIRIT: A Semantic Transparency and Compliance Stack." In: *Proceedings of the 14th International Conference on Semantic Systems*. CEUR Workshop Proceedings 2198. 2018, Article 119, pp. 1–4. URL: http://ceur-ws.org/Vol-2198/paper_119.pdf.

[234] Attila Altay Yavuz and Peng Ning. "BAF: An Efficient Publicly Verifiable Secure Audit Logging Scheme for Distributed Systems." In: *Proceedings of the 25th Annual Computer Security Applications Conference*. IEEE. 2009, pp. 219–228. DOI: 10.1109/acsac.2009.28.

[235] Tal Z. Zarsky. “The Trouble With Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making.” *Science, Technology, & Human Values* 41.1 (2016), pp. 118–132. DOI: 10.1177/0162243915605575.

[236] Angeliki Zavou, Vasilis Pappas, Vasileios P. Kemerlis, Michalis Polychronakis, Georgios Portokalidis, and Angelos D. Keromytis. “Cloudopsy: An Autopsy of Data Flows in the Cloud.” In: *Proceedings of the 1st International Conference on Human Aspects of Information Security, Privacy, and Trust*. Lecture Notes in Computer Science 8030. Springer, 2013, pp. 366–375. DOI: 10.1007/978-3-642-39345-7_39.

[237] Liang Zhang, Haibin Kan, Yang Xu, and Jinhao Ran. “Revocable Data Sharing Methodology Based on SGX and Blockchain.” In: *Proceedings of the 15th International Conference on Network and System Security*. Lecture Notes in Computer Science 13041. Springer, 2021, pp. 61–78. DOI: 10.1007/978-3-030-92708-0_4.

[238] Qingqin Zhang. *Workplace Surveillance and Protection of Worker's Privacy in Covid-19*. 2021. URL: https://www.qmul.ac.uk/law/media/law/docs/events/Qingqin-Zhang.pdf.

[239] Jianying Zhou and Dieter Gollmann. “Observations on Non-repudiation.” In: *Proceedings of the 3rd International Conference on the Theory and Application of Cryptology and Information Security*. Lecture Notes in Computer Science 1163. Springer, 1996, pp. 133–144. DOI: 10.1007/bfb0034842.

[240] Guy Zyskind, Oz Nathan, and Alex ‘Sandy’ Pentland. “Decentralizing Privacy: Using Blockchain to Protect Personal Data.” In: *Proceedings of the 2015 IEEE S&P International Workshop on Privacy Engineering*. IEEE. 2015, pp. 180–184. DOI: 10.1109/spw.2015.27.