A problem-based approach to the advancement of heuristics for socio-technical evaluation

Thomas Herrmann, Isa Jahnke & Alexander Nolte

BEHAVIOUR & INFORMATION TECHNOLOGY
https://doi.org/10.1080/0144929X.2021.1972157

A problem-based approach to the advancement of heuristics for socio-technical evaluation

Thomas Herrmann [a], Isa Jahnke [b] and Alexander Nolte[c,d]

[a]Ruhr-University of Bochum, Department of Information and Technology Management (IMTM), Bochum, Germany; [b]University of Missouri, School of Information Science and Learning Technologies, Columbia, MO, USA; [c]University of Tartu, Institute of Computer Science, Tartu, Estonia; [d]Carnegie Mellon University, Institute for Software Research, Pittsburgh, PA, USA

ABSTRACT

With ubiquitous, mobile computing, health care systems, and smart factories, socio-technical phenomena continue to emerge that challenge traditional design and evaluation methods. We perceive such phenomena as the intertwinement of technical artifacts and social practices. Previous work shows that there is no sufficient method to evaluate the quality of this *socio-technical intertwinement*. Hence, our goal was to develop socio-technical heuristics, in short ST-heuristics, that can be applied by individuals to detect issues. Drawing inspiration from the success of usability heuristics in the field of human–computer interaction, we first applied a literature review to develop an initial set of ST-heuristics derived from six domains comprising groupware/computer-support cooperative work, job design, usability, socio-technical design principles, privacy, and process design. We then conducted two studies to evaluate and improve this set using empirical data from 13 cases from health care, industry, and engineering education fields. In total, we analysed 306 problems. The results substantiate a final set of eight ST-heuristics which allow for evaluating the socio-technical intertwinement in situ. We perceive the contribution of this work as a starting point for evaluators to uncover crucial issues and to improve current practice. We discuss the developed set of ST-heuristics within existing literature.

ARTICLE HISTORY
Received 17 January 2021
Accepted 19 August 2021

KEYWORDS
Heuristics; socio-technical systems; evaluation; system design; work systems; social practice

1. Introduction

With the increasing emergence and adoption of applications, such as electronic health care, e-commerce, social networks, ubiquitous computing, and smart factories, the socio-technical perspective has gained more and more relevance. We characterise these emerging phenomena with an increasing complexity and contingency of intertwined social practices and technical artifacts. They cannot be fully analysed or modelled; once you start describing them, and before you come to an end, the intertwinement has already evolved and changed. This complexity arises from the different nature of multiple elements that interact with each other (Shin 2014). The dynamics of socio-technical intertwinement undergo a continuous evolution in a complex setting that is characterised by dynamic changes, uncertainty, and ambiguity (Herrmann et al. 2017) (see Section 2 for related work).

Independently of how scholars may conceptualise the complexity and contingency of these phenomena, e.g. as socio-technical intertwinement or organisational systems, the challenge is to run a type of formative *evaluation* that can inform ongoing cycles of design. Once you design a system, how do you know the quality is sufficient? You need evaluation methods to judge that. In general, evaluation is the act of making a judgment about a system or product while it is being developed or after it is deployed and in-use and has the goal of detecting potential issues (Reeves and Hedberg 2003). Evaluation methods make it possible to identify severe problems that can then be addressed by continuous and iterative cycles of agile improvement. This, however, leaves designers and managers with the challenge of not only designing but also evaluating the appropriateness or quality of the socio-technical intertwinement after such systems have been put into practice. For example, designers and managers might want to know whether a system works in a given context or how it can be improved in terms of its effectiveness, efficiency, or appeal.

Methods of how to evaluate a socio-technical intertwinement's quality are still under-researched. Existing evaluation approaches focus on how interactive systems match human tasks while other approaches address the design and distribution of tasks (Kleiner 2006; Hertzum 2010; Imanghaliyeva et al. 2020). Other literature focuses on specific domains or sectors, and goes into the depth of specifics, but does not provide a general set for socio-technical evaluations or framework that

CONTACT Thomas Herrmann thomas.herrmann@rub.de Ruhr-University of Bochum, Department of Information and Technology Management (IMTM), Universitätsstr. 150, 47058 Bochum, Germany

can be used for all kinds of socio-technical intertwinements (e.g. Ferreira et al. 2019; Nolte et al. 2018; Moallemi et al. 2017; Lowe, Chiu, and Oreszczyn 2018). In addition, different cases published by Demir et al. (2019) or Ackermann et al. (2017) revealed a set of problems that would not have been discovered using the aforementioned evaluation methods. Some of these problems include, for example, contingency issues with the communicative assistance of help desks, flexibility regarding the assigning and scheduling of tasks, involvement of informal roles and the negotiation of their rights and duties, and support of continuous learning.

We draw our inspiration on how to evaluate socio-technical intertwinements from the method of *heuristics* (Herrmann et al. 2016). Heuristic-based usability inspection has been proven successful and efficient (Wilson 2013). Heuristics are *rules of thumb* to detect issues in technology usage and are applied along workflows of task handling, help to uncover serious shortcomings, and uncover usability and user experience problems. Typical examples of heuristics are published in Nielsen and Molich (1990) or Shneiderman's golden rules (Shneiderman and Plaisant 2005). Here we aimed to develop a set of heuristics for evaluating the socio-technical intertwinement – not just the evaluation of its technology usage but also considering the social or organisational practices. We call them socio-technical heuristics or ST-heuristics. The following **research questions** (RQ) guided our work:

(1) What does a manageable set of heuristics that help to evaluate the *intertwinement* between social practices and technical artifacts look like (RQ1)?
(2) How sufficiently specific are these heuristics with respect to their scope, conciseness, and understandability (RQ2)?
(3) How does the new set of heuristics for evaluating the socio-technical intertwinement – in short *ST-heuristics*– go beyond existing heuristics (RQ3)?

To answer them, we followed a formative research approach as proposed by Quiñones, Rusu, and Rusu (2018) in the context of usability or user experience heuristics. First, we conducted a literature review to analyse existing design principles from six domains. We then conducted two empirical studies to refine the design of the ST-heuristics and explored their practicability based on multiple real-world cases. The final set of eight ST-heuristics will be discussed.

2. Background and related work

In this section, we first introduce the various approaches of *socio-technical intertwinement* and related work (2.1). We then describe evaluation concepts and heuristics approaches as a promising option for evaluating socio-technical intertwinement (2.2).

2.1. Related work and theoretical concepts

In early works, the concept of socio-technical intertwinement was described with socio-technical systems (STS) and initially based on the observation that the development of organisations cannot be understood without referencing the technology used (Trist and Murray 1993 referring to Eric Trist 1950). Later, literature emphasised that organisations should not simply be regarded as technical systems into which replaceable individuals are fitted (Emery, Thorsrud, and Thorsrud 2013 original from 1969) but should 'give equal weight to social and technical issues' (Mumford 2000, 125).

Since then, the discourse of socio-technical systems has gained new momentum during the last decade. For example Ghaffarian (2011) states that the STS-perspective has lost its relevance as design guidance but is still a highly relevant analytical approach. Baxter and Sommerville (2011) argue to strengthen the STS view in software engineering by integrating different disciplines such as computer-supported cooperative work (CSCW), human-computer interaction (HCI), or business process reengineering. They advocate for overcoming the separation of software and usability design on the one hand and the management of change on the other hand; and emphasise the relevance of considering the intertwinement of both aspects.

Recently, Shin and colleagues have conducted a series of studies that employ a socio-technical framework that emphasises the relevance of the combination of social and technical components. They argue that the complexity of a socio-technical system is related to the distinct nature of the elements that interact with each other (Shin and Ibahrine 2020; Shin 2020; Shin and Jung 2012; Shin, Choo, and Beom 2011).

The socio-technical view has practical as well as theoretical implications. On the practical side, it is applied to improve the design of technical systems in a wide span of areas such as ubiquitous computing (Shin 2019), health care (Ackermann et al. 2017), maintenance repair (Vinck 2019), and others. The early approaches to support and evaluate socio-technical design (Andriessen 2003) have been continued. For instance, MEESTAR has been developed to evaluate socio-technical arrangement in the context of nursing (Wutzkowsky and Böckmann 2018), and Imanghaliyeva et al. (2020) propose a synthesis of the socio-technical principles that has been developed over the years. Shin and colleagues argue that fairness, transparency, and accountability are the most relevant issues to be satisfied in the context of socio-technical systems (Shin and Ibahrine 2020; Shin, Choo, and Beom 2011).

On the theoretical side, the socio-technical view is also present in approaches such as the concept of socio-technical resilience (Amir and Kant 2018) or the discourse on sociomateriality (Leonardi 2012). Related to the latter, Leonardi differentiates between the social subsystem (including roles, hierarchies, communication networks, and others) and the technical subsystem that he characterises as imbrication of human (social) agency and material agency. We suggest that intertwinement should refer to the interplay between the social and technical sub-system as well as to the imbrication of social and material agency. Scholars referring to Orlikowski (2007) use the term *entanglement* for the intertwinement of the social and the material. We however perceive the concept of entanglement to not necessarily constitute a symmetry between the social and the material though, as proposed in the Actor-Network-Theory (Booth et al. 2016; Law 2009). Communities such as CSCW and HCI perceive social practices and technical artifacts as inevitably *intertwined*. Once the technology is launched and integrated in a certain context (e.g. organisation, communities, social practice), the social practices and the technical artifacts merge into a form of a system in which the two parts cannot be separated anymore; the one affects the other and vice versa. This intertwinement varies from one system to another– not all systems have the same intertwinement of social practices and technical artifacts. We suggest that these differences in the intertwinement of social practices and technical artifacts should be a subject of inspection and potential improvement. From our theoretical point of view, the intertwinement combines well-structured with less structured or formal with informal phenomena. The variety of these combinations leads to increased contingency of socio-technical systems in the sense of Luhmann (1996), which is characterized by variability, particularity, mutability and uncertainty, and serves as a basis for continuous evolution (Fischer and Herrmann 2015).

Figure 1 illustrates how we perceive the socio-technical intertwinement. Human–computer interaction plays a central role in this intertwinement of social and organisational practices with technical artifacts. From this perspective, socio-technical intertwinement can be described as intertwined organisational practices of task-handling in an ecology of tasks (Carroll and Campbell 1989; Kirsch, Troxler, and Ulich 1995) and the usage of technical artifacts or infrastructure. Such organisational practices are part of social practices that inevitably require human communication (Luhmann and Knodt 1996) between various roles (Kling 2007) and include both formal and informal tasks (Herrmann et al. 2017).

This view shifts the focus away from the user and the technical artifacts to a more sociological and systemic understanding of social practices. The close intertwinement between the components shown in Figure 1 and the continuous evolution of this intertwinement can be a subject of evaluation.

In Appendix 1, we describe the example of predictive maintenance (PM), its socio-technical intertwinement, and how heuristics can be useful to detect potential issues.

2.2. Evaluating socio-technical intertwinements and usefulness of heuristics

Related work points to challenges in evaluation approaches. Bruce et al. differentiate between situated evaluation and standard (summative and formative) approaches. '*Situated evaluation* is an approach to articulating the emergence of innovations through practice, assuming that innovations are mutually constituted by social practice and some external input' (Bruce, Rubin, and An 2010, 687). They focus on social practices that emerge with a technical innovation and the different uses that can evolve with them. Bruce et al. are aligned with Eason (1989) as they point out that the design of a good or valuable socio-technical system is never finished; that draws attention to formative evaluation methods. As Reeves and Hedberg (2003) write, formative evaluation methods are designed to collect data while a product is being developed with the intention of improving it. The results of a formative evaluation provide feedback on how the different components of a program or system are working and lead to decisions regarding what needs to be enhanced, what needs to be deleted or added, and what needs to be revised.

Several authors propose means or methods for socio-technical evaluation. For example, Shin (2019), Krenn (2015) and Nelles et al. (2017) propose a method to evaluate Internet of the things or human-robot collaboration (Behrenbruch et al. 2014). Andriessen (2003) develops a set of seven evaluation heuristics (e.g. technical efficacy, context match, and adaption) which are derived from theoretical considerations such as activity theory, communication theories, structuration theory, and technology acceptance model and refer to the four domains of HCI, communication, group work, and organisational change.

These approaches, however, do not consistently consider that the social practices, in which the use of technical artifacts is embedded, should also be subject to evaluation and redesign. The evaluation of technical

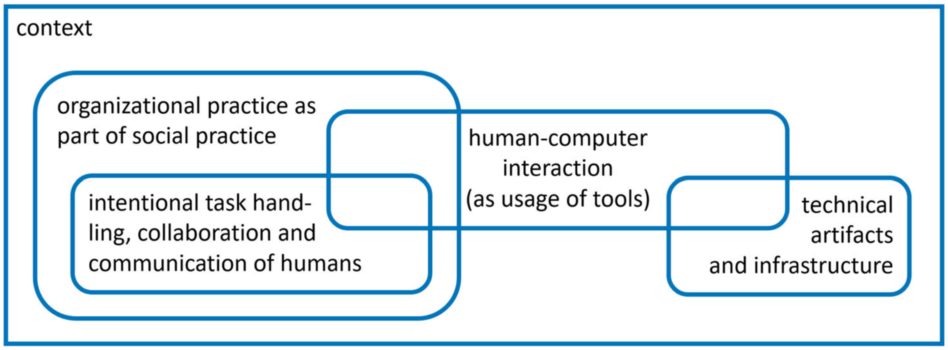

Figure 1. Socio-technical intertwinement of social practice and technical artifacts through human-computer interaction.

artifacts mainly revolves around their usage for work tasks. Other methods range from cognitive walk-throughs, think-aloud, and focus groups, collecting data about user satisfaction and task-based performance methods measuring the user's task completion rate and time with the goal to evaluate and improve ease of use, error frequency and severity, and efficiency of technical artifacts (Schmidt et al. 2020). In summary, these methods cover some aspects of the socio-technical inter-twinement, but neglect to evaluate its organisational context, social practices, or social dynamics (Bødker 2015).

Few methods address or evaluate the fit of technical artifacts into an organisation, mainly in the context of labour. Examples are participatory design approaches such as MUST (Kensing, Simonsen, and Bodker 1998) and ETHICS (Mumford 1983). They focus on the phase of designing socio-technical systems rather than evaluating. A method that combines socio-technical design and evaluation is the Socio-Technical Walk-Through (STWT) (Herrmann et al. 2004) that consists of a series of facilitated workshops. The facilitator employs a set of guiding questions to inspire a group of stakeholders to discuss the features of an STS that is represented by process diagrams. What is missing are hints or heuristics that help develop these guiding questions.

Overall, the related literature indicates a gap of evaluating socio-technical intertwinements beyond the exclusive evaluation of technical artifacts and their usage and neglects the evaluation of the social practice and its context. Hence, we focus on an integrated approach that guides the evaluation of the socio-technical intertwinement instead of principles that aim at supporting design (Imanghaliyeva et al. 2020).

2.2.1. Heuristic-based inspection and evaluation

We apply the term, *heuristic*, as it has been influenced by cognitive psychology. 'Heuristics are rules of thumb for reasoning, a simplification, or educated guess that reduces or limits the search for solutions in domains that are difficult and poorly understood' (Interaction Design Foundation 2021, 1).

Heuristics have been proved useful to evaluate systems, mainly to quickly detect the most severe usability issues in human–computer interaction (Wilson 2013). They help to efficiently identify the most serious problems and help to draft design recommendations. Experts or users inspect the features of an interactive system step-by-step by applying a list of items (heuristics). Typical heuristics that have been used are those of Nielsen (1994b) or the International Standard Organization (ISO 9241–110, 2006). They were developed to allow evaluators to assess an interactive system quickly without running extensive user tests. From the perspective of advanced HCI research and socio-technical design, it is reasonable to extend the evaluation of the usability of technical artifacts by including the social practice and broader organisational context, for example the quality work as outlined by Muller and McClard (1995) and Cockton et al. (2008). Muller et al. point out that the focus on usability has its shortcomings and that an extension, which, for example, takes the support of quality work and the users' skills into account, can be useful. Consequently, we here propose an approach that widens the existing heuristics' focus from a pure technology view and its usability to a broader socio-technical view by also evaluating the context of organisational conditions and social practices.

2.2.2. Existing sets of heuristics in STS-related domains

In order to prepare our literature search and the integrated set of new heuristics (see Section 4.1), we identified various relevant domains or disciplines that already have published sets of categories, criteria, principles, or guidelines. We then used these existing sets and synthesised them into a new set of socio-technical heuristics. (See Section 3 for details.) The six domains that span different disciplines were

- human–computer interaction (HCI) and usability,
- computer-supported cooperative work (CSCW)/groupware,
- process (re)design,
- socio-technical design,
- principles of job design, and
- privacy.

The relevance of these six domains is mirrored in Baxter and Summerville's (2011) work on socio-technical systems engineering. In addition, the domains were cross-checked with other literature that lists design aspects of how individuals collaborate using technology in an organisational or societal context. We aimed to focus on the most influential research work in the domains. In the literature review, we first searched for already existent principles (Section 2.4.1–2.4.6) that we then used later in our empirical study.

It is important to note that prior work does not always utilise the term *heuristics*. Instead, one will find terms such as principles, categories, design guidelines, or golden rules. They all refer to the same idea of heuristics in that they provide strategies to make decisions on how to improve the system where rational choices are possible (Interaction Design Foundation 2021). In this study, we use them as synonyms for *heuristics*. The heuristics, that we refer to in the following sub-sections, might serve a variety of purposes while we mainly consider them with respect to their potential contribution to socio-technical evaluation.

2.2.2.1. Heuristics found in literature of human–computer interaction (HCI). In HCI literature, heuristics mainly cover design or evaluation principles and go back to the origins of heuristics-based evaluation that can be seen in the work of Nielsen and Molich (1990), who developed a set of 10 heuristics for usability assessment (Nielsen 1994b). Comparable to these heuristics are the principles of dialogue design, which were later extended and converted into an ISO standard (ISO 9241–110:2006). In addition, there are other prominent collections of principles or guidelines in the field of usability research that can be used to evaluate interactive systems, such as Shneiderman's Golden Rules (Shneiderman and Plaisant 2005), and the principles of Dix et al. (2003) and Tognazzini (2014). Since HCI represents the link between human activities, social interaction, and technical artifacts (Figure 1), it is relevant for socio-technical heuristics.

2.2.2.2. Heuristics found in the field of computer-supported cooperative work (CSCW). The field of CSCW published heuristics mainly utilising the related term 'principles' that cover the social dimension and social practices. Criteria for the evaluation of groupware solutions are already available (Baker, Greenberg, and Gutwin 2001: Greenberg et al. 2000; Herrmann, Wulf, and Hartmann 1996). The criteria are concerned with the technical support of collaborative and cooperative work processes. CSCW, for example, introduces the concept of *awareness* (Dourish and Bellotti 1992) that extends the HCI principle of *visibility*[1] (Nielsen 1994a) to the question of what others have done and are doing and with whom one collaborates.

2.2.2.3. Heuristics found in the field of process (re)design. The field of process design shows heuristics as published principles for developing workflow systems. It is noticeable that the strictest form of coordination of collaborative processes happens through workflow management systems. It is, therefore, relevant to search for criteria or principles about the interaction of technology and organisation in the workflow management research context, whereby the perspective of process redesign is most relevant. Reijers and Mansar (2005) outline principles that deal with trade-offs between quality, cost, time, and flexibility of business processes. These principles are relevant for developing STS heuristics. For example, they focus on the customer perspective (e.g. to reduce the number of contacts with customers and third parties) and on factors external to an organisation (e.g. outsourcing a business process in whole or parts).

2.2.2.4. Heuristics found in the field of socio-technical design. There are several socio-technical design principles, that we consider as potential heuristics, such as the ones by Cherns (1987), Clegg (2000), and Eason (1989), that inform designers while developing an STS. These principles mainly refer to the process of design and development of an STS and less to the evaluation of STS. For example, Clegg asserts formulations, such as 'design should reflect the needs of (…) users' and 'core processes should be integrated.' These principles are potential candidates that we considered in our study for developing ST-heuristics.

2.2.2.5. Heuristics found in the field of job design and work. Enid Mumford (1983) combines information technology (IT) development and application with socio-technical principles that we consider as potential heuristics. She introduces five types of fits between the needs of working people or groups and the technical and organisational measures: knowledge fit, psychological fit, efficiency fit, task structure fit, and ethical fit. Mumford's approach clearly shows links to job redesign criteria by referring to the relevance of job satisfaction. The socio-technical perspective looks at the conditions of human work for individuals and groups that are shaped by the interplay between technology, organisation, and human characteristics in task processing. In this vein, we looked for further literature in the domain of job redesign, in which the work of Hackman and Oldham (Hackman and Oldham 1975) is particularly influential. A direct reference to IT-based information processing and job design is made by Dunckel's criteria (Dunckel 1989). Grote et al. (2000) present an elaborated method called KOMPASS for evaluating the interplay between humans and technology.

2.2.2.6. Heuristics found in the field of privacy. Relevant heuristics or principles can also be found in literature on privacy. The aspects of data protection and privacy are of great importance for work design in Europe (and Germany in particular) and in the context of social network systems. However, they are hardly mentioned in the literature on socio-technical principles mentioned above. Only in the early CSCW discussion can corresponding analyses be found (Clement 1994). It is appropriate to include data protection criteria for potential inclusion in ST-heuristics (Rost and Bock 2011). For example, criteria such as confidentiality influences how visibility and mutual awareness should be implemented.

3. Methodology

The aim of this study was to develop, improve, and consolidate the design of ST-heuristics. Using the approach proposed by Quiñones, Rusu, and Rusu (2018) as a foundation, we conducted the study in six phases. Figure 2 provides an overview. We outline the data sources that we used as a basis for our study in this section before elaborating on our approach for each of the six phases in Sections 3.1–3.6.

For this study, we recruited seven experts (E1 to E7) with different backgrounds, including computer science, information systems, sociology, ergonomics, and process management. All of them had prior experience related to the implementation of projects in the field of socio-technical systems.

We used two main data sources. The first data source consisted of seminal works about principles, criteria, or rules in six relevant domains as outlined previously in Section 2.3. As a second data source, we created a *problem database* related to socio-technical systems (Herrmann and Nierhoff 2019). This database contained 306 problems from 13 select cases, which we split into two groups to use during different parts or stages of our study. Table 1 lists the cases. Part 1 consisted of cases #1 to #9, which included a total of 223 problems, while Part 2 consisted of cases #10 to #13, which included 83 problems. The case selection was based on the following criteria:

- Cases should cover different domains (e.g. health care, industry, and education).
- Each case describes the use of a novel technology in the respective domain (e.g. high degrees of autonomy or algorithmically inferred decisions).
- Cases present a range of different socio-technical arrangements.

To identify the different problems within each case, we analysed different materials that included process models, project descriptions, transcripts of interviews with project members, documented conversations, and artifacts with written comments from different stakeholders. Each identified problem led to one entry in the database. These entries were derived from statements about problems and needs or from system features that were introduced to solve a problem. Table 1 provides an overview of the cases we considered and the numbers of problems that were derived from each case. In addition to the derived problems, the database also contains short descriptions for each case to provide additional context for the problem description, as demonstrated by the example of Case #4 in Table 2.

3.1. Phase 1: exploration

For this phase, we first selected 17 seminal publications discussed in the six domains we identified in Section 2. The selection was based on the following criteria:

- The article provides a list of aspects that are characterised as design guidelines, principles, or heuristics and that have relevance for their respective domain as well as for socio-technical design. We refer to those aspects as *literature items*.
- Those aspects are well introduced and usually have clear scientific impact.
- The identified article does not cover a special, restricted area but refers to a domain as a whole, such as HCI, CSCW, etc.

Table 3 contains an overview of the publications that were selected from each domain.

3.2. Phase 2: descriptive

Each of the identified 17 publications contains a number of heuristics or design principles that their authors have

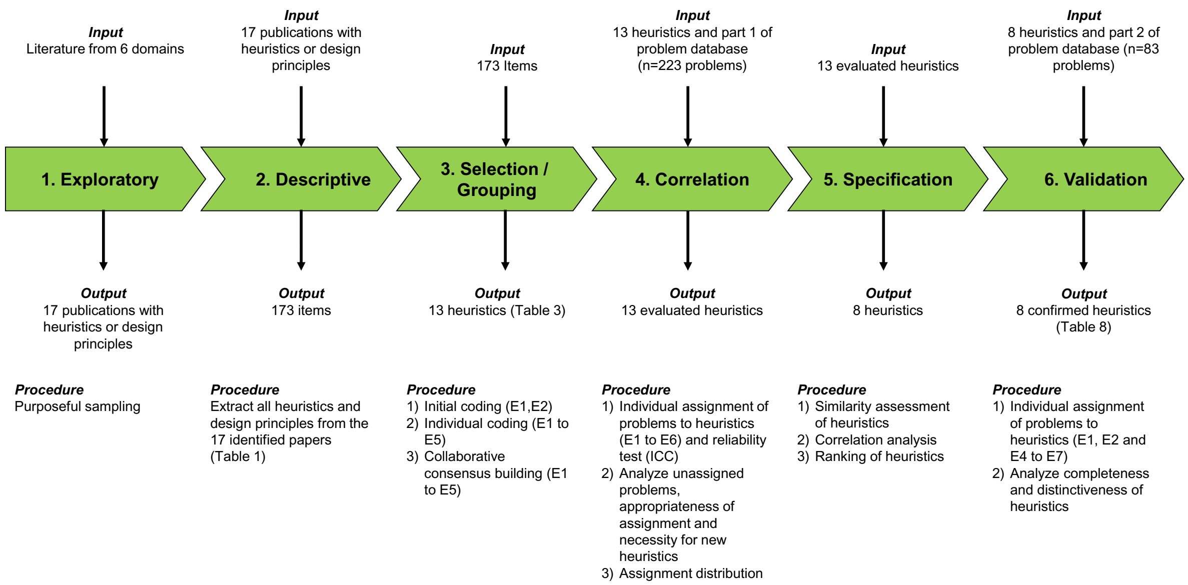

Figure 2. Study procedure (adapted from Quiñones, Rusu, and Rusu 2018).

identified and listed as relevant for the respective domain and that we considered important for the evaluation of socio-technical systems. We retrieved these heuristics and principles, which resulted in a total number of 173 items that served as a basis for our initial set of heuristics. The last column of Table 3 indicates the number of items that were extracted from each of the 17 publications we selected in Phase 1.

3.3. Phase 3: selection and grouping

To arrive at a manageable set of heuristics, we conducted a repetitive procedure of open and axial coding of the 173 identified literature items by using thematic analysis and the constant comparative method (Creswell 2013).

In general, we compared each segment of data (literature item) with the other segments to detect similarities and differences. Similar items were grouped together into one category. Open coding and axial coding were used across our data. The categories were tentatively given a name (Merriam and Tisdell 2015). The emerging bottom-up semantic codes (categories) became the label of a *theme* (a preliminary heuristic), ultimately leading to a set of different categories.

In detail, we started with open coding, meaning we coded the first item, then the second item, the third, and so forth until all 173 items were coded. As appropriate, literature items were coded using the emerging

Table 1. Overview of the 13 cases with numbers of problems (for generating the problem database).

Case number and name	Description	# of problems
1: Dent (Mårell-Olsson and Jahnke 2019)	Smart glasses for coordination of dentistry students and supervisors during practical training (workplace learning)	13
2: CreativB (Nierhoff and Herrmann 2017)	Continuous assessment of working climate regarding creativity (organisation studies)	11
3: LeXM (Jahnke et al. 2020)	Supporting orientation for new students via augmented reality technology (college orientation)	19
4: ElevatEd (Demir et al. 2019)	Data-based strategic improvement planning at a school (schools)	64
5: PeTEX (Jahnke et al. 2010)	Remotely controlled experimentation learning system for engineering students (engineering education)	12
6: Service4-Home (Herrmann, Prilla, and Nolte 2016)	Coordination of health care services for older adults (health care)	59
7: Web-update	Continuous updates of small organisational units' websites (small business, cooperative work)	5
8: eHealth (Ackermann et al. 2017)	Digital systems in health care (health care technologies)	29
9: Talk-Reflect (Prilla and Herrmann 2017)	Supporting the reflection of conversations with relatives of stroke patients (health Care)	11
10: Future-work-lab	Demonstrator-based use cases of future work lab (smart factories)	10
11: Predictive- maintenance (Herrmann 2020)	Predictive maintenance in the automotive industry (smart factories)	45
12: Manufacturing	Smart support system for assembly lines (smart factories)	10
13: Production of dental implant technology	Dynamic serial production of pieces for dental implants (smart factories)	18

Notes. We used the cases in two different study iterations. Cases 1 to 9 in Part 1 and cases 10 to 13 in Part 2.

Table 2. Case #4, ElevatEd (excerpt).

- Educational systems in the United States undergo reforms that involve the integration of data-intensive improvement processes known as Strategic Improvement Plans (SIPs). SIPs have the goal to improve teaching and especially learning outcomes. For example, they might examine what School 1 can do to improve the grades of a math class from average B to A grade […]. Schools turn to digital systems to set goals, create interventions for improving teaching and learning, use and analyse student data, and monitor and report SIPs. Challenges when launching digital systems include the integration of a highly diverse set of data sources and identifying the participants (end users) who will actually work with the new processes. This study explored how teachers and principals carry out SIPs in their schools including the technical tools they used. Examples of problems that were added into the problem database include:
- Unclear how to archive school improvement plans and for how long (ElevatEd P3-10)
- The wording of the IT system as developed by the software team is unclear to its actual users (e.g. the term *room* is actually a folder) (ElevatEd P1-2-22a)

categories from the previous set of items or added as a new category. This constant comparative method was used to compare the preliminary results of the categories. The data analysis was both iterative and recursive. The initial categories (set of coded items) helped to develop in-depth descriptions for each category as they were constantly refined to best represent the data. Afterwards, we sorted the categories from the open coding via the use of axial coding. Here, all coding for a category was closely examined to identify linkages and connections between them.

The previously described process was done iteratively by exploring the semantic overlap between different items. This process took place in two steps. In the first step, two experts (E1, E2) coded a set of categories which then served as a basis for a second step of coding. The second step of coding was to be conducted by five experts (E1 to E5).

To ensure the reliability of the results of this phase, emerging findings from individual analyses were discussed and compared. We checked data by peer conversations and applied a communicative discourse to check the quality of the coding (Creswell 2013). More specifically, when the experts had the same items coded into different categories, peer discussion about those items was initiated. To facilitate mutual agreement during the discussion, the decision was made that one item could be assigned to two or three different categories as long as the distinction between the categories was not blurred. For example, the item *aesthetic and minimalist design* was initially assigned to four different categories by the five experts: A, C, G, J. Eventually, after the discussion, this item was assigned to only two categories: Support of Information Exchange (F) and Visibility, Awareness, Feedback (J). In addition to the coding, we also developed a description for each category that translated to one

Table 3. Number of literature items found in the publications as heuristics, principles, or guidelines across six domains.

Six domains	Literature	# literature items
Human-computer interaction	Nielsen 1994a incl. discoverability Tognazzini 2014	11
	(Shneiderman and Plaisant 2005)	9
	Dix et al. 2003	14
	ISO 9241–110:2006	7
Computer-supported cooperative work	Greenberg et al. 2000	5
	Baker, Greenberg, and Gutwin 2001	7
	Herrmann, Wulf, and Hartmann 1996	8
Process redesign	Reijers and Limanmansar 2005	27
Socio-technical design	Clegg 2000	15
	Cherns 1987	11
	Eason 1989	10
Principles of job design	Mumford 1983	5
	Hackman and Oldham 1975	7
	Dunckel 1989	10
	Grote et al. 2000	16 (merged from 18)
Privacy	Rost and Bock 2011	6
	Clement 1994	5
	Total	**173**

heuristic. A total of 13 preliminary heuristics were coded (results in Section 4.1).

3.4. Phase 4: correlation

Following the lead of Quinones et al., in Phase 4, we investigated the match between our 13 preliminary heuristics and the problem database. As a starting point, we asked six experts[2] (E1 to E6) to individually assign each of the 223 problems contained in Part 1 of the problem database (cases 1 to 9 in Table 1) to the initial set of 13 heuristics. In addition, the experts were asked to answer the following questions for each problem:

1. Do you understand the problem? (Rate from 1 to 7.)
2. Which heuristic fits the problem best? (If no heuristic seems to match, the rater can select the option 'additional heuristic needed'.)
3. How confident are you in your choice? (Rate from 1 to 7.)
4. Are there other heuristics that also fit the problem? (0 to n secondary heuristics can be assigned.)

The experts also had the option to add a comment to each problem assignment. They were explicitly asked to add a comment if they thought there was a need for a new or modified heuristic. The reason behind allowing secondary heuristics to be assigned was to offer the experts an opportunity to express that they were not confident with just assigning one heuristic. That opportunity also allowed us to subsequently detect co-assignments and semantic overlaps of heuristics.

We analysed the results from this study iteration by using three questions.

- Is there sufficient reliability among expert judgements?
- To what extent is it possible to assign every socio-technical problem to at least one heuristic?
- How evenly or unevenly are the problems assigned among the heuristics?

In the first step, we analysed the individual assignments by calculating the Intraclass Correlation Coefficient (ICC), both overall and for each heuristic, to determine the reliability of the assignments within the expert group. The ICC was calculated for the primary assignments only and also by replacing primary assignments with those secondary assignments that complied with the majority of experts' opinions. ICC measures the reliability defined as the extent to which measurements can be replicated, in other words, both the degree of

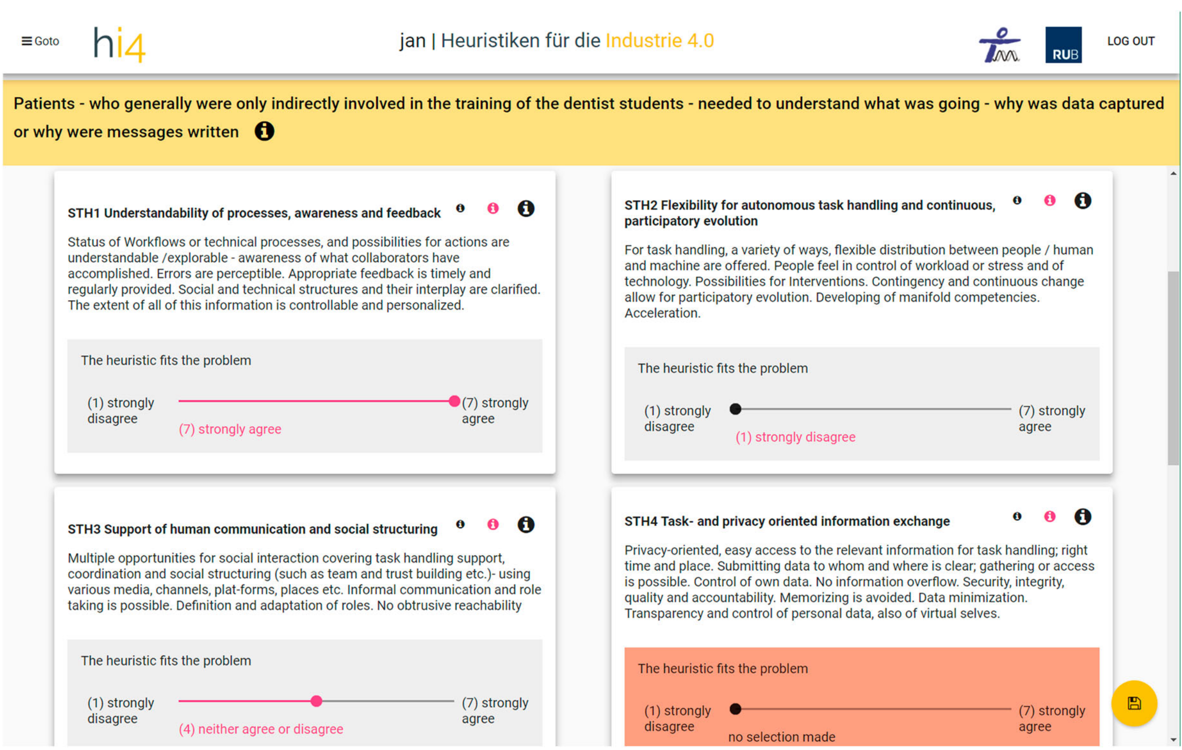

Figure 3. Screen shot from Step 3 rating assignment for one of the six experts. Note. Regarding the problem that is shown at the top of the screen, the rater matched Heuristic 1 as 'strongly fits (7/7)' while Heuristic 2 does not fit at all (1/7). He or she was unsure about Heuristic 3 (4/7) and had not yet rated the problem for Heuristic 4.

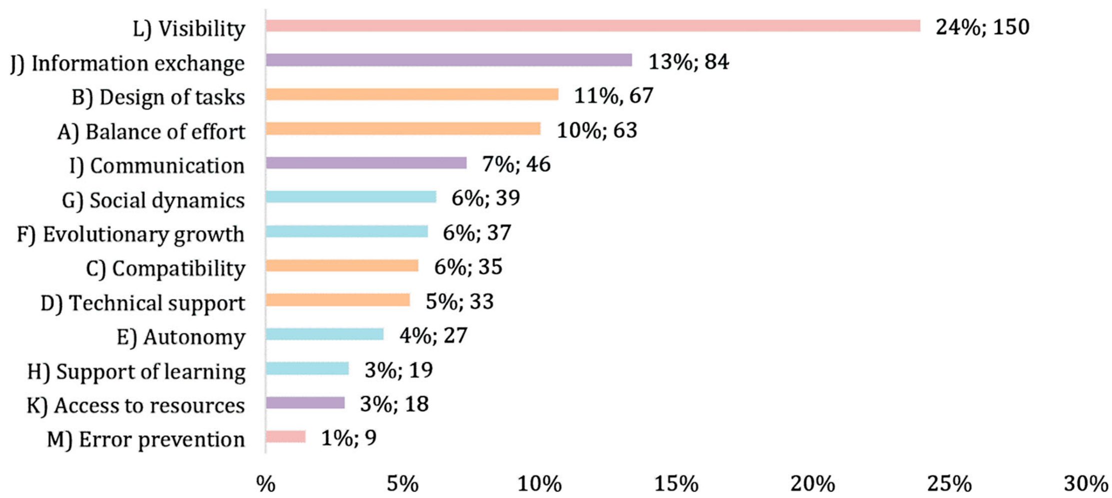

Figure 4. Ranking of how often heuristics were assigned to problems (mean = 7.62 and SD = 5.98 based on percentages).

correlation of the measurements and agreement between measurements (Koo and Li 2016).

In the second step, we counted problems where the experts (a) did not assign a heuristic, (b) felt unsure whether their assignment was appropriate, or (c) thought a new heuristic should be offered.

In the third step, we considered whether the numbers of assignments were evenly distributed between the heuristics. To do so, we selected the 50% of the problems where the assignments were made with the highest confidence and consistency and calculated how often each heuristic was assigned to them. The confidence was derived from the third questionnaire item mentioned above. To evaluate the consistency, we ranked the problems with an ordering method where a problem, *case x*, is ranked higher than *case y* if we found more pairs of two experts that assigned the same heuristic to x than y.

Each assignment of a heuristic to a problem was counted as one point. If an expert co-assigned several heuristics to the same problem, each of these assignments counted one point divided by the number of assigned heuristics. The sum of these points per heuristic was used as an indicator for how often the heuristic was used in assignments (Figure 4).

This procedure resulted in a set of 13 heuristics which served as the basis for the following phase.

3.5. Phase 5: specification

Based on the previous phases, we reduced the 13 heuristics to a smaller set since some of them were only assigned to a few cases. We selected those heuristics that were assigned fewest and viewed them as potential candidates for merging. To identify heuristics for which a merge was reasonable, we conducted a correlation analysis based on the same fifty percent of problems described in Section 3.4 (third step) which were ranked highest with respect to confidence and consistency. The correlation analysis focused on the experts and how they assigned different heuristics to the same problem.

Based on the analysis, we selected the heuristics with the highest significant correlations as a starting point. Subsequently, we identified the most promising candidates for merging of heuristics within the set of the nine lower ranked heuristics. The final decision whether and how the heuristics were completely or partially merged was based on the analysis of their semantic overlap and conceptual similarities between the heuristics. The result of this phase is the specification of a set of eight socio-technical heuristics.

3.6. Phase 6: validation

In this phase, we applied a two-step approach. For the first step, we again asked six experts[3] (E1, E2 and E4 to E7) to individually assign the remaining 83 problems contained in Part 2 of the problem database (light grey rows in Table 1) to the reduced set of eight heuristics. For this, we asked the experts to rate how well each of the eight heuristics matched a certain problem (rating from 1 to 7, 7=best match). Experts were thus asked to consider every heuristic and determine whether they agreed or not that it matched with the problem. Figure 3 shows an overview of the interface we used for the assignments. This approach stands in contrast to Phase 4, where experts only selected heuristics that fit a problem without stating whether the remaining heuristics sufficiently matched the problem as well or not. For the second step, we focused on the completeness and distinctiveness of the heuristics, and thus we selected methods to answer the following questions.

- Is there sufficient agreement between the expert judgements? Therefore, we calculated the Intraclass Correlation Coefficient (ICC) to assess the reliability of the expert assignments.
- Is the refined set of eight newly phrased heuristics sufficient to characterise the problems? Therefore, we counted the problems to which an expert did not agree on a level of 5 or higher that any of the heuristics fits the problem (Figure 3).
- Is a further merging of the heuristics reasonable or necessary? Therefore, we conducted a correlation analysis.
- How relevant are the eight heuristics in comparison to each other? Therefore, we counted the number of times a heuristic was assigned to a problem by every expert with an agreement level of at least 5 or higher (from 1 to 7, 1=strongly disagree, 7=strongly agree).

To keep this analysis compatible with our evaluation in Phase 4, we again ranked the problems with respect to the consistency between the experts' assignments and selected the 41 cases above the median of the ranking position.

4. Results

This section presents the findings from our study. It is organised along the methodological six phases of Figure 2 introduced earlier. The outcome of the Phase 1 and Phase 2 has been already presented in Table 3. Thus, we continue in Section 4.1 with the outcome of Phase 3.

4.1. Results of phase 3: selection and grouping

From the grouping and selection in Phase 3, all 173 literature items were coded into categories, named as preliminary heuristics. Every heuristic is labelled with a name and is described by a set of three to five aspects. Each of these aspects consists of similar coded items. As an example, Table 4 shows the 13 literature items of *Heuristic A: Balance of effort and experienced benefits, values, and goals*.

The description of Heuristic A consists of the following aspects.

- Interests, needs, values, goals, mindsets of, or problems owned by participants must be recognised and presented in the self-descriptions of the socio-technical process.
- The tasks to be carried out must be significant and meaningful with respect to the context (organisations, communities, etc.) of the socio-technical process.
- Efficiency where increased effort must lead to increased benefit with respect to the interests, etc. Tasks, and support to carry them out, must be functional with respect to these benefits.

Table 4. 13 items grouped in Heuristic A: Balance of effort and experienced benefits, values, goals.

Literature	Domain	Items of 'Balance of effort and experienced benefits, values, goals'
Clegg 2000, 465	Socio-technical design (STD)	'Design should reflect the needs of the business, its users, and their managers.'
Clegg 2000, 465	STD	Values and mindsets are central to design.
Dunckel 1989, 73	Job design (JD)	Körperliche Aktivität erfordern und ermöglichen. [Require and enable physical activity. (Translated by the authors.)]
Eason 1989, 47	STD	' … to serve the functional needs of the organisation by serving the functional needs of individual users … '
Eason 1989, 44	STD	'The successful exploitation of information technology depends upon the ability and willingness of the employees of an organisation to use the appropriate technology to engage in worthwhile tasks.'
Eason 1989, 45	STD	'The design target must be to create a socio-technical system capable of serving organisational goals, not to create a technical system capable of delivering technical services.'
Eason 1989, 47	STD	' … the socio-technical developments are directed at major organisational purposes where there are opportunities to be taken or problems to be solved'
Hackman and Oldham 1975, 161	JD	'Task significance. The degree to which the job has a substantial impact on the lives or work of other people—whether in the immediate organisation or in the external environment.'
Mumford 1983, 44	JD	'The PSYCHOLOGICAL 'fit'. Seeks to further personal interests, e.g. to have a sense of achievement, recognition, responsibility, status.'
Mumford 1983, 44	JD	'The EFFICIENCY 'fit'. Seeks an equitable effort-reward bargain, and controls, including supervisory ones, which are acceptable. Seeks efficient support services such as information, technical aids, supervisory help.'
Mumford 1983, 44	JD	'The ETHICAL (social value) 'fit': Seeks to work for an employer whose values do not contravene personal values.'
(Shneiderman and Plaisant 2005, 74)	Human-computer interaction (HCI)	'Cater to universal usability. Recognise the needs of diverse users and design for plasticity, facilitating transformation of content.'
(Shneiderman and Plaisant 2005, 66)	HCI	'Determine users' skill levels.'

Table 5. The initial set of 13 socio-technical heuristics.

Heuristic	Name (description)	# of literature items coded	# of coded items used in other heuristic(s)
A	Balance of effort and experienced benefits, values, and goals	13	B(2), C(1), D(1), F(1)
B	Suitable design of tasks and workflows	37	A(2), C(3), D(2), E(3), G(2), H(2), I(3), L(1),
C	Congruence of components and compatibility with reality	17	A(1), B(3), D(1), E(2), F(2), G(1), H(1), I(1), M(1)
D	Adequate, seamlessly integrated technical support	10	A(1), B(2), C(1), E(1), I(1), J(2), L(1), M(1)
E	Support of autonomy and flexibility	35	B(3), C(2), D(1), F(2), I(3), J(3), K(1), L(1), M(1)
F	Support of adaptation, change, and evolutionary growth	15	A(1), C(2), E(2), G(1), H(2),
G	Dealing with social dynamics	6	B(2), C(1), F(1), G(1)
H	Support of learning and development of competencies	15	B(2), C(1), F(2), L(2), M(2)
I	Support of human communication, cooperation, and coordination	23	B(3), C(1), D(1), E(3), G(2), J(4), L(2), M(1)
J	Support of proper information exchange/access	17	D(2), E(3), I(4), K(1), L(1), M(2)
K	Appropriate access to resources	6	E(1), J(1), M(1)
L	Visibility, awareness, and feedback	22	B(1), D(1), E(1), H(2), I(2), J(1), M(1)
M	Error prevention and support of error handling	15	C(1), D(1), E(1), H(2), I(1), J(2), K(1), L(1)

Note. Literature items with multiple meanings were coded into more than one heuristic.

- Individual differences (e.g. skill levels) must be considered by the tasks offered to the participants.

In total, the 173 items were coded into 13 different categories from Heuristic A to Heuristic M. Table 5 lists the preliminary 13 heuristics. It also includes the number of literature items coded and number of coded items used in other heuristics. This means, for example, in the case of the first line in Table 5, that 13 literature items are assigned to heuristic A and two of these 13 are also assigned to B, one to C, one to D, and one to F.

Overall, from the total of 173 available literature items, 171 were coded into at least one category (preliminary heuristic). The remaining two were meta-principles that did not fit into the proposed categories. Nine of these were coded to three different heuristics, 42 to two heuristics, and 120 items were coded to exactly one heuristic. The number of literature items coded into a heuristic varied from six to 37 (Table 5).

4.2. Results of phase 4: correlation

Phase 4 had the goal to analyse the quality of the 13 heuristics. Overall, the experts made 2144 assignments including 833 assignments of one or more secondary heuristics, making the average per problem 9.61 and the average per expert per problem 1.60.

To check the reliability of the measurements (i.e. assigning the heuristics to problems), we applied the Intraclass Correlation Coefficient (Koo and Li 2016).

Table 6. ICC results with six raters for Phase 4.

	ICC	95% Confidence Interval		F test with true value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Avg. Measures	.863	.832	.889	7.425	222	1110	.000

ICC estimates and their 95% confidence intervals were calculated using SPSS statistical package Version 26 (SPSS Inc, Chicago, IL) based on a mean-rating (k = 6), absolute-agreement, 2-way random model (see Table 6). The Intraclass Correlation Coefficient is ICC (*2*, *6*) = .863 (group of six raters), which indicates very good reliability of the measurements. Measurements here refer to the coded 223 problems assigned to 13 heuristics.

The relative high ICC reflects a high degree of rater and measurement agreement but also relates to a strong variability among the sampled subjects (cases), great number of subjects (problems), and great number of raters. Table 7 presents the ICC for every heuristic.

The individual ICC-analysis indicates that 10 of 13 heuristics tend towards a moderate agreement (0.5 - 0.75). There are three heuristics (A, G, L) that have a good agreement. The heuristic assignment was agreed upon by at least three experts for 158 (70.9%) problems, and four or more experts choose the same heuristic for 36.8% of the problems. For the second step of data analysis, we identified three types of criteria indicating that the set of heuristics or their descriptions were possibly incomplete or inappropriate.

(a) The experts asked for a new or modified heuristic by selecting the option ‘new heuristic,’ by writing a comment that asks for modifying the heuristics, or by writing a comment that asks for a new heuristic.

(b) The experts did not assign any heuristic.
(c) The experts clearly felt unsure with respect to the assignment they made.

Table 8 provides an overview of the three types of criteria (a-c), where criteria (a) has several indicators for incompleteness of the heuristics set. For indicator 1 of criteria (a), we found that for 22 problems, exactly one

Table 7. ICC results for every heuristic.

Heuristic	A	B	C	D	E	F	G	H	i	j	k	L	M
ICC (2, 6) with second-dary	.790	.586	.610	.697	.652	.695	.830	.660	.672	.670	.611	.795	.515

expert thought that a new heuristic was needed. Overall, for 26 problems (11.6%), a new heuristic was required. However, for only one problem (0.45%), three or more experts suggested creating a new heuristic The experts could also express, by use of comments, that they were not confident with the offered heuristics. They asked for modifications in 57 cases (indicator 4) and asked for a new heuristic in nine cases. However, there was no single problem where three or more experts asked for new heuristics or for a modification. For 63% of all problems, at least one expert felt uncertain when assigning a heuristic (indicator 5). However, only in 5.38% of the cases, three or more experts expressed this uncertainty for the same problem.

To support the final specification of the heuristics, we ranked the 13 heuristics based on how often they were assigned. In order to calculate the frequency, we focused on the subset of problems for which the experts were in relative high agreement (method Section 3.4, Step 1). There were 106 problems assigned to heuristics by six experts (total of 636 possible points) that we selected by applying the following thresholds: the experts' understanding of the problem (> = 4), confidence of assignment (> = 4), and the consistency ranking (> = 2,219, median). The ranking results in Figure 4 show a clear difference between the frequency of assignments in which *visibility* scored 150 points while *error prevention* only nine points.

4.3. Results of phase 5: specification

From Figure 4, we identified those nine heuristics that were below the percentage of average assignment (7.62). They are indicated by bold characters in Table 9, which displays the correlation analysis.

Applying the classification of Dancey and Reidy (2011) from 0.1–0.3 as weak, 0.4–0.6 as moderate, and 0.7–0.9 as strong, most of the correlations between the remaining nine heuristics were below weak and/or not significant. The significant but only weak or moderate correlations between *Error prevention*, *Access to resources*, and *Technical support* suggest that they can be merged into one single heuristic labelled *Technical Support* (Figure 5). This decision also appears reasonable from a qualitative point of view since most guidelines require that technology should be designed to be robust against errors, and technology also helps to prevent errors or to overcome their effects. Furthermore, in most cases of the problem descriptions, resources were associated with technological means. However, error prevention has more than only a technical dimension. We know from our projects and from literature (Edmondson 2004) that the way tasks are designed is also related to the avoidance of errors to support efficiency. Thus, efficiency and the avoidance of unnecessary steps, as they are related to error handling, are mentioned in Heuristic 7 (See Table 10).

We found another significant weak correlation between *autonomy* and *evolution* (0.35***). We grouped both together, thus creating a new heuristic labelled *Flexibility* (Table 10). This decision also complies with the semantical aspects of both heuristics since autonomy for carrying out adaptations by the involved people and their flexibility are prerequisites for continuous change and evolution. And, vice versa, offering options for evolution includes the support of flexibility.

Table 8. Indicators for incompleteness of the heuristic set.

Indicator	Criteria (a), (b), and (c)	Number of problems indicated by …				All affected problems		
#	The majority of the experts …	1 expert	2 experts	3 experts	4 or more	Absolute	Proportional (n=223)	3 or more experts (aggregated)
	(a) … required a new or modified heuristic.							
1	Request for a new primary heuristic	22	3	1	0	26	11.66%	0.45%
2	Request for a new aspect describing a heuristic	65	9	1	0	75	33.63%	0.45%
3	Request for a new secondary heuristic	6	0	0	0	6	2.69%	0,00%
	Sum of 1–3	**93**	**12**	**2**	**0**	**107**	**48%**	**0.9%**
4	Comment asking for modification of a heuristic	52	5	0	0	57	25.56%	0.00%
5	Comment asking for a new heuristic	9	1	0	0	10	4.48%	0.00%
6	(b) … did not assign a primary heuristic.	8	0	0	0	8	3.59%	0.00%
7	(c) … felt uncertain <4* (between 1 and 7) while matching.	88	41	10	2	141	63.23%	5.38%

Note. One expert is excluded as a statistical outlier (means of 2.70 vs. means of 4.85 of the 5).

Table 9. Correlation analysis.

		A	B	C	D	E	F	G	H	I	J	K	L	M
A	Balance													
B	Task design	0,13												
C	Compatibility	0,12	0.14*											
D	Technical Support	0,03	0,11	0,08										
E	Autonomy	0,08	0.35***	0,11	0,02									
F	Evolution	0,06	0.18**	0,05	0,06	**0.39*** **								
G	Social dynamics	0,03	0,12	0,02	−0,06	−0,01	**0.14***							
H	Learning	0,03	−0,02	0,12	−0,03	0,02	0,07	0,02						
I	Communication	−0,04	0.25***	0	0,07	−0,01	0	**0.15***	0,03					
J	Information exchange	−0,06	0,02	0,06	0,1	−0,03	−0,03	−0,08	0,13	0.24***				
K	Resources	−0,08	0,04	−0,02	**0.27*** **	0,06	0,01	−0,03	0,1	−0,04	0.27***			
L	Visibility	0,02	0,07	−0,08	−0,06	−0,01	−0,01	−0,01	0,03	0,08	0.34***	0,1		
M	Error prevention	−0,06	−0,06	0,1	**0.53*** **	−0,04	−0,04	0,01	0,03	0,12	0,07	**0.21*** **	0,02	

Notes. *** =*p*<.001; **=*p*< .01; *=*p*<.05; the 2nd column lists the abbreviations of the initial heuristics which are described in detail in Table 5.

In addition, *social dynamics* is related to *evolution* since the dynamics are driven by social interaction, and the patterns of social interaction, such as the negotiation of role taking, are subject to ongoing change. However, this correlation is only weak (0.14*), and while still significant, only on a 95% level. *Social dynamics* is also weakly correlated with *communication* (0.15*). Therefore, we add *social dynamics* to *communication* because of their semantic and conceptual similarity; the change of social relations mainly takes place by acts of communication according to Luhmann and Knodt (1996). Typical examples are the negotiation of duties and rights of roles (Table 10, Heuristic 3).

The remaining two heuristics—*learning support* and *compatibility*—are not significantly correlated to the other considered heuristics but have a clear conceptual overlap. The items integrated into *compatibility* require that tasks, technology, organisational procedures for support, competencies, and skill of people as well as customers' requirements fit together. This requires, as one aspect amongst others, that people are promoted and supported for the continuous development of their competencies and skills to manage newly developing challenges.

Figure 5 demonstrates how the 13 heuristics have been merged into a set of eight. The upper layer shows the most frequently assigned heuristics (L, A, B, J). The size of the circles indicates the frequency of assignments, as shown in Figure 4. The lower layers show heuristics that were assigned less than average (H, G, M, K). The oval lines represent the new heuristics (labelled in bold). *Autonomy* and *evolution* (E, F) are put together; *support of learning* is merged into *compatibility* (H to C); *social dynamics* is absorbed into *communication* (G to I); and *error prevention* and *access to resources* are moved into technical support (M and K to D). This is also indicated by the bold arrows while the thinner ones indicate the distribution of side aspects. Table 10 documents the final set of eight heuristics and answers RQ1: What does a manageable set of heuristics for the evaluation of socio-technical processes look like?

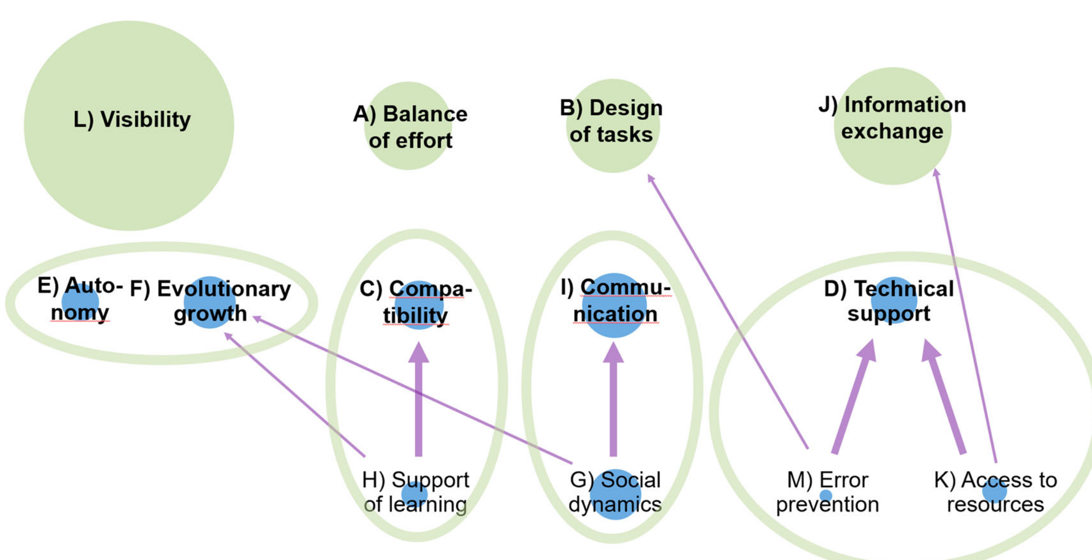

Figure 5. Reorganisation of the 13 heuristics into eight heuristics.

Table 10. The final set of eight heuristics for socio-technical evaluation (ST-heuristics).

#	ST-Heuristics with short description
1	***Visibility* about task handling and feedback about its success** Focused information is continuously offered about the progress of technical processes and, as far as permitted, about collaborative workflows. This helps to understand what further steps are possible or not and why as well as how well the expectations of others are met.
2	***Flexibility* for variable task handling leading to a participatory evolution of the system** One can vary options of task handling and can flexibly decide about technology usage, time management, sharing of tasks, etc. Consequently, one can develop a wide range of competencies that support participation in the ongoing evolution of the whole system.
3	***Communication* support for task handling and social interaction** By means of technical and spatial support for communication, a person can be reached to an influenceable extent for purposes of task handling and coordination This support is intertwined with negotiating duties and rights of roles, including values, so that reciprocal reliability can be developed.
4	**Purpose-orientated *information exchange* for facilitating mental work** To support task handling, information is purposefully exchanged via technical means, updated, kept available, and minimised. This implies technical linking of information and the emergence of personal profiles that must be visible and exchanged in compliance with privacy regulations.
5	***Balance of effort* and experienced benefit by organisational structuring of tasks** Organizational structuring of tasks supports a proportional balance between individuals' effort and experienced benefit. Tasks are assigned to people, pooled, and technically supported in a way that makes sense and is fun for people. Tasks comply with individuals' technical, social, and physical competencies while also supporting health. Thus, a sustaining balance of efforts and personal benefits is pursued.
6	***Compatibility* between requirements, development of competencies, and the system's features** Technical and organisational features of the system are continuously adjusted to work with each other. Within clarified limits, they meet outside requirements in a way that is based on the development of competencies and proactive help for dealing with changing challenges.
7	***Efficiency*-oriented allocation of tasks for pursuing holistic goals** By appropriate sequencing, integration, and distribution of tasks between humans and technology, seamless collaboration is supported. Unnecessary steps or wasting resources is avoided. If needed, an increase of efficiency can be realised.
8	***Supportive technology* and resources for productive and flawless work** Technology and further resources support work and collaboration and consider the intertwining of criteria, such as technology acceptance, usability and accessibility for different users, avoiding consequences of mistakes and of misuse, security, and constant updating.

Note. A more explicit description with examples is available for each heuristic (https://hi4.iaw.ruhr-uni-bochum.de/#!/manual).

4.4. Findings of phase 6: validation

In this phase, we aimed to assess the feasibility of the newly designed set of eight heuristics using 83 problems with the goal to understand whether the set is sufficient and whether further merging is reasonable or not.

The following analysis is based on 82 problems since the data collection did not work in one case (originally, there were 83 problems in the database). Table 11 shows the results of the intercoder-reliability using the Intraclass Correlation Coefficient that is ICC(2, 6) = .799 (group of six raters). Those problems were selected into the analysis with the experts' highest agreement per problem ('the heuristic fit to the problem,' scale 1–7, 7=strongly agree). It points towards a good to very good agreement (Koo and Li 2016). In 91.6% of the problems, at least three of six experts assigned the same heuristics, and in 79.5%, at least four experts assigned the same heuristic.

Table 12 presents the ICC for every heuristic. It shows that seven of the eight heuristics, received a higher interrater agreement than the previous 13 heuristics and that the exception, Heuristic 7 received a moderate agreement. This data indicates a good to very good agreement and an improvement from the 13 heuristics to the eight heuristics.

If an expert did not assign any heuristic to a certain problem with a score higher than four, we took this as an indicator that he or she thought that there was no feasible heuristic to characterise the problem. For 72 cases (88%), all six experts have assigned at least one heuristic with an agreement level higher than four (scale from 1 to 7, 1=strongly disagree, 7=strongly agree). For the remaining 10 cases (12%), only one expert had not selected a single heuristic with an agreement-level higher than four.

Regarding an agreement-level of more than five, six experts also assigned heuristics for 47 cases (57%). For 35 cases (43%), there was one expert who did not assign any heuristic with an agreement level higher than five, and for seven cases (9%), two experts did not assign any heuristic with an agreement level higher than five.

With respect to the question of whether a further merging of the heuristics to build a smaller set is reasonable, a correlation analysis did not reveal any clear hint to do so, as shown in Table 13. Only four correlations are positive. Three of these positive correlations are only weak in accordance with the classification of Dancey and Reidy (2011).

To calculate a ranking of the relevance of heuristics, we used the method of counting the

Table 11. ICC results with six raters for Phase 6.

		95% Confidence Interval		F test with true value 0			
	ICC	Lower Bound	Upper Bound	Value	df1	df2	Sig
Avg. Measures	.799	.723	.860	4.975	80	400	.000

Table 12. ICC results for every heuristic in the new set of eight.

Heuristic	1	2	3	4	5	6	7	8
ICC (2, 6) with secondary	.791	.842	.869	.797	.789	.846	.571	.805

reason for integrating these aspects into other heuristics is that their relevance and understanding depends on the context of a heuristic's topic. Privacy, for example, is important in the context of visibility, communication support, and information exchange.

Table 13. Correlation analysis between different heuristics (∗∗∗ $p < 0.001$, ∗∗ $p < 0.01$, ∗ $p < 0.05$).

		STH1	STH2	STH3	STH4	STH5	STH6	STH7	STH8
STH1	Visibility								
STH2	Flexibility	−0.32**							
STH3	Communication	0.01	−0.14						
STH4	Information exchange	0.28*	−0.33**	−0.04					
STH5	Balance	−0.17	−0.02	−0.24*	−0.13				
STH6	Compatibility	−0.22	−0.01	−0.09	−0.33**	−0.19			
STH7	Efficiency	−0.22	0.28*	−0.15	−0.17	0.24*	−0.23		
STH8	Technical Support	−0.29*	−0.02	−0.22	−0.03	−0.09	−0.02	−0.07	

Note. These socio-technical heuristics (STH) were selected into the analysis with the experts' highest agreement per problem ('the heuristic fit to the problem,' scale 1–7, 7=strongly agree).

assignments as described in Section 3.3. Figure 6 shows the ranking. The analysis is based on the 50% of heuristic assignments with the highest consensus between the experts. In other words, value for the consistency ranking was not smaller than the median of 3.22.

5. Discussion

RQ1 asks about a manageable set of heuristics to evaluate the intertwinement between social practices and technical artifacts. Our study shows that the final set of eight heuristics sufficiently cover the scope of the relevant problems. Out of a set of 173 items, we first created and evaluated a set of 13 heuristics, which was eventually condensed to a set of eight. The scope of the eight heuristics can be considered sufficiently complete. While the number of indicators requiring a new heuristic was modest for the set of 13 heuristics (Table 8), the number of these indicators became marginal for the set of eight, where at least five of six experts found a heuristic for each problem (Section 4.4). For the 13 heuristics, in 48% of the problems (Table 8, Section 4.2), a new heuristic or a new aspect of its description was requested by at least one of the six experts, but there was no case where such a request was explicitly made by a majority (four or more) of the experts. For the condensed eight heuristics (Section 4.4), there were only nine problems (11%) where a single expert made such a request, but for all problems, at least five experts assigned a heuristic. We take this as evidence that the final heuristics sufficiently cover the scope of relevant problems.

Based on the design rationale for the condensed set of heuristics (Section 4.3), there are distinct topics that are addressed by various heuristics. These are mainly privacy, suitability for learning, and error prevention. The

With respect to conciseness and understandability (RQ2), one indicator is the certainty-level that was indicated by the experts for every assignment, or the level of agreement between experts. In Phase 4 (Table 9), for 63% of all 223 problems, at least one expert's certainty about his/her assignment was below a rating of four on a scale of one to seven (Section 4.2). This improved to only 12% of all 82 problems for the newly designed eight heuristics (Section 4.4). Furthermore, the comparison between the set of 13 and eight heuristics reveals that the percentage of problems increased where at least four experts selected the same heuristic from 64.6% in Phase 4–84.5% in Phase 6. Our interpretation of this finding is that the heuristics achieved a sufficient level of understandability. A further indication is the improvement of the interrater agreement where seven of the eight final heuristics have a good agreement (Table 11) while this was the case for only three out of 13 in Phase 4 (Table 6).

The conciseness of the heuristics is evident by how well the heuristics can be distinguished from each other. The distinctiveness can be analysed with respect to the development from Phase 3 to Phase 6. In Phase 3, the majority of items, 120 out of 173 (69%), was assigned to only one heuristic, 42 to two heuristics, and nine to three different heuristics. In Phase 5, the correlation analysis reveals 13 positive, significant but weak correlations (16.66% of 78 possible correlations) while only two weak correlations (0.07% of 28 possible) were found for the eight heuristics (*visibility* and *information exchange*; *balance of effort* and *efficiency*).

We take this development as an indication that the semantic overlaps between the heuristics were reduced more and more over the course of the six phases and that the last eight heuristics are sufficiently and clearly separate from each other.

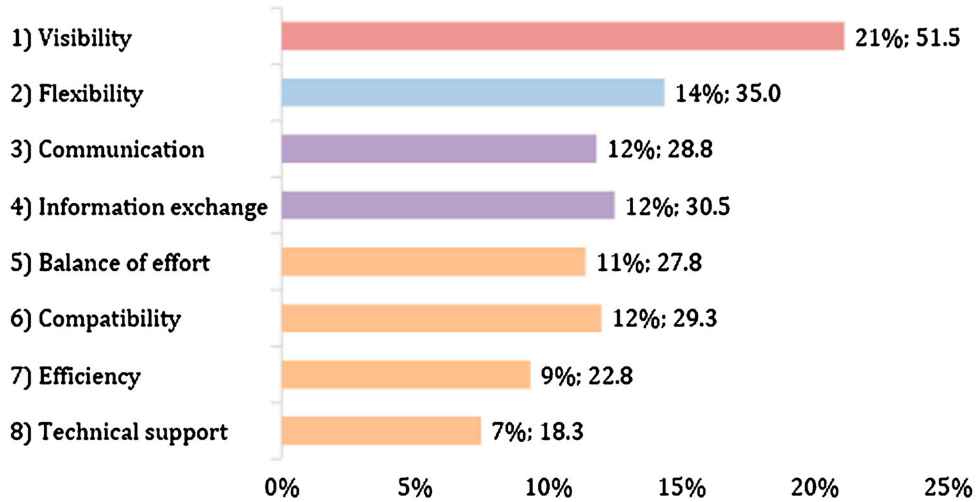

Figure 6. Ranking heuristics based on heuristic assignments for 42 problems by six experts. Note. Each assignment of a heuristic to a problem was counted as one point. If a problem was assigned to several heuristics with an agreement level higher than 4, each of those cases counted one point divided by the number of assignments. (m=12.2, SD=4.1).

5.1. Advancing existing approaches of socio-technical evaluation

RQ3 asks whether the new heuristics go beyond existing heuristics. Literature regarding socio-technical evaluation indicates the main difference between our new socio-technical heuristics approach and existing ones. The existing approaches developed sets of evaluation heuristics or criteria for certain domains. They go into depth for their sector. For example, Carayannis et al. (2018) developed a tourist destination evaluation index, Lowe, Chiu, and Oreszczyn (2018) focused on construction and buildings, Nolte et al. (2018) on volunteer groups, and Moallemi et al. (2017) and Schnuelle et al. (2019) discuss evaluation criteria for the energy and electricity sector. Other examples contain the socio-technical evaluation of a social credit application (Ferreira et al. 2019), cryptocurrency (Desmond, Lacey, and Salmon 2019), and cyber-physical production systems (Noehring et al. 2019). These evaluation approaches focus on specific socio-technical systems in a certain domain, but they cannot be applied beyond their domain. That is the main difference in our study. Our new ST-heuristics do not focus on a certain domain or sector, but rather provide a general approach for all socio-technical intertwinements, where social practices, technology and human–computer interaction are intertwined. Our study suggests that the here developed ST-heuristics can be seen as an advancement of Nielsen's approach (Nielsen 1994b) in that they support the evaluation of a variety of different systems.

The here developed ST-heuristics offer an extension of the various sets of principles that are present in the investigated domains. This can be demonstrated by the additions to the heuristics of human–computer interaction, e.g. the aspect of visibility (Nielsen 1994a). While conventional usability heuristics focus on the visibility of the status of the technology, our heuristic of *visibility* (Heuristic 1) also includes the visibility of the progress of the people on whom one depends when collaborating and goes beyond technological visibility.

Our new set of ST-heuristics also expand the domain of CSCW. For example, our new heuristic *communication* (Heuristic 3) is expanded from the CSCW discourse on awareness support (Dourish and Bellotti 1992), which requires that the status and further possibilities of a technically supported collaboration is made visible. However, the way we frame the heuristics of visibility and communication (Heuristics 1 and 3) expands to the concept of awareness in that our understanding also includes the reasons behind certain situations or activities as well as the degree to which the expectations of others are met. The latter aspect is related to insights of job redesign and job satisfaction since giving feedback to workers is a central measure of increasing motivation. At the same time, the heuristic demands the limitation of the visibility of the activities of others with respect to privacy.

Furthermore, our heuristic of *communication support* (Heuristic 3) also adds new aspects to the listed CSCW principles (Baker, Greenberg, and Gutwin 2001; Greenberg et al. 2000) since the relevance of social dynamics is emphasised. For instance, social dynamics refer to the negotiation of roles and their rights and duties as well as to the dynamics of role taking (Jahnke, Ritterskamp, and Herrmann 2005). Furthermore, the integration of informal roles and the transition from informal to formal roles complements conventional criteria for workplace redesign in the domain of job design and work.

Technical Support (Heuristic 8) is closely related to usability since this heuristic refers to the technological characteristics of a socio-technical system. However, the inclusion of IT-Security, avoidance of misuse, and constant updating of tools or technologies in our new heuristic provide additional aspects compared to the principles that already existed. Additionally, technology is considered in the context of further resources such as the material to be processed. The fact that the heuristic of *technical support* appears in our study with less relevance reflects the focus on the broader socio-technical scope and the fact that several technical aspects are integrated in other heuristics.

If we compare our results with other literature that provides a more holistic view on socio-technical systems, we find additional interesting differences. For example, most aspects that Andriessen (2003) addresses are included in our heuristics. However, his heuristics refer to the technology and consider requirements for its compliance with the organisational and human environment. By contrast, we refer to the socio-technical system as a *whole* where social practices are also a subject of the evaluation, as the list of the eight ST-heuristics show (Table 10).

In addition, Andriessen (2003) does not cover the aspect of *visibility* (Heuristic 1), which we found to be one of the most prominent heuristics. This might result from a theoretical blind spot compared to our bottom-up approach in which most heuristic aspects were derived from existing principles that refer to practical problems. Furthermore, Andriessen mentions the aspect of adaptability of systems but not the necessity of support for ongoing evolution as we do with the second heuristic (Heuristic 2, *flexibility*) by indicating flexibility as a prerequisite and promotor of ongoing evolution. Similar to Andriessen, the method MEESTAR emphasises self-determination and participation as relevant principles, but not the support of ongoing evolution.

MEESTAR is compatible with our approach since it refers to socio-technical arrangements as a whole. It emerged in the context of technical support for care for needy persons but is also transferred to other areas such as collaborative robots (Nelles et al. 2017). The context of care includes specific heuristics such as avoiding a negative influence of technology on people's self-conception. This aspect of MEESTAR is only indirectly referenced by our approach with the Heuristic 5 (*balance between effort and experienced benefit*), which addresses the interests and values pursued by the involved people. The benefit orientation of Heuristic 5 promotes additional effort as long as it contributes to people's and groups' interests and values from the viewpoint of their self-conception.

Shin emphasised that fairness, transparency, and accountability are the most relevant issues (Shin 2020; Shin and Ibahrine 2020). While transparency can be immediately aligned with visibility, the other issues are indirectly relatable with Heuristic 4 (*information exchange*) and Heuristic 5 (*balance*). Furthermore, the whole set of the new heuristics is relevant as a prerequisite to realise fairness and accountability. A recent synthesis of socio-technical principles (Imanghaliyeva et al. 2020) is rather oriented on recommendations for designers and less on purposes of evaluation. Imanghaliyeva et al. pursue a holistic approach and include aspects, such as designing design and flux, that point to a need for ongoing evolution of socio-technical systems. However, the synthesis does not include an orientation toward values and benefits as our Heuristics 5, 6, 7 do. It is also noticeable that Imanghaliyeva et al. approach does not include *visibility* (our Heuristic 1).

5.2. Implications

The eight ST-heuristics can be used for further research and practice in the field of user experience and socio-technical systems development.

Our analysis indicated that evaluators who use our heuristics to evaluate socio-technical intertwinements do not need to be experts with respect to every aspect. It is sufficient for them to be experts or practitioners who are familiar with certain aspects of the system, as it is the case with managers, UX-designers, software engineers, operative forces, or members of the work council.

Nielsen (1994a) found that more than one evaluator is necessary to identify the majority of serious problems of an interactive system. We assume that this is particularly true for the evaluation of socio-technical intertwinements. Different experts will have different foci on each heuristic that depend on their experience and professional background. To avoid blind spots and avoid the need of intensive training to become familiar with all the details of the ST-heuristics, it is reasonable to involve several evaluators so that their perspectives combine to create a holistic view of the system under evaluation. What supports the detection of relevant problems of the quality of a socio-technical intertwinement is the fact that the same problem can be detected from different points of view and by applying more than only one heuristic. In the end, it is not significant which heuristic leads to the discovery of the problem. However, it is crucial that the number of different perspectives (of experts and of heuristics) adds to the probability that the problem is identified.

Similar to existing usability or user experience (UX) labs, our ST-heuristics are a method for investigating and developing broader socio-technical ecologies. Furthermore, the heuristics can be employed to identify information gaps about the STS. That means that by applying the heuristics, evaluators find certain issues where they recognise that they have to ask for further information before they can completely understand whether the system is in compliance with a certain heuristic. Overall, the ST-heuristics can additionally serve as guidance, like the items of a questionnaire, to help asking relevant information, as one tries to understand how an STS works (Herrmann et al. 2004).

For researchers, our study reveals new research questions that are relevant and can be investigated in the future.

- **Theory:** How do the multiple aspects included in the heuristics and their relationship contribute to theoretical considerations? They can refer to the discussion about the symmetry versus asymmetry of the social and technical dimension or to the overlap between material agency and social agency.
- **Method**: Which procedure of a heuristic-based inspection or walk-through proves to be most promising for applying the proposed heuristics? For instance, should the whole system be walked through eight times, each time with one heuristic as a focus?
- **Generalizability**: Do different areas of socio-technical systems (e.g. health care with manufacturing) need an extension by adding different descriptions of details to the heuristics?

5.3. Limitations

The aim of our study was to develop and evaluate a set of heuristics to identify problems within socio-technical systems. We followed the methodology proposed by Quiñones, Rusu, and Rusu (2018) as a basis for the development of our heuristics. Limitations include a focus on six selected domains as a basis to extract principles that formed the basis for our study. There may, however, be other relevant areas we did not consider. To address this limitation, we considered a vast volume of literature from a variety of different domains (Table 3) ranging from process design to privacy. Then, we evaluated the developed heuristics based on a problem database relying on the assumption that the selection of cases sufficiently represents the scope of possible socio-technical systems. To mitigate this limitation, we intentionally selected problems from different domains.

Furthermore, the problems included in the database were extracted by different individuals based on selected material, workflow diagrams we have developed by ourselves, or interviews that we have conducted. Some problems were directly spelled out by interviewees or by documented comments. Other problems were indirectly derived by analysing certain features that were introduced to react to a problem. To mitigate potential biases, we asked multiple individuals to contribute to the database. Finally, the analysis was conducted by manually assigning heuristics to problems which might induce interpreter bias. To mitigate this limitation, we asked seven experts from different disciplines to independently conduct the assignment before calculating their agreement scores and solving potential issues collaboratively. Moreover, we provided the possibility to suggest adding additional heuristics during all steps of the process to ensure that the proposed heuristics are complete.

6. Conclusion

Literature analysis reveals that there is a lack of heuristics to evaluate socio-technical intertwinements. The goal of this study was to design an advanced set of ST-heuristics, as shown in Table 10, that are useful for evaluating and detecting problems in socio-technical systems. These eight heuristics are, in short,

- **visibility** about task handling and feedback about its success;
- **flexibility** for variable task handling leading to a participatory evolution of the system;
- **communication support** for task handling and social interaction;
- Purpose-oriented **information exchange** for facilitating mental work;
- **balance of effort** and experienced benefit by organisational structuring of tasks;
- **compatibility** between requirements, development of human competencies, and the system's features;
- **efficiency**-oriented allocation of tasks for pursuing holistic goals; and
- **supportive technology** and resources for productive and flawless work.

Results show a manageable set of eight ST-heuristics that are sufficiently specified toward scope, conciseness, and understandability. The advanced ST-heuristics extend existing heuristics as they can be found in domains, such as HCI, CSCW, process (re)design, socio-technical design, job design, and privacy. In particular, our ST-heuristics do not focus on technical systems only (e.g. assembly robot or technology interfaces) but on the intertwinement of social and technical dimensions (e.g. the setting up of a robot by people in a factory). Overall, it is not our aim to replace other methods, such as usability inspection with the ST-heuristics, but rather to complement them. An obvious example is the aspect of visibility that is included within usability inspection methods but neglected in other holistic approaches for socio-technical evaluation (Andriessen 2003; Imanghaliyeva et al. 2020; Wutzkowsky and Böckmann 2018).

The ST-heuristics provide a foundation to identify starting points for discussing improvement potentials. If certain problems are detected, it helps to initiate a discourse with the involved decision makers. We suggest that an STS cannot be designed and implemented in a first draft in such a way that it completely complies with the heuristics. Rather, several cycles of evaluation and improvement are necessary. The advantages of heuristics lie in their potential to not only be applied by experts but also by non-experts in practical settings. They aid evaluators in achieving goals such as:

- pragmatic and fast evaluation for detecting most serious problems;
- discovery of problems that would not be recognised without the heuristics;
- finding problems by systematizing the search with the heuristics (like using a checklist);
- naming problems and referring to them in discussions; and
- allowing evaluators to understand that they need more information.

In this sense, heuristics support a systematic and agile evaluation process or intuitive assessment of the quality of socio-technical systems by offering various perspectives and avoiding blind spots.

In future work, we will explore the effectiveness of the ST-heuristics by investigating how users apply them in practice.

Notes

1. The visibility principle says, ‘The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.’ (https://www.nngroup.com/articles/ten-usability-heuristics/).
2. We added one individual to the group of experts for this phase because of her/his background in computer science and STS design, which increased the group to six experts rather than the originally planned group of five.
3. One expert had to be replaced because the original expert was not available during the time period of this phase.

Acknowledgment

We are grateful to Dr. Rainer Skrotzki, who helped during the grouping and discussions of the items and heuristic assignments. We thank Rolf Molich for his comments on the first version of the heuristics and Felix Thewes for the technical and organisational support of the empirical work with the problem database. Many thanks go to Dr. Jan Nierhoff for data collection and commenting on previous versions.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was funded by the Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen, Germany.

ORCID

Thomas Herrmann http://orcid.org/0000-0002-9270-4501
Isa Jahnke http://orcid.org/0000-0001-6255-6416

References

Ackermann, M. S., S. P. Goggins, T. Herrmann, M. Prilla, and C. Stary. 2017. *Designing Healthcare That Works – A Socio-technical Approach*. London: Academic Press.

Amir, S., and V. Kant. 2018. “Sociotechnical Resilience: A Preliminary Concept.” *Risk Analysis* 38 (1): 8–16.

Andriessen, J. H. E. 2003. *Working with Groupware*. London: Springer. doi:10.1007/978-1-4471-0067-6.

Baker, K., S. Greenberg, and C. Gutwin. 2001. “Heuristic Evaluation of Groupware Based on the Mechanics of Collaboration.” In *Engineering for Human-Computer Interaction*, Vol. 2254, edited by M. R. Little, and L. Nigay, 123–139. Berlin Heidelberg: Springer. doi:10.1007/3-540-45348-2_14.

Baxter, G., and I. Sommerville. 2011. “Socio-technical Systems: From Design Methods to Systems Engineering.” *Interacting with Computers* 23 (1): 4–17. doi:10.1016/j.intcom.2010.07.003.

Behrenbruch, K., O. Kieselmann, M. Schuldt, M. Söllner, and L. Schmidt. 2014. “System EvaluationEvaluation.” In *Socio-technical Design of Ubiquitous Computing Systems*, edited by K. David, K. Geihs, J. M. Leimeister, A. Roßnagel, L. Schmidt, G. Stumme, and A. Wacker, 297–319. Springer International Publishing. doi:10.1007/978-3-319-05044-7_17.

Bødker, S. 2015. "Third-wave HCI, 10 Years Later—Participation and Sharing." *Interactions* 22 (5): 24–31.

Booth, R. G., M.-A. Andrusyszyn, C. Iwasiw, L. Donelle, and D. Compeau. 2016. "Actor-Network Theory as a Sociotechnical Lens to Explore the Relationship of Nurses and Technology in Practice: Methodological Considerations for Nursing Research." *Nursing Inquiry* 23 (2): 109–120.

Bruce, B. C., A. Rubin, and J. An. 2010. "Situated Evaluation of Socio-Technical Systems." In *Social Computing: Concepts, Methodologies, Tools, and Applications*, edited by B. C. Bruce, A. Rubin, and J. An, 2211–2225. IGI Global.

Carayannis, E. G., F. A. F. Ferreira, P. Bento, J. J. M. Ferreira, M. S. Jalali, and B. M. Q. Fernandes. 2018. "Developing a Socio-Technical Evaluation Index for Tourist Destination Competitiveness Using Cognitive Mapping and MCDA." *Technological Forecasting and Social Change* 131: 147–158. doi:10.1016/j.techfore.2018.01.015.

Carroll, J. M., and R. L. Campbell. 1989. "Artifacts as Psychological Theories: The Case of Human-Computer Interaction." *Behaviour & Information Technology* 8 (4): 247–256.

Cherns, A. 1987. "Principles of Sociotechnical Design Revisted." *Human Relations* 40 (3): 153–162.

Clegg, C. W. 2000. "Sociotechnical Principles for System Design." *Applied Ergonomics* 31 (5): 463–477. doi:10.1016/s0003-6870(00)00009-0.

Clement, A. 1994. "Considering Privacy in the Development of Multi-Media Communications." *Computer Supported Cooperative Work (CSCW)* 2 (1): 67–88. doi:10.1007/BF00749284.

Cockton, G., D. Lavery, and A. Woolrych. 2008. "Inspection-based Evaluations." In *The Human-Computer Interaction Handbook*, edited by A. Sears and J. Jacko. Boca Raton: CRC Press.

Creswell, J. W. 2013. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage Publications Incorporated.

Dancey, C., and J. Reidy. 2011. *Statistical Without Math for Psychology*. Harlow: Pearson Education.

Demir, F., S. M. Kim, N. Current, and I. Jahnke. 2019. Strategic improvement planning in schools: A sociotechnical approach for understanding current practices and design recommendations: *Management in Education*. doi:10.1177/0892020619847681.

Desmond, D. B., D. Lacey, and P. Salmon. 2019. "Evaluating Cryptocurrency Laundering as a Complex Socio-Technical System: A Systematic Literature Review." *Journal of Money Laundering Control* 22 (3): 480–497. doi:10.1108/JMLC-10-2018-0063.

Dix, A., J. E. Finlay, G. D. Abowd, and R. Beale. 2003. *Human-Computer Interaction*, 3rd ed. Essex: Pearson.

Dourish, P., and V. Bellotti. 1992. "Awareness and Coordination in Shared Workspaces." *Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work - CSCW '92*, 107–114. doi:10.1145/143457.143468.

Dunckel, H. 1989. *Arbeitspsychologische Kriterien zur Beurteilung und Gestaltung von Arbeitsaufgaben im Zusammenhang mit EDV-Systemen*. B.G.Teubner. http://dl.gi.de/handle/20.500.12116/6328.

Eason, K. D. 1989. *Information Technology And Organisational Change*. CRC Press. doi:10.1201/9781482275469.

Edmondson, A. C. 2004. "Learning from Mistakes is Easier Said than Done: Group and Organizational Influences on the Detection and Correction of Human Error." *The Journal of Applied Behavioral Science* 40 (1): 66–90. doi:10.1177/0021886304263849.

Emery, F. E., E. Thorsrud, and E. Thorsrud. 2013. *Form and Content in Industrial Democracy: Some Experiences from Norway and Other European Countries*. Routledge. doi:10.4324/9781315013695.

Ferreira, F. A. F., J. P. Esperança, M. A. S. Xavier, R. L. Costa, and B. Pérez-Gladish. 2019. "A Socio-Technical Approach to the Evaluation of Social Credit Applications." *Journal of the Operational Research Society* 70 (10): 1801–1816. doi:10.1080/01605682.2017.1415650.

Fischer, G., and T. Herrmann. 2015. "Meta-design: Transforming and Enriching the Design and Use of Socio-technical Systems." In *Designing Socially Embedded Technologies in the Real-World*, edited by V. Wulf, K. Schmidt, and D. Randall, 79–109. London: Springer. doi:10.1007/978-1-4471-6720-4_6.

Ghaffarian, V. 2011. "The new Stream of Socio-Technical Approach and Main Stream Information Systems Research." *Procedia Computer Science* 3: 1499–1511. doi:10.1016/j.procs.2011.01.039.

Greenberg, S., G. Fitzpatrick, C. Gutwin, and S. Kaplan. 2000. "Adapting the Locales Framework for Heuristic Evaluation of Groupware." *Australasian Journal of Information Systems* 7 (2), doi:10.3127/ajis.v7i2.267.

Grote, G., C. Ryser, T. Wāler, A. Windischer, and S. Weik. 2000. "KOMPASS: A Method for Complementary Function Allocation in Automated Work Systems." *International Journal of Human-Computer Studies* 52 (2): 267–287. doi:10.1006/ijhc.1999.0289.

Hackman, J. R., and G. R. Oldham. 1975. "Development of the Job Diagnostic Survey." *Journal of Applied Psychology* 60 (2): 159–170. doi:10.1037/h0076546.

Herrmann, T. 2020. "Socio-Technical Design of Hybrid Intelligence Systems–The Case of Predictive Maintenance." *International Conference on Human-Computer Interaction*, 298–309.

Herrmann, T., M. Ackerman, S. Goggins, C. Stary, and M. Prilla. 2017. Designing Health Care That Works—Socio-technical Conclusions. In *Designing Healthcare That Works: A Socio-technical Approach* (pp. 187–203). Academic Press (Elsevier). doi:10.1016/B978-0-12-812583-0.00011-0.

Herrmann, T., I. Jahnke, and A. Nolte. 2016. "Evaluating Socio-Technical Systems with Heuristics – a Feasible Approach?." Proceedings of STPIS16. http://ceur-ws.org/Vol-1604/Paper9.pdf

Herrmann, T., G. Kunau, K. U. Loser, and N. Menold. 2004. "Socio-technical Walkthrough: Designing Technology Along Work Processes." *Proceedings of the Eighth Conference on Participatory Design: Artful Integration: Interweaving Media, Materials and Practices* 1: 132–141.

Herrmann, T., and J. Nierhoff. 2019. *Heuristiken zur Evaluation digitalisierter Arbeit bei Industrie 4.0 und KI-basierten Systemen aus soziotechnischer Perspektive*. http://www.fgw-nrw.de/studien/industrie16.html, FGW, Düsseldorf.

Herrmann, T., M. Prilla, and A. Nolte. 2016. "Socio-Technical Process Design—The Case of Coordinated Service Delivery for Elderly People." In *Blurring the Boundaries Through Digital Innovation*, edited by F. D'Ascenzo, M. Magni, A. Lazazzara, and S. Za, Vol. 19, 217–229. Springer International Publishing. doi:10.1007/978-3-319-38974-5_17.

Herrmann, T., V. Wulf, and A. Hartmann. 1996. "Requirements for the Human Centred Design of Groupware." *The Design of Computer Supported Cooperative Work and Groupware Systems* 12: 77–100.

Hertzum, M. 2010. "Images of Usability." *International Journal of Human-Computer Interaction* 26 (6): 567–600. doi:10.1080/10447311003781300.

Imanghaliyeva, A. A., P. Thompson, P. Salmon, and N. A. Stanton. 2020. "A Synthesis of Sociotechnical Principles for System Design." *International Conference on Applied Human Factors and Ergonomics*, 665–676.

Interaction Design Foundation. 2021. *Heuristics and Heuristic Evaluation*. https://www.interaction-design.org/literature/book/the-glossary-of-human-computer-interaction/heuristics-and-heuristic-evaluation.

Jahnke, I., M. Meinke-Kroll, M. Todd, and A. Nolte. 2020. "Exploring Artifact-Generated Learning with Digital Technologies: Advancing Active Learning with Co-design in Higher Education Across Disciplines." *Technology, Knowledge and Learning*. doi:10.1007/s10758-020-09473-3.

Jahnke, I., C. Ritterskamp, and T. Herrmann. 2005. "Sociotechnical Roles for Sociotechnical Systems—A Perspective from Social and Computer Sciences." *AAAi Fall Symposium Proceedings*, 68–75.

Jahnke, I., C. Terkowsky, C. Pleul, and A. E. Tekkaya. 2010. *Online Learning with Remote-Configured Experiments*. Gesellschaft für Informatik e.V. http://dl.gi.de/handle/20.500.12116/4730.

Kensing, F., J. Simonsen, and K. Bodker. 1998. "MUST: A Method for Participatory Design." *Human–Computer Interaction* 13 (2): 167–198. doi:10.1207/s15327051hci1302_3.

Kirsch, C., P. Troxler, and E. Ulich. 1995. "Integration of People, Technology and Organization: The European Approach." In *Symbiosis of Human and Artifact*, edited by Y. Anzai, K. Ogawa, and H. Mori, Vol. 20, 957–961. Elsevier. doi:10.1016/S0921-2647(06)80337-0.

Kleiner, B. M. 2006. "Macroergonomics: Analysis and Design of Work Systems." *Applied Ergonomics* 37 (1): 81–89. doi:10.1016/j.apergo.2005.07.006.

Kling, R. 2007. "What Is Social Informatics and Why Does It Matter?" *The Information Society* 23 (4): 205–220. doi:10.1080/01972240701441556.

Koo, T. K., and M. Y. Li. 2016. "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research." *Journal of Chiropractic Medicine* 15 (2): 155–163.

Krenn, B. 2015. "Robot: Multiuse Tool and Ethical Agent." In *A Construction Manual for Robots' Ethical Systems: Requirements, Methods, Implementations*, edited by R. Trappl, 11–29. Springer International Publishing. doi:10.1007/978-3-319-21548-8_2.

Law, John. 2009. "Actor Network Theory and Material Semiotics." *Social Theory,* 141–158.

Leonardi, P. M. 2012. "Materiality, Sociomateriality, and Socio-Technical Systems: What do These Terms Mean? How are They Different? Do we Need Them." *Materiality and Organizing: Social Interaction in a Technological World,* 25–48.

Lowe, R., L. F. Chiu, and T. Oreszczyn. 2018. "Socio-technical Case Study Method in Building Performance Evaluation." *Building Research & Information* 46 (5): 469–484. doi:10.1080/09613218.2017.1361275.

Luhmann, N. 1996. *Social Systems* (J. B. Jr & D. Baecker, Trans.; 1st Edition). Stanford CA: Stanford University Press.

Mårell-Olsson, E., and I. Jahnke. 2019. "Wearable Technology in a Dentistry Study Program: Potential and Challenges of Smart Glasses for Learning at the Workplace." In *Perspectives on Wearable Enhanced Learning (WELL): Current Trends, Research, and Practice*, edited by I. Buchem, R. Klamma, and F. Wild, 433–451. Springer International Publishing. doi:10.1007/978-3-319-64301-4_20.

Merriam, S. B., and E. J. Tisdell. 2015. *Qualitative Research: A Guide to Design and Implementation* (4th ed.). *John Wiley & Sons*.

Moallemi, E. A., L. Aye, F. J. de Haan, and J. M. Webb. 2017. "A Dual Narrative-Modelling Approach for Evaluating Socio-Technical Transitions in Electricity Sectors." *Journal of Cleaner Production* 162: 1210–1224. doi:10.1016/j.jclepro.2017.06.118.

Muller, M. J., and A. McClard. 1995. "Validating an Extension to Participatory Heuristic Evaluation: Quality of Work and Quality of Work Life." *Conference Companion on Human Factors in Computing Systems*, 115–116. doi:10.1145/223355.223457.

Mumford, E. 1983. *Designing Human Systems for New Technology: The ETHICS Method*. Manchester: Manchester Business School.

Mumford, E. 2000. "A Socio-Technical Approach to Systems Design." *Requirements Engineering* 5 (2): 125–133. doi:10.1007/PL00010345.

Nelles, J., S. Kohns, J. Spies, C. Bröhl, C. Brandl, A. Mertens, and C. Schlick. 2017. *Best-Practice Approach for a Solution-Oriented Technology Assessment: Ethical, Legal, and Social Issues in the Context of Human-Robot Collaboration* (pp. 1–14). Springer. doi:10.1007/978-3-662-53305-5_1.

Nielsen, J. 1994a. "Heuristic Evaluation." In *Usability Inspection Methods*, edited by J. Nielsen and R. L. Mack, 25–62. John Wiley & Sons, Inc.

Nielsen, J. 1994b. "Enhancing the Explanatory Power of Usability Heuristics." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 152–158. doi.org/10.1145/191666.191729.

Nielsen, J., and R. Molich. 1990. "Heuristic Evaluation of User Interfaces." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 249–256. doi:10.1145/97243.97281.

Nierhoff, J., and T. Herrmann. 2017. "Data Elicitation for Continuous Awareness of Team Climate Characteristics to Improve Organizations' Creativity." *Proceedings of the 50th Hawaii International Conference on System Sciences.*

Noehring, F., R. Woestmann, T. Wienzek, and J. Deuse. 2019. "Socio-Technical Capability Assessment to Support Implementation of Cyber-Physical Production Systems in Line with People and Organization." In *Advances in Human Factors and Systems Interaction*, edited by I. L. Nunes, 299–311. Springer International Publishing. doi:10.1007/978-3-319-94334-3_30.

Nolte, A., I. Jahnke, I. A. Chounta, and T. Herrmann. 2018. "Supporting Collaboration in Small Volunteer Groups with Socio-Technical Heuristics." Proceedings of 16th European Conference on Computer-Supported Cooperative Work-Exploratory Papers. European Society for Socially Embedded Technologies (EUSSET).

Orlikowski, W. J. 2007. "Sociomaterial Practices: Exploring Technology at Work." *Organization Studies* 28 (9): 1435–1448.

Prilla, M., and T. Herrmann. 2017. "Challenges for Socio-Technical Design in Health Care: Lessons Learned From Designing Reflection Support." In *Designing Healthcare That Works. A Socio-Technical Approach*, edited by Ackerman, M. S., S. P. Goggins, T. Herrmann, M. Prilla, and C. Stary, 149–166. London: Academic Press (Elsevier).

Quiñones, D., C. Rusu, and V. Rusu. 2018. "A Methodology to Develop Usability/User Experience Heuristics." *Computer Standards & Interfaces* 59: 109–129.

Reeves, T. C., and J. G. Hedberg. 2003. *Interactive Learning Systems Evaluation*. Englewood Cliffs, NJ: Educational Technology Publications.

Reijers, H., and S. Limanmansar. 2005. "Best Practices in Business Process Redesign: An Overview and Qualitative Evaluation of Successful Redesign Heuristics." *Omega* 33 (4): 283–306. doi:10.1016/j.omega.2004.04.012.

Rost, M., and K. Bock. 2011. "Privacy By Design und die Neuen Schutzziele." *Datenschutz und Datensicherheit - DuD* 35 (1): 30–35. doi:10.1007/s11623-011-0009-y.

Schmidt, M., Y. Earnshaw, A. A. Tawfik, and I. Jahnke. 2020. "Methods of User Centered Design and Evaluation for Learning Designers." In *Learner and User Experience Research*. EdTech Books. https://edtechbooks.org/ux/ucd_methods_for_lx.

Schnuelle, C., J. Thoeming, T. Wassermann, P. Thier, A. von Gleich, and S. Goessling-Reisemann. 2019. "Socio-technical-economic Assessment of Power-to-X: Potentials and Limitations for an Integration Into the German Energy System." *Energy Research & Social Science* 51: 187–197. doi:10.1016/j.erss.2019.01.017.

Shin, D. 2014. "A Socio-Technical Framework for Internet-of-Things Design: A Human-Centered Design for the Internet of Things." *Telematics and Informatics* 31 (4): 519–531. doi:10.1016/j.tele.2014.02.003.

Shin, D. 2019. "A Living lab as Socio-Technical Ecosystem: Evaluating the Korean Living lab of Internet of Things." *Government Information Quarterly* 36 (2): 264–275. doi:10.1016/j.giq.2018.08.001.

Shin, D. 2020. "Expanding the Role of Trust in the Experience of Algorithmic Journalism: User Sensemaking of Algorithmic Heuristics in Korean Users." *Journalism Practice*, 1–24. doi:10.1080/17512786.2020.1841018.

Shin, D.-H., H. Choo, and K. Beom. 2011. "Socio-Technical Dynamics in the Development of Next Generation Mobile Network: Translation Beyond 3G." *Technological Forecasting and Social Change* 78 (3): 514–525. doi:10.1016/j.techfore.2010.04.017.

Shin, D., and M. Ibahrine. 2020. "The Socio-Technical Assemblages of Blockchain System: How Blockchains are Framed and how the Framing Reflects Societal Contexts." *Digital Policy, Regulation and Governance* 22 (3): 245–263. doi:10.1108/DPRG-11-2019-0095.

Shin, D.-H., and J. Jung. 2012. "Socio-Technical Analysis of Korea's Broadband Convergence Network: Big Plans, Big Projects, Big Prospects?" *Telecommunications Policy* 36 (7): 579–593. doi:10.1016/j.telpol.2012.03.003.

Shneiderman, B., and C. Plaisant. 2005. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. 4th ed. Boston: Pearson.

Tognazzini, B. 2014, March 6. First Principles of Interaction Design (Revised & Expanded). *AskTog*. https://asktog.com/atc/principles-of-interaction-design/.

Trist, Eric. 1950. "The Relations of Social and Technical Systems in Coal-Mining." *Paper presented to the British Psychological Society, Industrial Section*.

Trist, E., and H. Murray. 1993. "Introduction to Vol. II." In *The Social Engagement of Social Science. A Tavistock Anthology*, 36–60. University of Pensylvania Press. https://www.upenn.edu/pennpress/book/1928.html.

Vinck, D. 2019. "Maintenance and Repair Work." *Engineering Studies* 11 (2): 153–167. doi:10.1080/19378629.2019.1655566.

Wilson, C. 2013. *User Interface Inspection Methods: A User-Centered Design Method*. Waltham, MA: Morgan Kaufmann.

Wutzkowsky, J., and B. Böckmann. 2018. "Using MEESTAR to Identify Ethical and Social Issues Implementing a Digital Patient-Centered Care Platform." *EHealth*, 278–285. doi:10.3233/978-1-61499-858-7-278.

Appendix 1. Example of a socio-technical intertwinement: a predictive maintenance system (PM system)

A typical example of socio-technical intertwinement is the application of predictive maintenance (PM) in smart factories where algorithms provide warnings that a machine part may fail in the near future (Herrmann 2020). These warnings allow the staff to plan maintenance and to avoid stress caused by unforeseeable breakdowns. The plant operators (users) do not blindly follow the warning notification but do critically evaluate them after an on-site inspection (Herrmann 2020). In order to do so, the applied software would support *visibility* and that the reasons for its output are explainable.

Furthermore, the decision as to whether a notification has been taken seriously is also influenced by communication among the members of the staff, including communication among experienced colleagues, in specialised workshops, and in team meetings. This also applies to the question of how the maintenance work is scheduled and carried out. Plant operators must be prepared to validate PM messages, know whom to talk to, know how to propose improvements to the PM system, and so forth. For the evaluation, additional heuristics are needed to help evaluate this socio-technical predictive maintenance system. For example, heuristics to evaluate its quality may address questions regarding how to train or motivate people, how the communication between colleagues is organised and supported, or if the decision-making authority for modifying the rules of predictive maintenance is sufficiently clarified.