SUPPORTING USERS IN UNDERSTANDING INTELLIGENT EVERYDAY SYSTEMS

DISSERTATION

an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität München

vorgelegt von

MALIN EIBAND
M.Sc. Medieninformatik

München, den 8. Juli 2019

Erstgutachter:	Prof. Dr. Heinrich Hußmann
Zweitgutachter:	Prof. Dr. Gerhard Fischer
Drittgutachter:	Dr. Simone Stumpf

Tag der mündlichen Prüfung: 31. Oktober 2019

ABSTRACT

Intelligent systems have permeated many areas of daily life like communication, search, decision-making, and navigation, and thus present an important meeting point of people and artificial intelligence in practice. These *intelligent everyday systems* are in focus of this thesis.

Intelligent everyday systems exhibit the characteristics of so-called *complex systems* as defined in cognitive science: They serve ill-defined user goals, change dynamically over time, and comprise a large number of interrelated variables whose dependencies are not transparent to users. Due to this complexity, intelligent everyday systems can violate established usability guidelines of user interface design like transparency, controllability and easy error correction. This may introduce uncertainty to interaction that users have to overcome in order to reach a goal. I introduce a perspective from cognitive science, where users do so through *knowledge*. The work presented in this thesis aims at assisting users in gaining this knowledge, or *supporting users in understanding intelligent everyday systems*, for example, through explanation, control, correction or feedback. To this end, the work included in this thesis makes three main contributions:

First, I present *a method for eliciting user need for support and informing adequate solutions* through practical user problems with intelligent everyday systems in daily interaction. In a first phase, the presented method uses passive data collection to extract user problems with intelligent everyday systems through a combination of automated and manual analyses. In the second phase, these problems are then enriched and validated through active data collection to derive solutions for support. In addition, I report on the application of this method to uncover user problems with four popular commercial intelligent everyday systems (Facebook, Netflix, Google Maps and Google Assistant).

Second, I introduce a *conceptual framework* for categorising and differentiating prevailing notions in the field of how users should be supported in understanding intelligent systems related to *what users seek to know*, *how they acquire knowledge*, and *what kind of knowledge they acquire*. The presented framework can be used to make these notions explicit and thus introduces an overarching structure that abstracts from the field's fractured terminological landscape. It aims at helping other researchers become aware of existing approaches and locate and reflect on their own work.

Third, I present a number of *case studies and arguments* as an exploration of how users can be supported in the face of real-world challenges and trade-offs. My research reflects two possible perspectives to approach this question, a *normative* and a *pragmatic* one. As part of a critical reflection on the normative perspective, the work shows that explanations without information can similarly foster user trust in a system compared to real explanations, and discusses how user support can be exploited to deceive users. From the pragmatic perspective emerges a stage-based participatory design process that incorporates different stakeholder needs and a study assessing how support can be interwoven with users' primary tasks.

In summary, this thesis adopts a perspective on interaction with intelligent everyday systems, where *understanding* is a fundamental process towards reaching a user-set goal. On this basis, I introduce a research agenda for future work that incorporates the presented contributions and also includes challenges beyond the scope of this work, such as considering user empowerment. I hope that this agenda, along with the presented method, framework and design exploration, will help future work to shape interaction with intelligent everyday systems in a way that allows people to use them better, and to better ends and outcomes.

ZUSAMMENFASSUNG

Intelligente Systeme haben Einzug in viele Bereiche des täglichen Lebens wie Kommunikation, Informationssuche, Entscheidungsfindung, und Navigation erhalten und stellen damit einen wichtigen Berührungspunkt von Menschen und künstlicher Intelligenz in der Praxis dar. Solche *intelligenten Alltagssysteme* stehen im Fokus dieser Arbeit.

Intelligente Alltagssysteme weisen die Charakteristika von sogenannten *komplexen Systemen* aus der Kognitionsforschung auf: Sie dienen unscharfen Nutzerzielen, verändern sich dynamisch über die Zeit, und beinhalten eine große Anzahl an miteinander verknüpften Variablen, deren Wechselbeziehungen für Nutzer nicht erkennbar sind. Auf Grund dieser Komplexität können intelligente Alltagssysteme bewährte Richtlinien zur Gestaltung von nutzerfreundlichen Benutzeroberflächen verletzen, beispielsweise Transparenz, Kontrollierbarkeit, und einfache Fehlerbehebung. Dies kann bei der Interaktion zu Unsicherheit führen, die Nutzer auf dem Weg zu einem Ziel überwinden müssen. Ich führe eine Perspektive aus der Kognitionsforschung ein, nach welcher Nutzer dies durch *Wissen* tun.

Die hier präsentierten Arbeiten haben zum Ziel, Nutzern beim Erlangen dieses Wissens zu helfen, oder *Nutzerverständnis von intelligenten Alltagssystemen zu unterstützen*, beispielsweise durch Erklärung, Kontrolle, Korrektur oder Rückmeldung an das System. Hierzu leisten die vorgestellten Arbeiten hauptsächlich drei Beiträge:

Ich präsentiere zunächst *eine Methode, um das Nutzerbedürnis nach Unterstützung zu ermitteln und entsprechende Lösungen zu informieren*. Die Methode identifiziert dazu praktische Nutzerprobleme mit intelligenten Alltagssystemen im täglichen Gebrauch. In einer ersten Phase werden diese Probleme auf Grund von passiver Datenerhebung unter Verwendung automatisierter und manueller Analysemethoden extrahiert. In der zweiten Phase werden die ermittelten Probleme durch aktive Datenerhebung angereichert und validiert, um Lösungen zur Unterstützung abzuleiten. Daneben berichte ich von der Anwendung dieser Methode, um Nutzerprobleme in vier verbreiteten kommerziellen intelligenten Alltagssystemen (Facebook, Netflix, Google Maps und Google Assistant) aufzudecken.

Danach führe ich ein *konzeptuelles Framework* ein, mit dem im Feld vorherrschende Annahmen, wie Nutzerverständnis von intelligenten Alltagssystemen unterstützt werden sollte, klassifiziert und differenziert werden können. Diese Annahmen beziehen sich darauf, welches Wissen Nutzer erlangen *wollen*, *wie* sie dieses Wissen erlangen, und um *welche Art* von Wissen es sich handelt. Durch das Framework können die jeweiligen Annahmen explizit gemacht werden. Es schafft so eine übergreifende Struktur, die von der Fülle und Diversität der im Feld verwendeten Begrifflichkeiten abstrahiert. Das Framework kann anderen Forschern dabei helfen, sich über bestehende Ansätze bewusst zu werden, und ihre eigene Arbeit zu verorten und zu reflektieren.

Zum Dritten bringe ich eine Reihe von *Fallbeispielen und Argumenten* an, die explorieren, wie Nutzer angesichts von Einschränkungen und Abwägungen in der Praxis unterstützt werden können. Meine Forschung spiegelt dabei zwei mögliche Sichtweisen auf diese Frage wider, eine *normative* und eine *pragmatische*. Im Zuge einer kritischen Betrachtung der normativen Sichtweise zeigt diese Arbeit, dass Erklärungen ohne Informationsgehalt in ähnlicher Weise Vertrauen in ein System hervorrufen können wie richtige Erklärungen. In diesem Zusammenhang wird weiterhin diskutiert, wie Unterstützung gezielt zur Täuschung von Nutzern missbraucht werden kann. Aus der pragmatischen Sichtweise geht in dieser Arbeit ein stufenförmiger partizipatorischer Designprozess hervor, der die verschie-

denen Interessen in der Praxis Beteiligter berücksichtigt. Zudem wird in einer Studie untersucht, wie Unterstützung von Verständnis mit der Primäraufgabe von Nutzern verknüpft werden kann.

Zusammenfassend nimmt diese Arbeit eine Perspektive auf Interaktion mit intelligenten Alltagssystemen ein, die *Verstehen* als grundlegenden Prozess auf dem Weg zu einem Nutzerziel begreift. Basierend darauf stelle ich eine Forschungsagenda vor, die die präsentierten Publikationen einschließt und zudem Herausforderungen über den Rahmen dieser Arbeit hinaus beinhaltet, wie beispielsweise die Einbeziehung von "Nutzer-Empowerment". Ich hoffe, dass diese Agenda, die vorgestellte Methode, das Framework und die Erkenntnisse aus der Exploration möglicher Designansätze zukünftiger Forschung hilft, Interaktion mit intelligenten Systemen im Alltag zu gestalten – so, dass Nutzer sie besser und zu besseren Zwecken verwenden können.

Acknowledgements

During my time as a PhD student, I had the honour and joy of meeting and working with people dedicated to research and friendship, and without whom this thesis would not exist.

First and foremost, thank you, **Heinrich**, for your support over the last four years, your everlasting loyalty and alliance to your PhD students, and for having a way with bringing people to the department that make it a fantastic place to work at. I also want to thank you, **Simone** and **Gerhard**, for reviewing this thesis and for being the most supportive external examiners a PhD student could wish for. Thank you, **Emanuel**, for persuading me to do a PhD and bringing me here in the first place, and a million thanks to all those who provided feedback on different drafts of this thesis, even last minute: Ceenu, Christian, Daniel (B.), Daniel (R.), Fabius, Robert, and Sarah (V.).

Furthermore, I would like to thank those amazing and crazy people without whom I would probably not have lasted long: Ceenu, Daniel (B.), Hanna, Nadja, Renate, and Sarah (V.) Thank you for sharing everyday PhD life with its bright and dark moments, for our (sometimes) excessive teatimes, for all the fun and laughter, for your support when I had one of my spleens before deadlines, for working and travelling together. You have truly made these last four years special and unforgettable to me. Thank you, **Ceenu**, for harbouring me, for thinking deeply about life and its challenges together, and for your honest nature. Thank you, **Daniel** (B.), for all the nerdy humour, for being a great roomie, and for the fruitful research collaborations (in the course of which we decorated our office with post-its). Sorry for dumping this conceptual framework one week before the deadline. Thank you, **Hanna**, for sharing our thoughts on so many topics, for bearing with me in good and bad times, for introducing me to yoga, and for being an amazing research partner. Thank you, **Nadja**, for the positive atmosphere you create, for caring deeply for those around you, and for being as much of a foodie as I am. Thank you, **Renate**, for taking care of me when I started my PhD, for so many deep talks and cups of tea, and for being an inspiring person. Thank you, **Sarah** (V.), for bearing with me and making the impossible possible together, for being the most organised person I know, and for your truly sympathetic nature.

I also would like to thank all those who have enriched my PhD life in so many ways: Thank you, **Basti**, for teaching me to ski (at least it was worth the try). Thank you, **Beat**, for inspiring discussions about artificial intelligence and for the artsy glow you radiate. Thank you, Best-Paper-**Christian**, for regularly checking on me (having a seemingly built-in sensor for when I was procrastinating). Thank you, **Daniel** (R.), for sharing my love for so many things (from Japanese food to dancing) and for being my gipsy buddy. Thank you, **Fabius**, for living through the ups and downs of (research) life and travelling together along the way. Thank you, **Florian** (Lachner), for your laughter and memorable parties. Thank you, **Kai**, for having my back and answering a lot of student tutor emails in the last four months. Thank you, **Mohamed/Khamis/Mo**, for teaching me Arabic (*el naharda el khamis* will be forever stuck in my head!), for surviving our first joint CHI project, and for all the fun when when we were sharing an office.

Moreover, I want to thank the most amazing team, **my current and former colleagues**, for travelling and working together, and for making our department a lively, supporting place: Alina, Axel, Changkun, Christina, David, Dennis, Fiona, Florian (B.), Florian (Lang), Gesa, Heiko, Henri, Jakob, Jingyi, Julie, Lewis, Linda, Lukas, Mariam, Matthias (H.), Matthias (S.), Michael (B.), Michael (C.), Pascal, Sarah (A. B.), Sarah (P.), Sarah (T.), Simon, Sylvia, Thomas, Tobi, Tonja, Ville, and Yoma.

A big thank you to the people without whom we would be (and often were) lost – **Rainer**, **Franziska**, and **Anja** – and our great navigators through the deep waters of our PhDs, **Albrecht**, **Andreas**, **Florian** (A.), and again, **Heinrich**.

Thank you to **my excellent students**, it was a pleasure working with you: Alex, Charlotte, David, Diana, Julian, Mareike, Martin, Michael, Philipp, Sandra, Sophia, Tanja, and Tim.

Finally, I want to thank **my family** for your patience over the last months when I didn't call in weeks, for turning up when I hadn't left the house for quite some time, and for all the love and encouragement.

And I deeply thank you, **Robert**, for all the songs and laughter, for the shared adventures, for your endless support, and for filling my heart with joy.

Table of Contents

1 Introduction 1

- 1.1 Thesis Statement . . . 1
- 1.2 Background and Definitions . . . 3
- 1.3 Summary and Overview of the Thesis . . . 7

2 Research Problems and Questions 9

- 2.1 Empirical Research: Need for User Support . . . 9
- 2.2 Conceptual Research: Conceptualisation of User Support . . . 11
- 2.3 Constructive Research: Real-world Applicability of User Support . . . 12

3 Contribution 13

- 3.1 Contributing Publications . . . 14
- 3.2 Empirical Research: A Method & Analysis to Elicit User Need for Support . . . 15
  - 3.2.1 A Methodological Exploration & Method to Elicit User Need for Support . . . 15
  - 3.2.2 An Analysis of User Need for Support . . . 18
- 3.3 Conceptual Research: A User-Centric Framework of User Support . . . 20
- 3.4 Constructive Research: A Normative & Pragmatic Design Exploration of User Support 22
  - 3.4.1 Normative: Placebo Effects & Dark Patterns of User Support . . . 23
  - 3.4.2 Pragmatic: A Participatory Design Process & Implicit User Support . . . 24
- 3.5 Research Methods and Limitations . . . 25

4 Discussion and Future Work 29

- 4.1 Supporting Users in Understanding Intelligent Everyday Systems: A Research Agenda 29
  - 4.1.1 Eliciting User Need for Support . . . 30
  - 4.1.2 Establishing Conceptual & Terminological Clarity for User Support . . . 31
  - 4.1.3 Considering Limitations of Understanding . . . 32
  - 4.1.4 Applying User Support in Practice . . . 33
  - 4.1.5 Investigating Effects of User Support . . . 35
- 4.2 Concluding Remarks . . . 38

List of Figures 39

List of Tables 41

Glossary 43

References 47

1 Introduction

1.1 Thesis Statement

Intelligent systems have found their way into almost every area of daily life: They help users in various application domains to exercise[1], to work more efficiently when answering their emails[2], to navigate to places and arrive there in time[3], to find relevant information[4] and items to their taste[5], to get updates from their friends[6], or to organise their schedule[7]. Many of these systems are used by more than a billion people every day [42] and thus form the space where humans and artificial intelligence meet *in practice* today.

All of these examples are what I introduce as *intelligent everyday systems*, which are in focus of this thesis. This term includes two properties, *everyday* and *intelligent*.

First, I define *everyday* systems as *systems embedded in a layperson's habitus and mediating everyday tasks and practices*. This delineates intelligent everyday systems from other intelligent systems that incorporate specific domain knowledge and/or involve high risk, such as intelligent systems for medical diagnosis.

Second, I draw on work from cognitive science to define *intelligent* systems as *complex systems* [25, 102]: They comprise a large number of interrelated variables whose dependencies are intransparent and dynamically change over time, dependent on or independent from user interaction. For example, as introduced later, intelligent fitness coaching generates personalised training plans based on data such as users' profiles and past training performance, other users' data, and inferred fitness level, but how or why a specific training plan came into being is hidden from users (see [P6]). Moreover, complex systems often involve ill-defined user goals. In intelligent fitness coaching, this goal might be *to build muscles*, or *to lose weight* – many people have experienced that such a goal is fuzzy and difficult to break down into clear steps.

Due to this complexity, intelligent everyday systems may violate established usability guidelines of user interface design like transparency, consistency, controllability and easy error correction [3, 20, 47]. From a user perspective, this may introduce "considerable uncertainty" [97] about getting from a current state to a goal state when interacting with an intelligent everyday system, for example, through faulty predictions and inferences users cannot readily correct [7].

Helping users to reach their goal is the overarching motivation of this thesis. In particular, I introduce a perspective from cognitive science, where the "barriers" between a current state and a goal state are

[1] https://www.freeletics.com/en/, accessed 23 August 2019.

[2] https://www.google.com/gmail/about/, accessed 23 August 2019.

[3] https://www.google.de/maps/, accessed 23 August 2019.

[4] https://www.google.com/search/about/, accessed 23 August 2019.

[5] https://www.netflix.com/de-en/, accessed 23 August 2019.

[6] https://www.facebook.com/, accessed 23 August 2019.

[7] https://assistant.google.com/, accessed 23 August 2019.

considered lack of *knowledge* [25]. Accordingly, I define *understanding* in terms of the *knowledge required to overcome these barriers*. Assisting users in gaining this knowledge, or *supporting users in understanding intelligent everyday systems*, for example, through explanation, control, correction or feedback, is the main objective of my work. Thus, I follow a pragmatic perspective on user interaction with intelligent everyday systems that strives to *helping users to use a system better and in more informed ways, and to better ends and outcomes* [P3].

To this aim, this thesis addresses empirical, conceptual, and constructive research, and makes three main contributions to the field:

1) *Need for user support* – I present *a method for eliciting user need for support and informing adequate solutions* through practical user problems with intelligent everyday systems in daily interaction. Despite the great research interest that intelligent systems have sparked in Human-Computer Interaction (HCI) and related fields, there is yet little empirical research on user need for support in real-world use. In a first phase, the presented method uses passive data collection to extract user problems with intelligent everyday systems through a combination of automated and manual analyses. In the second phase, these problems are then enriched and validated through active data collection to inform solutions for support. In addition, I report on the application of this method to uncover user problems with four popular commercial intelligent everyday systems (Facebook, Netflix, Google Maps and Google Assistant).

2) *Conceptualisation of user support* – Second, I introduce a *conceptual framework* for categorising and differentiating prevailing notions in the field of how users should be supported in understanding intelligent systems. While prior work on system *transparency*, *intelligibility*, *scrutability*, and so on, is linked by this shared goal, a structured literature review revealed divergent implicit assumptions in terms of *what users seek to know*, *how they acquire knowledge*, and *what kind of knowledge they acquire*. The presented framework can be used to make these assumptions explicit and thus introduces an overarching structure that abstracts from the field's fractured terminological landscape. It aims at helping other researchers become aware of existing notions on how to support users in understanding intelligent systems and locate and reflect on their own work.

3) *Real-world applicability of user support* – I present a number of *case studies and arguments* as an exploration of how users can be supported in the face of real-world challenges and trade-offs. These papers reflect two potential perspectives to approach this question, a *normative* and a *pragmatic* one. As part of a critical reflection on the normative perspective, the work shows that explanations without information can similarly foster user trust in a system compared to real explanations, and discusses how means for user support can be exploited to deceive users. From the pragmatic perspective emerges a stage-based participatory design process that incorporates different stakeholder needs and a study assessing how support can be interwoven with users' primary tasks.

In summary, this thesis adopts a perspective on interaction with intelligent everyday systems, where *understanding* is a fundamental process towards reaching a user-set goal. On this basis, I introduce a research agenda for future work that incorporates the presented contributions and also includes challenges beyond the scope of this work, such as considering user empowerment. I hope that this agenda, along with the presented method, framework and design exploration, will help future work to shape interaction with intelligent everyday systems in a way that allows people to use them better, and to better ends and outcomes.

1.2 Background and Definitions

This section presents my definition of terms relevant to the objective of this thesis and their relation to one another. Figure 1.1 provides a graphical summary of these foundations. In particular, I transfer *complex problem solving* as defined in cognitive science [25, 85, 102] as a theoretical grounding of my work, which I will explain in detail below. This transfer introduces a user-centric perspective on interaction with intelligent systems which abstracts from the technical details of a system.

Users In the context of my work, the term *users* describes *end-users*, more precisely, *laypersons* who are familiar with using technology in the form of commercial products and services, but not necessarily with technical details. Moreover, in contrast to experts [3], they do not operate within a particular professional domain.

Intelligent Systems From a technical perspective, intelligent systems are often defined in terms of their algorithmic reasoning abilities or their behaviour [96]. However, this thesis focuses on the *user* side of interaction with intelligent systems. For my work, I therefore define intelligent systems with respect to the system *characteristics* and *how these characteristics relate to the user and surface in interaction*. In particular, I draw on work from cognitive science [25, 85, 102] to define intelligent systems as *complex systems*. Complex systems exhibit five distinct characteristics [33]. In the following, I present these characteristics one by one and illustrate them with the intelligent fitness application investigated in my work [P6][8].

1. *Complexity*: Complex systems comprise a *large number of variables* (input, output and hidden variables not represented in the interface). For example, the fitness application generates personalised training plans based on, among others, each individual user's profile (e.g., height, weight, BMI), past training performance, performance level, preferences (e.g., number of training days per week), goal (e.g., gain muscle or lose weight). A user's height and weight are examples for input variables, the BMI is a hidden variable calculated from the other two, and the training plan is an output variable.

2. *Interconnectedness*: The variables of the system are *interrelated* in a way that they impact each other. For example, a user's performance level is derived from past training performances.

3. *Opacity*: The interconnectedness of the system variables is *not transparent*, so that it is often not possible to see if and how a variable impacts another. Likewise, in the fitness application, users receive the system output, that is, their training plan, but the data used and how the system takes this data into account for generating a plan is hidden.

4. *Dynamics*: The system state *changes dynamically* over time, dependent on or independent from an action taken. For example, the training plans in the fitness application might change based

[8] I refer to the mid 2017 version of the application.

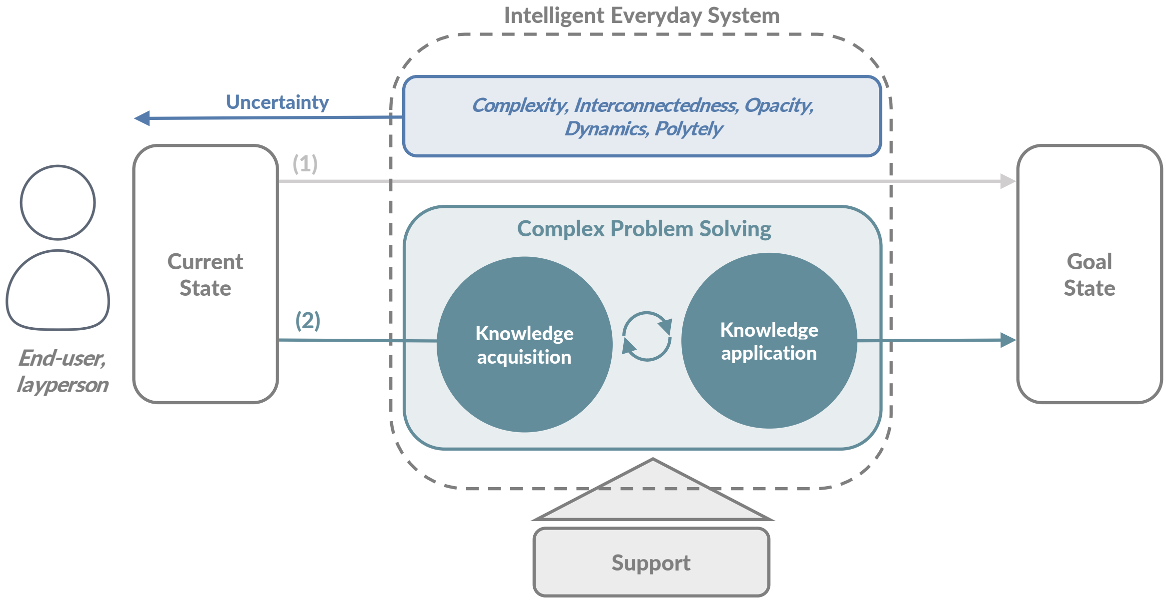

Figure 1.1: Summary of the theoretical foundations of my work. I define interaction with intelligent everyday systems as *complex problem solving* in situations where users do not know how to reach their goal (2) (in contrast to (1), where the goal can be reached). Users then engage in *knowledge acquisition* and *knowledge application* – supporting this process is the overarching objective of this thesis.

on user feedback (e.g., that a particular exercise has to be avoided due to an injury), or relative to other users' performance with the same fitness level.

5. *Polytely*[9]: Complex systems often serve *multiple, ill-defined goals* that might possibly conflict. For example, the goals *losing weight* and *building muscle* are fuzzy, difficult to break down into clear steps, and may partially conflict.

This illustration shows that the characteristics of complex systems can be readily transferred to intelligent systems: Intelligent systems access and rely on large datasets, and the factors impacting system calculations may be manifold and interconnected [9], but they are mostly hidden from the interface. Moreover, system states are in constant flux and evolve over time. This might happen either explicitly through user input, or implicitly and thus independent from interaction. In the literature on intelligent systems, this distinction is known as *adaptability* and *adaptivity* [27]. Finally, user goals in intelligent systems are often ill-defined and incompletely specified [18]. In summary, when I relate to the complexity of intelligent systems, I refer to these five characteristics.

Intelligent Everyday Systems I define intelligent everyday systems as *intelligent systems embedded in a user's habitus that mediate everyday tasks and practices*. Everyday tasks may either be completely (e.g., routing in navigation) or in parts (e.g., email response suggestions) delegated to the system. Applications investigated in this thesis include health and well-being (intelligent fitness coaching), navigation (Google Maps), entertainment (Netflix), and voice assistance (Google

[9] From Greek *poly-* and *-tel-* meaning "many goals".

Assistant) [P6, P8, P9]. As a subgroup of intelligent systems, intelligent everyday systems share the characteristics of complex systems.

Interaction with Intelligent Everyday Systems In this thesis, I assume that the purpose of any interaction with an intelligent everyday systems is to *reach a user-set goal*. Goals may consist of subgoals and have different levels of abstraction, such as writing an email, getting updates from friends, or getting fit. In particular, my work assesses situations in which a user *has a goal, but does not know how to reach it* when interacting with an intelligent everyday systems. Such situations are referred to as *problems* in cognitive science [21]. The process of finding a way to get from a current state to a goal state is called *problem solving* [74]. Problem solving is considered to be *complex* if it concerns the "goal-oriented control of [...] complex systems" [25]. This complexity then manifests itself in *uncertainty* from the perspective of someone confronted with the system [85] – the *user* in the context of my work.

Interpreting intelligent systems as *complex systems* thus allows to frame interaction with intelligent systems – and consequently also with intelligent everyday systems – as *complex problem solving* in situations in which users do not know how to get from a current state to a desired goal state. This framing permits to define and formalise how users approach a system in these situations, which I will explain next.

Understanding The "barriers" between a current state and a goal state are considered "lack of knowledge" in cognitive science [25]. The process of getting to a goal state (i.e., solving a problem) therefore typically consists of two phases: *knowledge acquisition* and *knowledge application* [25]. Importantly, this process is descriptive (how people *do* behave), not normative (how they *should* behave). Moreover, it is influenced by people's motivations and needs [39] as well as cognitive factors such as domain knowledge [109].

Transferred to the context of my work, this means that users pass through these two phases when interacting with an intelligent everyday systems and do not know how to reach a goal, as shown in Figure 1.1.

1. *Knowledge acquisition*: In the first phase, people acquire explicit and implicit (also known as tacit) knowledge [88] about the problem at hand. In analogy to HCI [83], *mental models* serve as representations of the knowledge a person possesses about the problem, and also include prior knowledge and experiences [25]. Transferred to the context of interaction with intelligent everyday systems, users explore the system and update their mental models according to the acquired knowledge.

2. *Knowledge application*: In the second phase, the acquired knowledge is applied while the consequences of the actions taken are being monitored. If it turns out that these actions are invalid to reach the goal, people switch back to the first phase or change the goal, depending on, for example, the effort involved and the importance of the goal [25]. In analogy, in interaction with intelligent everyday systems, users apply the knowledge they have acquired through system exploration and observe the resulting effects to determine whether their goal has been reached.

The knowledge users acquire in this process is what I define as *understanding* in this thesis.

The literature on complex problem solving distinguishes between three types of knowledge: *input-output* knowledge, *structural* knowledge and *strategic* knowledge [100]. Input-output knowledge represents knowledge about how an input and system output are related. Structural knowledge is knowledge about the causal relations between the system variables. Strategic knowledge is knowledge about how to proceed (i.e., which sequence of steps to take) to reach a goal. Transferred to the context of intelligent everyday systems, users might acquire knowledge about how an input is related to an output, which processes have led to an output, or how the system output can be controlled.

Support Based on these foundations, I define support as *assisting users in gaining the knowledge required to reach their goal*. This reflects a pragmatic perspective [P7] I adopt throughout my work which seeks to *help people to use a system better and in more informed ways* [P3], as explained later in more detail in Section 3.4.

I assume that support is based on the *transfer of information* between system and user [48]. Approaches to the nature of this transfer in the literature are highly diverse and are therefore subject to the research presented in this thesis. In a larger perspective, they can be broadly classified based on the *direction* of information transfer. From *system to user*, most approaches either target *explanation* or *experimentation* (e.g., [13, 44, 59, 62, 63, 69, 78, 91, 92])). From *user to system*, approaches aim at *control*, *correction* and *feedback* (e.g., [57, 58, 59]).

Finding a way to categorise and distinguish different notions of user support is part of my work and will be explained in more detail in Sections 2.2 and 3.3.

Relation to other Concepts and Terms Here, I clarify the relation of these theoretical foundations to other concepts and terms in the field. In particular, reoccurring discussions over the course of the presented research have surfaced two aspects that I will elaborate on below: First, the relation of intelligent everyday systems to *other* common groups of intelligent systems in the literature, and second, the relation of the complex problem solving perspective to Norman's *seven stages of action* [82].

Intelligent Everyday Systems and other Groups of Intelligent Systems There are many classifications and definitions of intelligent systems in the literature to delineate different types of systems. Most prominently, research classifies intelligent systems as *interactive machine learning* systems (e.g., [3]), *context-aware* systems (e.g., [58, 69]), or *interactive intelligent systems* (e.g., [8]). All of these categorisations highlight different aspects of the system and their relation to the user. My definition includes the *embedding in a user's habitus*. In this sense, intelligent everyday systems could be seen to make use of the *context* users operate in. For example, they may integrate contextual information such as user preferences, location, browsing history, style of writing, but also information from other sources, such as data from other users. Consequently, intelligent everyday systems as defined in this thesis can be classified in a broader sense as *context-aware systems* [12].

Complex Problem Solving and Norman's Seven Stages of Action Framing interaction with intelligent systems as a problem solving process is *user-centric*, which allows to abstract from the technical details of an intelligent system as well as the plethora of definitions, concepts and terms one finds in the literature. Instead, it focuses on what *any* intelligent system looks like from a user's perspective. Problem solving models how humans in general approach situations in which they do not see an obvious way to reach a particular goal – they react by acquiring and applying knowledge. This knowledge can be both explicit and implicit and thus also includes experiences ("X has always worked") that may not be verbalisable.

As such, it carries similarities to Norman's seven stages of action [82]. Getting to a certain goal state is central to both models. Norman's *gulf of execution* and *gulf of evaluation* similarly describe a lack of knowledge on the user side and thus potential points of action for user support. Moreover, both models draw on the construct of *user mental models* as a representation of the knowledge a person possesses about the system and interaction process. Also, they share a pragmatic perspective on user support: Knowledge about a system helps users to reach a goal, but is not an end in itself.

However, they differ in what they assume as the obstacle between a current state and a goal state. While Norman's work roots his gulfs in (missing) *affordances or system representations*, the problem solving perspective explains uncertainty in interaction through the *characteristics* of the system. Sure enough, the touch point between user and system is always the interface. However, I argue that in interaction with intelligent systems, there are cases in which affordances or system representations alone cannot account for users not being able to reach a goal. For example, in the case of the intelligent fitness application presented later [P6], interviews revealed that users sometimes completed the training schedule for a whole week in one single day. This was not due to usability problems with the app (e.g., that they did not know how and where to generate their training schedule), but rather because they did not understand why they were given a seemingly relaxed workout. The reasons can be found in the *underlying* system complexity, not in a poorly designed interface as such.

The problem solving perspective thus introduces complexity as an important part of interaction with intelligent systems and helps to explain its consequences from a user point of view.

1.3 Summary and Overview of the Thesis

In this introduction, I have clarified the overarching objective of this thesis: Assisting users in gaining the knowledge necessary to reach their goal when interacting with intelligent everyday systems. Moreover, I have introduced *complex problem solving* as a theoretical foundation to ground the presented work.

Chapter 2 introduces the research problems addressed in this thesis as well as the guiding research questions that emerge from these problems. Chapter 3 presents my respective contribution to each guiding research question. Finally, I discuss and reflect on my contribution in Chapter 4 as part of a research agenda for future work on supporting users in understanding intelligent everyday systems.

2

Research Problems and Questions

Interaction with intelligent systems has sparked great interest within the HCI community and related research areas in the last years. In this chapter, I reveal open research problems in the field and derive the research questions guiding my work. These questions are on an abstract level since they span more detailed research questions that can be found in the respective publications.

I draw on Laudan's taxonomy [65] of scientific research adapted for HCI by Oulasvirta and Hornbæk [86] as an overarching structure. The authors distinguish between three types of research problems towards scientific progress: *empirical*, *conceptual*, and *constructive*.

Empirical research is "aimed at creating or elaborating descriptions of real-world phenomena related to human use of computing". It consists of three subtypes: *unknown phenomena*, *unknown factors*, and *unknown effects*. The first part of my work copes with such *unknown phenomena and factors* by exploring actual user need for being supported in understanding intelligent everyday systems. I do so by capturing practical user problems with intelligent everyday systems in daily interaction as reported by users.

Conceptual research is "aimed at explaining previously unconnected phenomena occurring in interaction". This research problem type, too, has three subtypes: *implausibility*, *inconsistency*, and *incompatibility*. The second part of my work aims to resolve *conceptual inconsistency* in prior work as to how users should be supported in understanding intelligent systems.

Constructive research is "aimed at producing understanding about the construction of an interactive artefact for some purpose in human use of computing". Again, the authors distinguish between three subtypes: *no known solution*, *partial, ineffective, or inefficient solution*, and *insufficient knowledge or resources for implementation or deployment*. The third part of my work targets the first two subtypes by investigating the applicability of supporting users in understanding intelligent everyday systems given real-world constraints.

As an overview and for quicker orientation, Table 2.1 lists the identified research problems and related guiding research questions.

2.1 Empirical Research: Need for User Support

The violation of established usability principles [80, 82] through intelligent systems is an often mentioned concern in the literature (e.g., [3, 4, 20, 47, 46]). Yet, it is not clear how this impacts interaction with intelligent everyday systems in *practice*, that is, *when* users actually need to be supported in understanding intelligent everyday systems to reach a goal. This carries the risk of developing concepts and solutions that are decoupled from real-world use.

On the one hand, the majority of prior work is based on laboratory or online settings and prototype solutions (e.g., [13, 44, 46, 49, 57, 58, 59, 69, 91, 92, 104]). For example, Kulesza et al. [59] assessed explanations in a text classification system, Koch et al. [57] looked at system-supported design ideation, and Kocielnik et al. [58] investigated how explanation interfaces impact user expectations

Research Problem [86]	Short Description	Research Question
Empirical	No established methods to elicit user need for support in daily use of intelligent everyday systems.	**RQ1a:** When do users need to be supported in understanding intelligent everyday systems in daily interaction?
	Empirical evidence of user need for support is sparse.	**RQ1b:** What are suitable methods to capture user need for being supported in understanding intelligent everyday systems in daily interaction?
Conceptual	Lack of conceptual clarity and a connection of prior approaches concerning different notions of user support.	**RQ2:** What are existing notions of user support for understanding intelligent systems in HCI and related fields, and how can they be conceptualised and distinguished?
Constructive	The complexity of real-world scenarios is not reflected in prior work, best practices for applying user support are missing.	**RQ3:** How can user support for understanding intelligent everyday systems be designed given real-world constraints and challenges?

Table 2.1: Identified research problems and related guiding research questions of this thesis.

in an intelligent scheduling assistant. While these studies have greatly advanced the field, they necessarily abstract from real-world system use so that it remains unclear to what extent the results can be transferred to daily interaction with intelligent everyday systems. Furthermore, work on actually *deployed* intelligent everyday systems has mostly focused on users' beliefs about and perceptions of the system (e.g., Facebook [7, 23, 24, 93] or AirBnB [50]). To the best of my knowledge, the only study investigating user need for support along the lines of this thesis has been presented by Bunt et al. [8] who analysed users' desire for explanation in a variety of intelligent everyday systems including Youtube, Facebook, and systems for text input. However, their study focused on need for explanation only and thus did not assess other approaches like experimentation or feedback, which motivates the need for further investigation.

Hence, the first guiding research question of my work is –

> ***RQ1a**: When do users need to be supported in understanding intelligent everyday systems in daily interaction?*

On the other hand, capturing user need for being supported when interacting with intelligent everyday systems is challenging from a methodological point of view. For example, prior work has observed that users lack *algorithmic awareness* [24], that is, awareness *that* they are interacting with an intelligent systems. Moreover, terminology like "algorithm" is often not in a layperson's vocabulary and influenced by cultural background [P11]. As a result, as Bucher [7] puts it, "accessing people's personal stories and experiences with data and algorithms can be tricky. Where do you go to gather stories about things algorithmic? [...]". Methods employed in the field are diverse and include logging (e.g., [59]), questionnaires and surveys (e.g., [24, 59, 92, 93, 104]), think-aloud tasks (e.g., [63]), diary studies (e.g., [8]), and interviews (e.g., [8, 23, 24, 70, 104]). However, there are currently no established methods for investigating user need for support with intelligent everyday systems in daily interaction.

This motivates the next guiding research questions of my work –

> ***RQ1b**: What are suitable methods to capture user need for being supported in understanding intelligent everyday systems in daily interaction?*

In summary, the first objective of this thesis is a large-scale assessment of user need for support and an exploration of suitable methods to do so.

2.2 Conceptual Research: Conceptualisation of User Support

As intelligent systems have penetrated everyday contexts, their complex nature, in particular their *opacity*, have given rise to increasing concern in research, industry, politics, and the general public. This has sparked a plethora of work in HCI and adjoining research areas. However, terminology in the field is currently highly divergent: Prior work aims at making intelligent systems *transparent* [92], *scrutable* [55], *intelligible* [68], *explainable* [35], *interpretable* [94], *accountable* [15], and so on. A recent HCI survey by Abdul et al. [1] shows the fractured landscape around many such terms. While all of them carry a notion of how users should best be supported in understanding intelligent systems, this notion often remains unarticulated in the presented prototypes or concepts.

For instance, showing users information about *how well* a system knows a certain domain [97] implies a different notion about users and user support than telling them *that* a system output had been derived by an intelligent system [92], which again is different from providing explanations about *why* a system output came into being [70]. The first example targets a rather abstract level of information, the second informs about the presence of an intelligent systems, the last one refers to a concrete outcome. Moreover, all these examples imply to *present information* to users. Other approaches suggest to tell them how to correct the system [63], or even to adopt a mixed-initiative approach [3].

As a result, the field still lacks conceptual clarity and a shared terminology. Work on interpretability has recently been criticised for unclear use of the term [18, 71], a survey on explainability in recommender systems found incompatible existing taxonomies [84], and discussions about system transparency and accountability revealed diverging assumptions (i.e., disclosing source code versus system auditing through experts) [22]. This impedes awareness of existing work, a structured development and discussion of new ideas, and iterative learning from prior research. Yet, conceptual clarity and the connection of diverse existing approaches is crucial to advance scholarship, as pointed out in a recent "roadmap" towards a rigorous science of interpretability [18].

Hence, the second objective of my work is to add structure and conceptual clarity to existing and future work in the field by addressing the guiding research question –

> ***RQ2:** What are existing notions of user support for understanding intelligent systems in HCI and related fields, and how can they be conceptualised and distinguished?*

2.3 Constructive Research: Real-world Applicability of User Support

Supporting users in understanding intelligent systems is challenging in real-world scenarios. Practitioners face constraints that differ from the mostly experimental settings employed by research in the field. For example, real-world design scenarios have to meet multiple stakeholders' needs, like those of users, designers, and developers [P6]. Moreover, companies may keep intentional secrecy to protect their intellectual property [9]. Also, the complexity of intelligent systems may only be made understandable to users to a certain extent [9]. In addition, from a design perspective, screen space is often limited and user interface interventions have to be in line with the overall user experience and corporate identity of a product [P6]. Finally, real-world settings come with a trade-off between costs and benefits of user support, for example, when accessing explanations in the interface interferes with the primary task of users [8]. These challenges and constraints have become more urgent in the face of recent legislation: The General Data Protection Regulation was enforced on 25 May 2018 in the European Union and provides users with what has been called a "right to explanation" [89] as well as a right to opt-out of algorithmic decision-making altogether [90].

However, best practices for applying user support in real-world scenarios are still missing to date. Most research settings neglect these practical constraints as the field is still evolving. Moreover, many existing solutions and guidelines are difficult to transfer to a concrete real-world scenario, since they remain either on an abstract level (e.g., in the form of "solution principles" [20]) or are presented in the form of very specific prototypes (e.g., [58, 59, 69]). Also, there is no agreement on crucial aspects of user support, such as the level of detail and abstraction of an explanation (e.g., completeness of information [59] versus hiding unnecessary details [19]), or what constitutes a "good" explanation in general [84] (see Section 2.2).

As Amershi et al. [4] have noted in a recent paper, potential design guidance remains often tacit in the literature. Their work surfaces prior insights and distils them into "generally applicable design guidelines for human-AI interaction". However, even though the presented guidelines were evaluated against existing products, it remains unclear how practitioners would apply them to find suitable approaches for their own systems.

Since HCI strives to create technology that benefits people, it is important to bridge this gap between research and practice. The last part of this thesis therefore explores how user support can be applied in real-world scenarios given the mentioned constraints and challenges. Thus, I derive the last guiding research question of my work –

> ***RQ3:*** *How can user support for understanding intelligent everyday systems be designed given real-world constraints and challenges?*

3
Contribution

This chapter summarises the primary contributions of this thesis. In order to locate them in the greater view on HCI as a research field, I draw on the classification of HCI knowledge types by Wobbrock and Kietz [111] which orthogonally complement the three research problems introduced earlier.

Table 3.1 provides a summary of my contributions, the type of HCI knowledge they represent and the guiding research questions they address.

Since all contributions resulted from joint work with others, I will use the scientific “we” when presenting them.

Research Question	Knowledge Type [111]	Primary Contribution	Contributing Paper
RQ1a	Methodological	An exploration of established *active* data collection methods to elicit user need for support.	[P5, P6, P11]
	Methodological	A two-phased research method for eliciting user need for support based on *passive* data collection, and linking them to adequate solutions for support.	[P8, P9]
RQ1b	Empirical	A large-scale analysis of user problems, user coping strategies and wishes for support in four commercial intelligent everyday systems.	[P8, P9]
RQ2	Survey & Theoretical	A user-centric framework for defining and distinguishing different notions of user support.	[P3]
RQ3	Empirical	A lab user study indicating that explanations without information contribute similarly to trust in a system as real explanations.	[P4]
	Empirical	An online and lab user study to assess exploration as an implicit way of support interwoven with users’ primary task.	[P2]
	Methodological	A participatory stage-based design process for integrating user support in complex real-world scenarios.	[P6]
	Opinion	A normative and pragmatic perspective as lenses on supporting users in understanding intelligent systems in real-world settings.	[P7]
	Opinion	A reflection on possible *dark design patterns* of user support that deceive users for the benefit of other parties.	[P1]

Table 3.1: Overview of the primary contributions presented in this thesis.

3.1 Contributing Publications

This dissertation is cumulative: It consists of a list of previously published research projects, which contribute to the overarching narrative of this thesis. When referring to one of these contributing publications in the text, the format "[Pn]" is used (e.g. [P6]).

[P3], [P6], [P8] and [P9] are highlighted as the core contributions of my work. [P1], [P2], [P4], [P5], [P7], and [P11] complement these contributions by shedding light on specific details of the addressed research problems. [P10] adds to the overarching discussion beyond the scope of my work.

[P9] received an *Outstanding Paper Award* (top 5 %) at the respective conference.

[P1] Chromik, M., Eiband, M., Völkel, S. T. and Buschek, D. Dark Patterns of Explainability, Transparency, and User Control for Intelligent Systems. In: *Explainable Smart Systems Workshop at the 24th International Conference on Intelligent User Interfaces (IUI '19)*. `http://ceur-ws.org/Vol-2327/IUI19WS-ExSS2019-7.pdf`. 2019 (cited on pp. 13, 14, 22, 23, 25, 33, 34, 58).

[P2] Eiband, M., Anlauff, C., Ordenewitz, T., Zürn, M. and Hussmann, H. 'Understanding Algorithms Through Exploration: Supporting Knowledge Acquisition in Primary Tasks'. In: *Proceedings of Mensch und Computer 2019*. MuC'19. ACM, 2019, pp. 127–136. DOI: `10.1145/3340764.3340772` (cited on pp. 13, 14, 22, 24–26, 33–35, 37, 58).

[P3] Eiband, M., Buschek, D. and Hussmann, H. **'How to Support Users in Understanding Intelligent Systems? Structuring the Discussion'**. Submitted to *ACM Transactions on Interactive Intelligent Systems (TiiS)*. arXiv preprint: `http://arxiv.org/abs/2001.08301` (cited on pp. 2, 6, 13, 14, 20–22, 26, 31, 37, 38, 58).

[P4] Eiband, M., Buschek, D., Kremer, A. and Hussmann, H. 'The Impact of Placebic Explanations on Trust in Intelligent Systems'. In: *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI EA '19. ACM, 2019, LBW0243:1–LBW0243:6. DOI: `10.1145/3290607.3312787` (cited on pp. 13, 14, 22, 23, 25, 26, 33, 36, 58).

[P5] Eiband, M., Khamis, M., Zezschwitz, E. von, Hussmann, H. and Alt, F. 'Understanding Shoulder Surfing in the Wild: Stories from Users and Observers'. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI '17. ACM, 2017, pp. 4254–4265. DOI: `10.1145/3025453.3025636` (cited on pp. 13–17, 26, 58).

[P6] Eiband, M., Schneider, H., Bilandzic, M., Fazekas-Con, J., Haug, M. and Hussmann, H. '**Bringing Transparency Design into Practice**'. In: *23rd International Conference on Intelligent User Interfaces*. IUI '18. ACM, 2018, pp. 211–223. DOI: `10.1145/3172944.3172961` (cited on pp. 1, 3, 5, 7, 12–17, 22, 24–26, 29–31, 34, 35, 37, 58).

[P7] Eiband, M., Schneider, H. and Buschek, D. Normative vs Pragmatic: Two Perspectives on the Design of Explanations in Intelligent Systems. In: *Explainable Smart Systems Workshop at the 23th International Conference on Intelligent User Interfaces (IUI '18)*. `https://www.medien.ifi.lmu.de/pubdb/publications/pub/eiband2018iuiworkshop/eiband2018iuiworkshop.pdf`. 2018 (cited on pp. 6, 13, 14, 22, 25, 33, 37, 58).

[P8] Eiband, M., Völkel, S. T., Buschek, D., Cook, S. and Hussmann, H. **'A Method and Analysis to Elicit User-reported Problems in Intelligent Everyday Applications'**. To appear in *Special Issue of ACM Transactions on Interactive Intelligent Systems (TiiS)*. arXiv preprint: `https://arxiv.org/abs/2002.01288`. 2019 (cited on pp. 5, 13–19, 26, 30–32, 35, 37, 38, 58).

[P9] Eiband, M., Völkel, S. T., Buschek, D., Cook, S. and Hussmann, H. '**When People and Algorithms Meet: User-reported Problems in Intelligent Everyday Applications**'. In: *Proceedings of the 24th International Conference on Intelligent User Interfaces*. IUI '19. ACM, 2019, pp. 96–106. DOI: `10.1145/3301275.3302262` (cited on pp. 5, 13–19, 26, 30–32, 35, 37, 38, 58).

[P10] Schneider, H., Eiband, M., Ullrich, D. and Butz, A. 'Empowerment in HCI – A Survey and Framework'. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. ACM, 2018, 244:1–244:14. DOI: `10.1145/3173574.3173818` (cited on pp. 14, 26, 36, 58).

[P11] Schneider, H., Lachner, F., Eiband, M., George, C., Shah, P., Parab, C., Kukreja, A., Hussmann, H. and Butz, A. 'Privacy and Personalization: The Story of a Cross-cultural Field Study'. In: *ACM Interactions* 25.3 (2018), pp. 52–55. DOI: `10.1145/3197571` (cited on pp. 10, 13–17, 26, 30, 35, 58).

3.2 Empirical Research: A Method and Analysis to Elicit User Need for Support

The first objective of this thesis is an assessment of user need for being supported in understanding intelligent everyday systems in daily use as well as an exploration of suitable methods to do so.

3.2.1 A Methodological Exploration and Method to Elicit User Need for Support

For presentation clarity, I start with my contribution to research question –

> ***RQ1b**: What are suitable methods to capture user need for being supported in understanding intelligent everyday systems in daily interaction?*

I first present insights related to an exploration of established methods for *active* data collection used in [P5], [P6] and [P11]. This exploration revealed challenges that are then addressed by a method introduced in [P8] and [P9] which makes use of *passive* data collection.

An Exploration of Active Data Collection Methods [P5, P6, P11] In [P5, P6, P11], we explored the applicability of methods for *active* data collection to elicit user need for support in three different studies[1]. In particular, we assessed Flanagan's *critical incident technique* (CIT)[2] [32], drawing tasks [17, 53], semi-structured interviews, and think-aloud sessions [79].

We first applied the CIT in the context of smartphone use and privacy [P5]. The results revealed that the CIT can provide insights into problems with technology use in *daily interaction*.

In a second study [P11], we therefore used the CIT in combination with drawing tasks to elicit mental models of personalisation and privacy in Germany and India in semi-structured interviews. This field trip revealed a crucial limitation of active data collection when applied to a more conceptual level: Terminology like "personalisation" and "algorithm" is often not in a layman's vocabulary and additionally influenced by cultural background.

In a third study [P6], we applied the same methods to elicit user mental models and problems during interaction, this time tied to a specific system, a commercial intelligent fitness application. We found that using the CIT to investigate user need for support with a specific system mitigated the terminological constraints, and participants' drawings revealed misconceptions about the system.

Overall, these three studies revealed opportunities and constraints of *active* data collection methods:

1) The CIT is suited to provide insights into problems with technology use in *daily interaction* [P5].
2) However, its power is limited when applied to a more conceptual level due to terminological constraints [P11]. This could also be observed for drawing tasks and semi-structured interviews. We thus expect these terminological constraints to affect methods for *active* data collection in general when assessing user need for support.
3) These terminological constraints could be mitigated when referring to specific prototypes and use cases [P6].

A Method for Eliciting User Need for Support Based on Passive Data Collection [P8, P9] To overcome the methodological constraints of active data collection, we developed a two-phased research method that uses *passive* data collection to elicit user need for support in intelligent everyday systems as presented below (see Figure 3.1 for an overview).

Related work suggests that this need is linked to *practical problems*[3] users experience [8]. Therefore, our method extracts such *user problems* in the first phase:

[1] While [P5] was not related to intelligent systems yet, it provided insights which advanced our approaches and is therefore included in this thesis.

[2] The *critical incident technique* is a method to capture critical events via detailed oral or written accounts and thus allows for "generating a comprehensive and detailed description of a content domain" [32]. An incident can be considered *critical* "if it makes a 'significant' contribution, either positively or negatively" [32] to an activity or phenomenon. Since experiences with automation often have an emotional connotation [67], this technique has been successfully adopted in HCI research to understand how users actually feel before, during or after interaction (see e.g., [41]).

[3] I generally distinguish between problems that are merely related to classic system usability (e.g., system bugs) and those that can be attributed to the complexity of intelligent systems as explained in Section 1.2. While the boundary between these two types of problems may sometimes be fuzzy, I use the term *user problems* in the latter sense throughout this thesis.

1) **Identifying user problems (passive data collection):** The first phase aims at extracting a set of concrete problems users have faced in interaction with intelligent systems in the past. It exploits already existing, extensive sources of reported user experiences with intelligent systems as a basis for analysis.

   (a) *Collecting user experiences* from *existing* sources that reflect real-world user experiences with intelligent systems such as reviews of apps with intelligent functionality (e.g., from app stores) in contrast to assessing them with a user study.

   (b) *Finding the main topics* in the collected user experiences, for example, through statistical topic modelling.

   (c) *Extracting user problems* from the main topics, for example, through manual analysis.

From there, active data collection is applied to inform adequate *support solutions* to the identified problems:

2) **Developing support solutions (active data collection):** The second phase is then directed at assessing users' coping strategies to the extracted problems and wishes for support with the aim of informing adequate support solutions. To this end, the problem set is presented to users for in-depth insights. At the same time, this second phase serves as validation for the results from the first one.

   (a) *Representing user problems* in a way that can be shown to (other) users of the group that shall be supported, for example, as short scenarios.

   (b) *Asking users about problems* based on these representations, for example, in the form of a questionnaire.

   (c) *Informing support solutions* based on the insights from the previous step.

Contribution to RQ1b: A methodological exploration [P5, P6, P11] and method [P8, P9] to elicit user need for support

The contribution of this thesis to RQ1b is twofold: It comprises (1) an exploration of established *active* data collection methods that revealed terminological constraints in the context of intelligent systems, which makes them difficult to apply beyond specific prototypes [P5, P6, P11]; and (2) a two-phased method for eliciting user need for support which collects *passive* data about existing real-world user experiences with intelligent systems (e.g., via app reviews) to identify *practical user problems*, and links them to support solutions [P8, P9]. This method allows to combine automated and manual data processing and can thus be used to analyse a large dataset across a variety of domains while considering subtle nuances in the data (see Section 3.2.2 for an application example). By using existing datasets, it circumvents terminological constraints of active data collection.

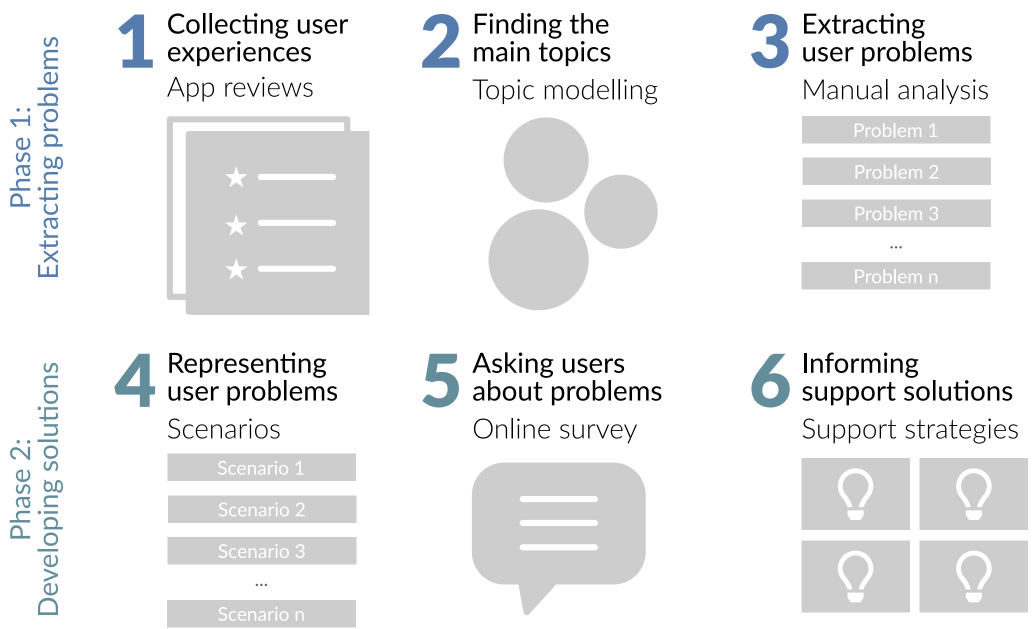

Figure 3.1: Our method for assessing user need for support in daily interaction with intelligent everyday systems. It consists of two phases with three steps each to (1) elicit practical user problems, and (2) inform adequate solutions for support.

3.2.2 An Analysis of User Need for Support

I next report on the application of the method presented in Section 3.2.1 to elicit user need for support in four commercial intelligent everyday systems (Facebook, Netflix, Google Maps and Google Assistant) [P8, P9], which contributes to RQ1a –

> ***RQ1a**: When do users need to be supported in understanding intelligent everyday systems in daily interaction?*

User Problems and Wishes for Support in Four Intelligent Everyday Systems [P8, P9] Applying the method presented in Section 3.2.1, we analysed user need for support based on 45,448 reviews of four apps on the Google Play Store that incorporate intelligent functionality, namely Facebook, Netflix, Google Maps, and Google Assistant.

We extracted user problems in daily use of these apps through topic modelling, sentiment analysis and manual coding. We then classified the identified problems along four categories of a basic pipeline for algorithmic decision-making: *knowledge base*, *algorithm*, *user choice*, and *user feedback*. The identified problems were presented as scenarios in a follow-up online survey to reveal how often participants had experienced a problem, how they had coped with it, and how they would like to have been supported by the system.

Overall, we found three reoccurring themes in the data, that reflect user need for support in daily interaction with these systems:

1) *Control*: Control issues and the wish for more and more fine-grained control overarched the results. Moreover, they indicated that the system should not alter a user decision – users want to have the last say. For example, our data revealed that the automated adjustment of Google Maps routing was often perceived as bothersome, in particular since it interfered with the primary driving task. Also, many users interacted with the system without making use of the intelligent features. For example, they navigated based on Google's map without the routing algorithm, looked for updates on their friends' profile on Facebook independent from the news feed, or browsed Netflix while ignoring the system recommendations. One option for more control could therefore be to design the system in a way that allows users to turn system intelligence on and off. In particular, this might help to "freeze" a user decision against further changes.

2) *Explanation*: We found an overall need for explanation of system workings in the analysis of reviews and the online survey, even though it was not as prevalent as the desire for more control. Our results show that users appreciate system suggestions, but want to make the final decision themselves on an informed basis. Our analysis revealed that this information should primarily describe the system suggestion *itself* in a comprehensive way, instead of explaining how and why it came into being. For example, users wanted more information about Netflix recommendations in order to assess their interest in a film rather than relying on their personal matching score. Suggestions should thus be presented with sufficient information to allow users themselves to assess the value of a suggested item (e.g. film, route, post in news feed).This adds a new perspective to prior work about explainability of intelligent systems, which so far has focused more on supporting users in understanding *why* a system decision has been made (e.g., [70]). Nevertheless, our results hint at potential benefits of interactive explanations, as suggested by Abdul et al. [1], in contrast to the predominantly static approaches presented in the literature (e.g., [52, 54]): Our participants wished for possibilities to try out different settings and observe effects in the algorithmic output, possibly "live", such as different ways to order the Facebook news feed.

3) *Correction and Feedback*: Options for feedback and corrections were a reoccurring theme in our data, but often sparse or difficult to find for users, or were not seen as helpful in their current state. For example, the binary feedback approach of "thumbs up" and "thumbs down" on Netflix was heavily criticised for lacking expressiveness. In contrast, a 5-level star rating was often mentioned as a more fine-grained, meaningful alternative. Moreover, a seemingly obvious, but crucial follow-up issue is to actually take feedback into account and confirming this to the user, as previously noted by Kulesza et al. [59].

Contribution to RQ1a: An analysis of user need for support [P8, P9]

Our work elicits user need for support by contributing a large-scale analysis of *practical problems* with intelligent everyday systems as reported by users [P8, P9]. These problems emerged from 45,448 reviews of four apps on the Google Play Store (Facebook, Netflix, Google Maps, and Google Assistant). We applied the research method presented in Section 3.2.1 using sentiment analysis, topic modelling and manual coding to extract user problems. We then assessed users' coping strategies and desired support through a follow-up online survey (N=286). The identified problems and strategies are related to the *knowledge base* and the *algorithm* of the apps, the *choice* users have, and the *feedback* mechanisms in place, and centred around *control*, *explanation*, and *correction and feedback*. These insights may inspire concrete design approaches and points of action for future work.

3.3 Conceptual Research: A User-Centric Framework of User Support

The second objective of this thesis is to introduce conceptual clarity and structure to the field to define, distinguish and conceptualise current and future approaches. To this end, I present [P3] which addresses the guiding research question –

> ***RQ2**: What are existing notions of user support for understanding intelligent systems in HCI and related fields, and how can they be conceptualised and distinguished?*

User Mindsets, User Involvement and Knowledge Outcomes [P3] Lim and Dey [68] established *user questions*, that is, questions about an intelligent systems and its workings such as *Why did the system do X?*, to surface and capture users' information demand. This approach has been taken up by subsequent work and has gained popularity as a way of rooting design decisions and solutions. For example, user questions are articulated in work by Kay and Kummerfeld [55] (*scrutability*), Kulesza et al. [59] (*end-user debugging*), or Rader et al. [92] (*transparency/accountability*).

We used such user questions as a lens to access and code the underlying notions of user support in prior work. In this way, we reviewed a corpus of about 250 papers. In particular, we included work on system *transparency*, *interpretability*, *scrutability*, *intelligibility*, *explainability*, *accountability*, *end-user debugging* and *interactive machine learning*. Our analysis surfaced three categories to define user support and differentiate current approaches in the literature. These three categories describe how users form a goal they want to reach in the first place (before interaction), how they are involved in interaction (during interaction), and the understanding, that is, the knowledge stemming from interaction or other sources (after interaction).

1) *User mindsets* – what users *seek* to know: The first category describes what psychology calls the "cognitive orientation" of people [36], based on which they form a goal and plan successive actions towards reaching that goal. Our review surfaced three such mindsets assumed or addressed in prior work: *utilitarian*, *interpretive*, and *critical*. A *utilitarian* mindset is directed towards *usability and utility*, for example, when users want to understand system recommendations to *better compare* products they are interested in [91]. An *interpretive* mindset carries a notion of *user perceptions and experience*, such as when users want to understand how they are being profiled [7]. Finally, a *critical* mindset stresses *ethical and legal reflection* about intelligent systems, for example, when user want to know if a system is fair [92].

2) *User involvement* – *how* users acquire knowledge: The second category concerns the nature of user involvement during interaction, that is, *in which way* users acquire knowledge. We propose two fundamental distinctions in the field: Users are given an *active* or *passive* role. For example, a design for an *active* user might include options for control, correction and feedback

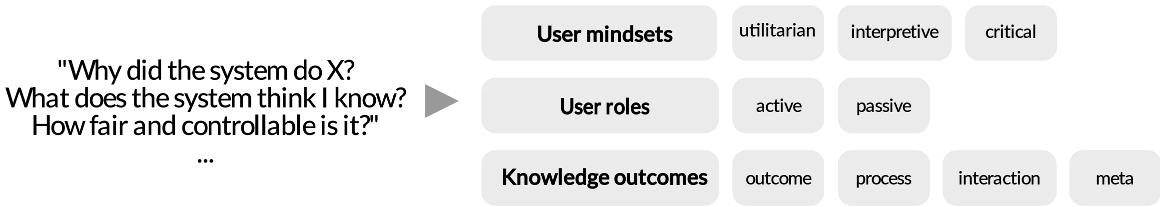

Figure 3.2: Our user-centric framework for defining and distinguishing different notions of user support as emerged through the lens of (implied) user questions. The three categories *user mindsets*, *user involvement*, and *knowledge outcomes* describe what users seek to know, how they acquire knowledge during interaction, and what kind of knowledge they acquire.

(e.g., [59]) and users are seen in active roles such as debuggers [61] or teachers [3]. A system design supporting a *passive* user lets users acquire knowledge by displaying information, but does not offer user feedback and corrections (e.g., [35]).

3) *Knowledge outcomes* – what *kind of knowledge* users acquire: The third category is the type of knowledge users acquire. We extract four knowledge types from the literature: (*outcome*, *process*, *interaction*, and *meta* knowledge). *Outcome* knowledge targets *individual instances* of an intelligent systems (e.g., understanding a specific movie recommendation). In contrast, *process* knowledge includes the system's *underlying model and reasoning* (e.g., the workings of a neural network that processes movie watching behaviour). *Interaction* knowledge describes understanding *how to get the system to do something*, and *meta* knowledge captures knowledge about a system *beyond interaction situations*, gained, for example via developer blogs or other media. These knowledge types, except for *meta knowledge*, can be mapped to those from complex problem solving as presented in Section 1.2, namely *input-output*, *structural* and *strategy* knowledge. Moreover, knowledge can have two qualities, *rigour* and *relevance*. *Rigor* refers to *completeness* of information, *relevance* to the information *required to reach a goal*.

Contribution to RQ2: A user-centric framework of user support [P3]

This thesis contributes a user-centric framework of support [P3] to introduce conceptual clarity and an overarching structure to the field. This framework is derived from a review of existing concepts and prototype solutions within diverse lines of research, and incorporates the categories *user mindsets*, *user involvement*, and *knowledge outcomes*. It aims to resolve conceptual ambiguity in the field by enabling researchers to clarify their notion of user support and become aware of those made in prior work. With this work, I hope to advance the ongoing discussion among scholars towards an overarching perspective on supporting users in understanding intelligent systems.

3.4 Constructive Research: A Normative and Pragmatic Design Exploration of User Support

The third objective of this thesis is to assess the applicability of user support given real-world constraints and challenges and thus address the last guiding research question –

> ***RQ3**: How can user support for understanding intelligent everyday systems be designed given real-world constraints and challenges?*

To approach this question, I present two perspectives that can be identified in the literature and the public discourse about intelligent systems, a *normative* and a *pragmatic* one [P7]. These perspectives serve as lenses for a design exploration in several case studies and arguments. [P1] and [P4] assess and discuss possible side-effects of the design tensions introduced by the normative perspective. [P2] and [P6] contribute to the pragmatic perspective through user-centric design approaches.

A Normative and Pragmatic Perspective on User Support [P7] One can identify two perspectives in the larger discussion about real-world applicability of user support, a *normative* and a *pragmatic* one:

1) A *normative* perspective on user support is based on *ethical and legal reasoning*. This view is reflected in recent legislation, in particular the General Data Protection Regulation enforced on 25 May 2018. For example, it has provided citizens of the European Union with what has been called a "right to explanation" [37], granting users access to "meaningful information about the logic involved, as well as the significance and the envisaged consequences of [automated decision-making] for the data subject" [89]. This perspective can also found in the academic literature [9, 15, 16, 45]. For example, Hildebrandt [45] claims that "[algorithmic] decisions that seriously affect individuals' capabilities must be constructed in ways that are comprehensible as well as contestable. If that is not possible, or, as long as this is not possible, such decisions are unlawful [...]". A normative perspective thus claims an *option* for support. It does not further specify what "meaningful information" is nor how it is going to be presented to users – most importantly, it should be *available* to users.

2) A *pragmatic* perspective focuses on *user need* for support to help people use a system "better and in more informed ways, and to better ends and outcomes" [P3]. It can be found in most of the work on interaction with intelligent systems presented in HCI (e.g., [59, 69]). From that perspective, it is the users who determine what "good" support is, not ethical reasoning. In contrast to the normative perspective, this claim is not necessarily met by a "right to explanation". For example, excessive explanations would rather be seen as information overload and thus an obstacle towards using a system better. Instead, user support has to provide information that is interesting and helpful for users in a particular situation or during a particular interaction, in an adequate level of detail, and integrated into the interface and user flow in a way so as to keep cognitive load as low as possible. From a pragmatic perspective, explanations detached from the interface and user flow are unlikely to be accessed by users in the first place, let alone seen as helpful. In face of real-world constraints, a pragmatic perspective strives to balance all stakeholders' interests and to find design trade-offs.

3.4.1 Normative: Placebo Effects and Dark Patterns of User Support

This thesis first explores possible design tensions introduced by normative desiderata in the face of real-world constraints and challenges. To this end, it contributes a user study on placebo effects of explanations [P4] and a collection of *dark design patterns* as a thought-experiment [P1].

Placebo Effects of User Support [P4] Prior work has criticised normative desiderata for user support, cautioning against a “transparency fallacy” [22] where practically irrelevant information is used to soothe users [22, 108, 113].

We investigated this critique for interaction with an intelligent everyday systems. In particular, we based our work on research from social psychology, which has demonstrated that people are more likely to comply to a request if they are presented with a justification – even if this justification conveys no information [64]. Using a prototype of a nutrition recommender, we conducted a lab study between three groups (each N=10) with different types of explanations (*no explanation*, *placebo explanation*, and *real explanation*). In line with [64], placebo explanations initiated a justification with “because/since/so that ...”, but did not reveal more information about the system than could be inferred from the study setup. Real explanations, on the other hand, provided details about the system and its decision-making.

Participants were asked to use the recommender to create a personalised meal based on a scenario they were given. We then assessed their trust in the system suggestions. Our results indicate that placebo explanations indeed invoked a level of trust similar to real explanations. They thus support concerns about “empty explanations” [108] as a psychological tool, and motivate further investigation of the role of cognitive processes and biases in human reasoning in interaction with intelligent systems.

Dark Patterns of User Support [P1] Prior work has raised concerns that normative claims for user support in intelligent systems may result in solutions that do not necessarily *benefit* users, given real-world constraints [22].

As a thought experiment, we argued that user support may even be exploited to purposefully *deceive* users for the benefit of *other* parties. Such carefully crafted deceptive interface design solutions have gained notoriety as *dark design patterns* [6]. We built on the work by Gray et al. [38] to present a collection of exemplary dark patterns for explainability, transparency and control. For example, we introduced the *information overload* pattern, where explanations are presented in a very lengthy format using technical language, similar to what we currently see in end user license agreements. Another example is the *explanation marketing* pattern, where an explanation for why a product was recommended is used to promote other products.

Hence, while legal efforts concerning the fairness of algorithmic data manipulation are urgently needed, we see a risk that these normative interventions might negatively impact interaction with intelligent systems in other ways. With our work, we contribute to a more nuanced view on the complexity of this interplay in practice.

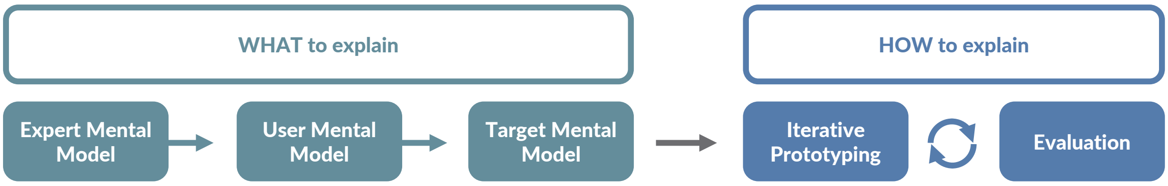

Figure 3.3: Our stage-based participatory design process for designing user support in real-world settings. The first three stages focus on *what to explain* in the system (information), the last two on *how to explain* (presentation).

3.4.2 Pragmatic: A Participatory Design Process and Implicit User Support

This thesis next explores how a pragmatic perspective may impact design approaches for user support. In particular, [P6] assesses how different stakeholder needs can be met in practice, and [P2] explores ways to interweave knowledge acquisition with users' primary task.

A Stage-based Participatory Design Process [P6] This thesis contributes a stage-based participatory process for designing user support (see Figure 3.3). This process takes the perspectives of users, designers and company stakeholders into account, and can be adapted to a specific real-world setting. Notably, our process has been taken up by subsequent work to design explanations for autonomous driving [110]. The process was developed and validated with a commercial intelligent fitness application in close cooperation with company staff. It is divided into two parts, where the first part defines the information that is to be transferred (*what to explain*), the second their presentation (*how to explain*). The stages are each guided by central underlying questions and involve different stakeholders. In total, the process incorporates seven stages:

1) ***What to explain?*** The first phase of our design process aims at defining the *information* to be transferred to users by synthesising expert and user knowledge, represented by their respective mental models.

    (a) *Expert mental model*: Crucial steps and features of the algorithm in place are summarised in a hypothetical "optimal" *expert mental model*.

    (b) *User mental model*: A blueprint *user mental model* is synthesised from the knowledge users currently possess about the system, and any deviations from the expert mental model are recorded.

    (c) *Target Mental Model*: Based on the differences between expert and user mental model, users select the information they perceive to be most relevant, interesting and helpful in their preferred level of detail. This information is added to the current user mental model and determines the focus of the next phase as *target mental model*.

2) ***How to explain?*** The second phase focuses on the *presentation* of the selected information in the system interface.

    (d) *Iterative prototyping*: Possible presentation formats that take corporate design guidelines and the user flow of the system into account are explored.

(e) *Design evaluation*: To evaluate the new system design, the user mental model is elicited based on the prototype and compared to the target mental model. If there are discrepancies between both models, stage 2 d) should be repeated.

With our work, we provide guidance to researchers and practitioners that meets the complex needs arising in real-world scenarios, and support the evolvement of best practices on realistic case-by-case learnings.

Implicit Approaches to User Support [P2] Many solutions for user support presented in prior work require users to take additional effort, for example for invoking pop-ups with information [13], or explicitly asking for explanation [69]. While this interface design on the one hand avoids information overload, it is not part of users' primary task and thus may disrupt them during interaction (see [8]).

We investigated *exploration* as a more implicit way of interweaving knowledge acquisition with users' primary task [P2]. We conducted a think-aloud study in the lab (N=10) as well as an MTurk online study (N=113) using a flight booking scenario in two different tasks. One group focused on finding the cheapest flight (knowledge acquisition as secondary task), the other on understanding the underlying system rules (knowledge acquisition as primary task).

Our results indicate that exploration, even as a secondary task, may contribute to knowledge about the underlying system workings. However, our study also suggests that the overall knowledge acquired through exploration is limited: It gives people an idea of how a system works, rather than teaching them concrete rules they can recall. We conclude that exploration has the potential to contribute to, but not substitute, existing more explicit design of explanation interfaces.

Contribution to RQ3: A Normative and Pragmatic Design Exploration of User Support [P1, P2, P4, P6, P7]

My work contributes an exploration of real-world applicability of user support through the lens of two perspectives, a normative and a pragmatic one, as identified in [P7]. A normative perspective stems from recent legislation and sees means for user support as prerequisite for deploying intelligent systems, while a pragmatic perspective aims at reconciling the different trade-offs involved in real-world settings. In summary, this exploration contributes (1) a user study showing that explanations can be used as a placebo to foster user trust [P4]; (2) a set of possible *dark design patterns*, where user support may purposefully deceive users for the benefit of other parties [P1]; (3) a stage-based participatory design process for user support [P6], which incorporates the needs of different stakeholders and can be adapted to a specific real-world setting; and (4) a user study that explores how user support can be interwoven with users' primary task [P2].

3.5 Research Methods and Limitations

The work presented in this thesis results from the application of diverse research methods. Table 3.2 provides an overview of all contributing papers and the concrete methods employed. Some publications comprise more than one study and are therefore listed several times. In the following, I describe general methodological limitations of my work that are particular to the context of intelligent everyday systems.

Contributing Paper	Research Method	Data Type	Assessment	Dataset
[P2]	Qualitative lab study, think aloud	Qualitative	User understanding of system rules as primary and secondary task.	10
[P2]	Online study	Quantitative	User understanding of system rules as primary and secondary task.	123
[P3]	Structured literature review	Qualitative	Review of existing work in HCI on the notion of *user support*.	250
[P4]	Prototype design and lab study	Qualitative & quantitative	Impact of different types of explanations on user trust.	30
[P5]	Critical incident technique, online survey	Qualitative	Real-world shoulder surfing scenarios in daily smartphone use.	174
[P6]	Participatory Action Design Research	Qualitative	Integration of explanations into a commercial fitness app.	40
[P8], [P9]	Topic Modelling	Qualitative	Extraction of user problems from app reviews.	45,448
[P8], [P9]	Online survey	Qualitative & quantitative	User problems, coping strategies and wishes for support for Facebook, Netflix, Google Maps and Google Assistant.	286
[P10]	Structured literature review	Qualitative	Review of existing work in HCI on the notion of *empowerment*.	54
[P11]	Critical incident technique, interviews, observations, drawing tasks	Qualitative	Participants' mental models on data privacy and personalisation.	16

Table 3.2: Overview of the research methods used in the publications contributing to this thesis along with a short summary and the dataset size.

Experiment Controllability versus Ecological Validity Investigating user understanding of intelligent systems comes with a trade-off between experiment controllability and ecological validity. In an experimental setting, it is often necessary to compare knowledge users have acquired to a ground truth in terms of the internal mechanisms of the system. This requires a clear model of these mechanisms which can be compared with the understanding users had developed during use. However, such a ground truth does not exist, or only on a more abstract level, for systems that dynamically evolve based on the data they operate on. In [P2, P4, P6], this problem was approached by reducing the system complexity to fixed behaviour, considering it a "snapshot" of an intelligent system. While this can be seen as an inherent limitation of the prototypes presented in my work, it accounted for the aforementioned trade-off in the experimental design.

Assessing Intelligent Everyday Systems as an Outsider Technical details about commercial intelligent everyday systems are mostly not publicly available. As part of the cooperation in [P6], insights into the workings of the underlying algorithm were given. However, in other instances [P8, P9], the presented work was conducted without the involvement of the companies whose products were investigated. Thus, it was not possible to verify if the problems users reported can be attributed to actual problems in the respective algorithm at the time of data collection and analysis. Different experiences might also be caused by different versions of the respective system. However, to inform user support it is the very problems that users *experience* that were of interest, be they caused by actual malfunction of the system or not.

4

Discussion and Future Work

The main objective of the work presented in this thesis is to assist users in gaining the knowledge necessary to reach their goal in interaction with intelligent everyday systems. I have introduced *complex problem solving* from cognitive science as a theoretical foundation of my work, which sets the stage for *empirical*, *conceptual* and *constructive* research to address this objective. As part of empirical research, I have presented a method to elicit *user need for support* through practical user problems with intelligent everyday systems in daily interaction and have reported on a large-scale analysis of such problems and users' coping strategies. The presented conceptual research makes explicit different underlying assumptions about user support in the field by introducing a *user-centric framework to conceptualise user support*. Finally, as part of constructive research, I have explored the *applicability of user support* given real-world constraints and challenges through the lens of a normative and a pragmatic perspective.

In this chapter, I introduce five research challenges that have emerged from my work and that I deem most important to be addressed for advancing the field. Together, these challenges form a research agenda for work on supporting users in understanding intelligent everyday systems. For each of these challenges, I discuss learnings from my work and critically reflect on its limitations, and point out possible pathways for future research. This agenda spans the insights gained in the presented thesis and can thus help other researchers as guidance and inspiration.

4.1 A Research Agenda for Supporting Users in Understanding Intelligent Everyday Systems

This thesis is guided by a pragmatic perspective on supporting users in understanding intelligent everyday systems as presented in Section 3.4, from the grounding in complex problem solving to the staged-based design process of user support introduced in [P6]. In focus of this perspective are the users and their need for support as it arises in everyday interaction. With this, I position my work in line with Fischer and Lemke [28] who claimed that "most computer users are not interested in computers per se, but they want to use the computer as a tool to [...] accomplish their tasks", and called for moving from human computer interaction to *human problem-domain interaction*. In the context of this thesis, this domain is the *everyday*: Support does not mean to make users understand the system itself, but should always be related to a user need and relevant to a user goal for solving everyday tasks.

This contrasts with prior work, in which solutions for support are often designed *first* and *then* evaluated as to their relevance (e.g., [57, 58, 69, 70, 91, 92]). While these studies have greatly advanced the field in terms of exploring possible approaches and demonstrating their positive effects, I argue that research should now move towards designing solutions for support that are driven by users' needs and goals to ensure that these solutions are coupled to real-world use.

Where to start? "The challenge in an information-rich world is not only to make information available to people at any time, at any place, and in any form, but specifically to say the 'right' thing at the 'right' time in the 'right' way" [26]. While this statement by Fischer reflects the overall requirements that come with the pragmatic perspective on supporting users in understanding intelligent everyday systems, I introduce five research challenges to make what he calls "right" more actionable and concrete. Some of these challenges have been explicitly addressed in my work, some have emerged from the presented approaches and studies. They can both be seen as a sequence of steps to be taken into account for designing solutions for support, or as individual research topics for future work.

4.1.1 Eliciting User Need for Support

The first challenge concerns the application of suitable *research methods and approaches* for capturing real-world use of intelligent everyday systems and eliciting user need for support. This identifies and delineates HCI issues as a basis for design solutions and ensures that these solutions adequately reflect real-world user needs.

Working towards Best Practices in Methodology The field needs to work towards best practices in terms of methodology to elicit user need for support in different scenarios. As an exploration of approaches in this thesis has shown (see Section 3.2.1), the actual choice of methods for eliciting user need for support depends on the concrete research endeavour. Eliciting user need for support based on a specific system, or a prototype of such a system, with a limited participant sample can be done through established methods like observations, interviews paired with the critical incident technique and drawing tasks [P6, P11]. Another approach suitable for this scenario has been presented by Bucher [7], who collected user statements on Twitter through manually selected search terms.

Assessing need for support on a more abstract level, on a larger scale, or over a variety of application domains requires different methods. One such method has been presented in this thesis [P8, P9]. It is based on the collection of *passive* data from existing sources to surface *practical user problems* with intelligent everyday systems. In my work, this method was applied to publicly available *user reviews*, which can be easily collected on a large scale through automated means, and which reflect users' attitude towards an application as well as their experiences [40, 73]. The user problems emerged from these reviews through a combination of automated preprocessing with sentiment analysis and topic modelling and subsequent manual coding. This data was then enriched and validated through a follow-up online survey using scenario-based questionnaires to assess users' coping strategies and their wishes for support. Being mostly qualitative, the presented analysis allows for deep insights into how people use intelligent everyday systems in their lives and which challenges they face during interaction. Also, by including automated means for data processing, it can account for a comparably large data set and cover a variety of application domains in contrast to methods that rely on purely manual analysis. A large-scale analysis of user need for support in four commercial intelligent everyday systems (Facebook, Netflix, Google Maps, and Google Assistant) [P8, P9] has demonstrated the utility of this method. However, since manual analysis nevertheless plays a crucial role in the analysis process, the presented method shares the limitations of all qualitative approaches: They are time-consuming. Future work should therefore explore if the manual analysis steps of the presented method can be (in parts) automated. Moreover, subsequent research might validate the applicability of the method in other scenarios, or explore its application to other datasets, such as data from issue-tracking systems.

Quantifying User Need for Support Future work should find approaches to quantify and measure user need for support. Adequate operationalisation of this need might pave the way for implicit data collection based on *how* a user interacts with a system. Prior work on so-called *behaviour-aware* systems has demonstrated that this is a promising way to adapt interfaces to individual users and thus improve interaction [10]. Similarly, the system could implicitly infer when to support the user with which information. For example, it could adapt to the knowledge a user has already acquired, or could intervene just-in-time when users face problems (e.g., in understanding a particular system suggestion). Such interventions and concepts have already been presented in early papers on mixed-initiative interaction principles (e.g., [31, 49]), and could enrich the landscape of solutions for user support. One exemplary approach from the literature are personalised explanations [81] that take into account contextual information such as user location.

In a larger perspective, such quantitative measures could also be used to optimise interface design for user understanding and to systematically evaluate different interface versions.

Considering Affective Factors Analysing practical problems to elicit user need for support is in line with the overall pragmatic perspective of this thesis and related work [8]. However, there are certainly other factors that impact such need – one of them is affect. Prior work has already presented insights in this direction. For example, Bucher [7] focused on the affective consequences of interaction with intelligent everyday systems, and Alvarado et al. [2] proposed the concept of *algorithmic experience* as a foundation for system design. In the work presented here, the app reviews collected for analysis in [P8] and [P9] indicated a relation between user need for support and negative affect – the reviews were therefore pre-processed via sentiment analysis before applying topic modelling to elicit user problems in the data. Future work should thus consider both practical need for support as it arises in interaction and how this need is coupled to affect. This might also provide starting points for measuring user need for support: Work on affective computing [87] could offer inspiration in this regard.

4.1.2 Establishing Conceptual and Terminological Clarity for User Support

The second challenge is to establish conceptual and terminological clarity for user support in HCI research. This allows to develop a common understanding beyond the plethora of different approaches, terms and concepts that have been presented in the field.

Considering Different Notions of User Support In [P3], I have categorised different notions of how users should be supported as have emerged from a survey of prior work: *user mindset*, *user involvement*, and *knowledge outcome*. Together, these categories form a conceptual framework for supporting users in understanding intelligent systems. This framework introduces a user-centric overarching structure to the discussion about user support that is independent from the inconsistent and varying terminology in the field. It allows to effectively analyse and describe commonalities and differences between approaches presented in prior work. Moreover, researchers can use it to make their own assumptions explicit. To the best of my knowledge, this presents the first research attempt in this direction.

As an illustration, I use this framework to locate the case study presented along with the participatory design process in [P6]. This research adopts a *utilitarian user mindset*: Supporting users is aimed at helping them train more efficiently and avoid injuries. *User involvement* is *passive* through displaying textual and animated explanations in the interface, for example about a user's performance level. As *knowledge outcome*, the adopted design targets *outcome* and *process knowledge* by giving users information about a suggested training plan, about the data the calculation was based on and about how it influenced the system output. This is restricted to information *relevant* to the above goals.

Future work should explore the advantages and disadvantages of the respective categories when designing support solutions in different interaction situations. For example, supporting a *utilitarian* mindset might probably be most helpful in daily interaction of low-cost systems for effective and efficient use. However, when high risk is involved, users might rather adopt an *interpretive* (e.g., wanting to assess the reasons for a system decision) or even *critical* mindset (e.g., wanting to know if a system is fair). User support could then be designed so as to target one particular or even several *user mindsets*, in the form of interface "modes" that users can switch between. Moreover, prior research has shown that people rather repeat actions physically than spend the mental resources on planning to avoid these repetitions [83]. *Active* user involvement in the form of interactive and explorative user support could therefore be used to reduce cognitive load during interaction. This also concerns interactive machine learning approaches (e.g., [59]), where the accuracy of system predictions and suggestions benefits from user correction and feedback. However, if users cannot explore the system, such as in time-critical tasks, a short explanation and thus *passive* user involvement might be more desirable. Furthermore, daily use of intelligent everyday systems might profit from supporting *interaction* knowledge, that is, knowing how to do something to influence the system. Also, [P8, P9] highlight the need for *outcome*, not *process* knowledge in intelligent everyday systems. Again, this could be different in other types of intelligent systems like domain-specific systems, where process knowledge is needed to assess if a system is trustworthy (e.g., [94]).

Finding Joint Definitions and Terms A literature mapping by Abdul et al. [1] has demonstrated the fractured landscape of terms and concepts currently prevalent in HCI and adjoining fields. Future research needs to work towards a joint terminology and clear definitions of terms and how they are interrelated. At the moment, definitions in the field are manifold and need to be distilled: For example, *transparency* is described as "providing users with ways to increase their understanding of how a system works" [14], "a way to see inside the truth of a system" [5], "a method to see, understand and govern complex systems in timely fashions" [5], or "visibility of system status" [103]. Likewise, future work should find consensus about interrelations between different terms and concepts. For example, Rader et al. [92] have linked explanations as "transparency mechanisms" to *transparency*, *interpretability*, *accountability*, and *correctability*. More research along these lines is needed to derive an overview of terms and concepts and their relations.

Notable work on terms and definitions has been presented by Doshi-Velez and Kim [18] with regard to *interpretability*, or by Nunes and Jannach [84] for *explanations*. These examples could form a basis for further research attempts towards more terminological clarity and rigour in the field.

4.1.3 Considering Limitations of Understanding

The third challenge refers to limitations of human understanding, which might diminish the effectivity of user support. I list two such limitations as have emerged most prominently from my work and the research involved.

Mitigating Human Biases The first limitation concerns *human biases* that may impact understanding of intelligent systems in unpredictable ways. The user study in [P2] revealed that mental models of a system may be *persistent* in a way that people rather try and fit observations to their existing knowledge than to update it according to experienced discrepancies. This phenomenon is empirically established as *cognitive lockup* [75, 56], where "operators [...] become rigidly stuck in one mode of behaviour [...], rather than exploring alternative possibilities during fault diagnosis and management".

These biases potentially present obstacles for users on their way to a goal, which is an even more urgent challenge for intelligent systems associated with high-risk where users have to make high-cost decisions, possibly under time-pressure. Future work needs to look into these biases and their effects in detail, and find ways of mitigating them. Recent work on user-centric explainable artificial intelligence [107], in which the authors highlight the importance of considering human biases and heuristics in explanations, might serve as a starting point.

Dealing with Human Reasoning and Feelings One argument brought forth in the literature is that even though the human brain has inspired the development of nowadays'intelligent systems, human-scale reasoning and styles of semantic interpretation are inherently different from that of machine reasoning [9]. This means that we humans may not have semantic concepts for the factors that caused a system result, or the imagination necessary to picture a large variable space. Explanation of artificial intelligence thus requires inherent abstraction from the algorithmic complexity [94, 95]. As a result, Burrell cautions that "explanations that bring forward a human-managable list of key criteria provide an understanding that is at best incomplete and at worst false reassurance".

From the pragmatic perspective presented in my work, user support should not aim for completeness of information. Rather, abstraction from the underlying system complexity may often be needed in order to design support in a way that is most helpful to users to reach their goal. Yet, the placebo explanations investigated in [P4] show that the mere wording of information can impact user trust in a system and thus might "falsely" reassure users. While effects on user feelings like trust and satisfaction were otherwise not in focus of this thesis, future work should explore the design tension between abstraction of information and such unfounded reassurance. The question of whether making users "feel" good when using an intelligent systems is desirable or not is further discussed in Section 4.1.5.

4.1.4 Applying User Support in Practice

Finding the right balance between the different trade-offs faced in real-world scenarios is the fourth research challenge. This ensures that all stakeholders' needs are met and that support benefits users during their actual task.

Considering Different Desiderata for Support I have presented two perspectives as a lens on how real-world applicability of user support may be approached: a normative and a pragmatic one [P7], each with their own desiderata. The *normative* perspective is based on ethical reasoning about intelligent systems (e.g., displaying explanations as a prerequisite for the usage of intelligent systems and a user right), the *pragmatic* one tries to balance the different needs that have to be met in practice (e.g., integrating explanations without hampering the usability of a system).

[P1] and [P4] critically reflect on side-effects when real-world challenges meet normative desiderata as expressed in recent legislation [89, 90]. [P4] shows that explanations without information can similarly invoke user trust in a nutrition recommendation system as real information. Such "empty" explanations may then be formally meet the "right to explanation", but will not help users in practice to reach a goal. On the same lines, [P1] presents a collection of possible *dark design patterns* for user support as a thought experiment, where user support is purposefully designed so as to benefit *other* parties. Such dark patterns might still be consistent with the normative perspective (e.g., providing "meaningful" information about the system), but might also be exploited to trick users.

On the other hand, [P2] and [P6] explore user support from the pragmatic perspective. This perspective is closely related to the principle of *satisficing* [101]: User support should be designed in a way that helps users to reach a goal, no less, no more. This is challenging in practice, since one has to face different trade-offs: Reflecting the complexity of the system versus minimising cognitive load for users, adding interface elements for user support versus distracting users from their actual task, taking user needs into account versus considering the needs of other stakeholders. In this sense, the stage-based participatory process in [P6] is a process for designing for satisficing. It allows to include all stakeholders' needs and to find design solutions for user support to meet these needs. It thus serves practitioners as a starting point for their work.

To conclude, the choice of perspective most likely depends on the type of system at hand. In the context of intelligent everyday systems in focus of my work, I have adopted the pragmatic perspective since these systems are mostly low-risk. In high-risk systems, for example, for medical diagnoses where human lives are at stake, the normative perspective might more appropriately guide system design. Also, the perspectives are on different ends of a spectrum – for some systems, a perspective "in between" might prove most sensible. For example, highly automated driving involves high-risk, but also has to integrate user support in a way that quickly gives an idea of what is happening.

Achieving Knowledge-Acquistion-by-Design Another challenge for real-world applicability of user support is its integration into interaction. Explanations, for example, are often decoupled from users' primary task (e.g., as pop-ups [13, 77]). Access then requires users to invest additional effort and thus directly competes with their actual task, as also noted by Bunt et al. [8].

In a larger perspective, instead of building interfaces and adding information about the system workings a posteriori, I argue that we as researchers should be rethinking interfaces for intelligent systems in the direction of what I call *knowledge-acquisition-by-design*:

1) *Design processes:* Designing user support should be a part of the system design process from the start. Otherwise, we will always run the risk that applied solutions will be decoupled from users' problems and needs or involve too much effort to be accessed. In [P6], I have presented one possible blueprint for such a process. This process emerged from a case study in cooperation with a commercial intelligent fitness application and has been taken up by subsequent work [110]. This indicates the validity of the process beyond the scope of the presented case and its applicability to other, similarly complex scenarios involving other types of systems. However, variants of the process may exist that are better suited for specific scenarios. Future work should therefore validate, extend and refine the process in different scenarios and application domains.

2) *Interface elements and interaction design:* We should rethink the design of interface elements and interaction in a way that closely interweaves knowledge acquisition with users' primary task. This means that the interface and possible interactions should already carry information

about the underlying system complexity. In [P2], I have presented a first step in this direction by assessing the potential of exploration for knowledge acquisition as part of users' primary task. The results show that such *implicit* user support can contribute to knowledge acquisition. This is in line with recent work calling for more interactive user support in contrast to many static approaches presented to date [1]. Yet, this is admittedly a first exploration of the topic, and certainly needs more research in the future. For example, *What if?* user questions [68] present an interesting anchor for interactive and implicit user support.

Another important aspect is the implementation of usable loops for correction and expressive feedback, as has emerged from our analysis in [P8] and [P9]. An obvious improvement would be to integrate options for correction and feedback in a way that is more accessible to users, such as directly beside a system suggestion instead of in a possibly mutlilayered menu.

Further inspiration could be found in HCI areas facing similar challenges, such as *Usable Security and Privacy*. For example, neither password entry nor privacy notices are part of users' primary task – researchers have therefore worked on approaches towards *usable security* [112] and *privacy-by-design* [76]. In particular, Schaub et al.'s design space for effective privacy notices [99] could provide interesting starting points in this direction.

4.1.5 Investigating Effects of User Support

The fifth research challenge comprises the possible effects of user support.

Evaluating User Understanding Future research should investigate how user understanding can best be evaluated. A widely adopted representation of this understanding in HCI are *mental models* [83], and this is no different in related work on supporting user understanding (e.g., [60, 62, 104]). This construct can also be found in complex problem solving, as presented in Section 1.2. Mental models stem from a constructivist perspective on knowledge, where knowledge entails individually constructed, subjective interpretations of the world that are based on experiences and assumptions [98]. Thus, mental models exist within the mind and can therefore not be readily inspected and measured.

In HCI, mental models are often inferred through interrogation of participants. Approaches include interviews and in situ surveys [104], multiple-choice questions [60], or questionnaires [62]. The presented work makes use of semi-structured interviews [P11], think-aloud sessions [P2, P6], multiple-choice questions [P2], and drawing tasks [P6, P11].

While such techniques are valid to gain a first understanding to elicit users' mental models [79], they also have their inherent limitations.

For example, Norman [83] claims that "you cannot simply go up to the person and ask. Verbal protocols taken while the person does a task will be informative, but incomplete". He roots this claim in possible discrepancies between people's beliefs and actions, the non-verbalisable nature of implicit knowledge, and social desirability effects in the experiment situation, where people tend to tell what they believe the researcher wants to hear.

For a more grounded analysis of the effectivity of user support, future research should therefore explore methods as well as their advantages and drawbacks to elicit mental models and evaluate user understanding. For example, Norman recommends to *implicitly* infer mental models by collecting data during experiment tasks. Inspiration could be found in work on complex problem solving, where a variety of standardised tasks and task environments has been presented (see e.g., [34]).

Considering Effects of Support beyond User Goals HCI research has often been divided into three so-called paradigms, or waves, each wave carrying their own background, approaches, claims and motivations [43]. Along these lines, my work shares the values of second-wave HCI, drawing on a model from cognitive science and focusing on information transfer between user and system.

However, user support could also be assessed through the lens of third-wave HCI, where "values in design" and meaning are in focus of interaction [43]. One of these values that can also be found in work on interaction with intelligent systems is *empowerment* (e.g., [3, 29, 30, 92]).

In an early paper, Fischer and Nakakoji have already called for user empowerment instead of replacement in interaction with intelligent systems. Moreover, Rader et al. [92] claim that "transparency can empower users to make informed choices about how they use an algorithmic decision-making system and judge its potential consequences".

However, research in other areas of HCI has cautioned against considering empowerment an unconditionally positive mission, since design approaches claiming to empower users might ultimately turn out *disempowering* [66, 106]. In the context of this work, providing explanations, for example, is a fine line between supporting users in reaching their goal and overwhelming them with information or interfering with their actual task.

To reflect in more depth on what we as researchers in the field strive for, I draw on the framework for user empowerment by Schneider et al. [P10], which defines empowerment along the lines of four categories. I illustrate these categories with examples from my work.

The first category describes the *concept of power* that underlies empowerment, namely if power is considered a growing ability (*power-to*) or a limited resource divided between two actors (*power-over*). Do we want to foster users' abilities through support or do we see it as a means to calibrate power between user and system (or system providers)? For example, in my work, *power-to* underlies the *pragmatic perspective*: Instead of assuming that intelligent systems take power away from users, they help users to reach a certain goal and thus cope with their everyday tasks and practices. I have also presented *power-over* in form of the normative perspective which aims to strengthen users' power towards an intelligent systems based on ethics and legislation. This notion can also be found in work on fairness and accountability of intelligent systems (e.g., [15, 16]), where algorithms are claimed to be "exercising power over individuals or policies" [15].

The second category is the targeted *psychological component*, namely if users *feel* empowered through the system, or if empowerment is based on *knowing* or *doing*. Do we envision to make users *feel* empowered when interacting with intelligent systems? Do we want them to gain *knowledge* about intelligent systems? Or do we strive to give them means for achieving their goals using an intelligent system (*doing*)? My work targets *doing via knowing* – knowledge is the means to reach a goal, which consequently affects *doing*. This contrasts the normative perspective, which does not express a particular function of knowledge, but rather sees *knowing* as an end in itself. My work has not addressed the *feeling* component, which opens up interesting questions in the context of interaction with intelligent systems – what does it mean to feel empowered when interacting with an intelligent system? [P4] points to possible *emotional* effects of user support: Placebic explanations, that is, explanations that do not contain any information, may invoke trust similar to real explanations. Similarly, Vaccaro et al. [105] have found that control settings for social media feeds function as placebos and increase user satisfaction when included in the interface, whether they work or not. Likewise, Holliday et al. [46] investigated users' perception of control with and without explanations. Their results show that the perception of control did not differ between groups. Why not, then, give users the feeling of, for example, control and explanation, when it increases trust and user satisfaction? In particular if a system output has "inconsequential" effects [18]? It exceeds the scope of this discussion to elaborate on this highly ethical question, but in the context of my work, addressing *feeling* alone would not be empowering: Knowledge is crucial to get from one state to a goal state – support that does not address *knowing* thus merely presents a psychological tool to soothe users, but does not empower them to reach their goal.

The third category describes the *persistence of empowerment*. Empowerment is *transient* if it happens during system use, and *persistent* if it lasts beyond. Is user empowerment tied to the intelligent system, or can it persist beyond interaction? In my work, I assume that empowerment happens during system use to reach a goal and is thus *transient*. Again, this differs from the normative perspective, where legislation envisions *persistent* user empowerment which is decoupled from system use.

The last category is the *design mindset* with which solutions for support are approached. This mindset can be *participatory* or *expert*. Do we involve users in the process of designing interaction with intelligent systems, or do we find design solutions based on our expertise? Most of my work adopts a *participatory* mindset: [P6] presents a participatory process for system design, [P8] and [P9] assess user wishes for support, and the pragmatic perspective presented in [P7] likely implies user involvement. On the other hand, three contributions adopt an *expert mindset*, namely [P2], the normative perspective presented in [P7], and [P3], which summarises such expert mindsets in the field.

Finding an answer to these questions is, of course, dependent on the design context and the domain. However, in the face of arguments that technology may deprive us of fundamental abilities such as navigation [11], it seems that this reflection is needed to ensure that design for interaction with intelligent everyday systems is indeed empowering.

4.2 Concluding Remarks

In a way, to live life successfully we are all obliged [to] observe, experience, think about, and understand and act in our worlds, and we do so continuously.

Susan A. Lynham, in Advances in Developing Human Resources, 2016

It has been argued that in systems with "inconsequential" effects, there is no need to support user understanding [18]. While most intelligent everyday systems are indeed low-risk and will not have "significant consequences for unacceptable results" [18], this thesis has introduced a different perspective that is reflected by the above quote [72] – understanding can be considered a fundamental process to successfully navigate the world we live in. Intelligent everyday systems have become part of this world, whether they help us to communicate, to handle information, or to find our way. In a larger view, supporting user understanding thus means supporting a problem solving process that we follow in everything we do, and this includes interaction with intelligent everyday systems. Reaching a goal is in focus of this process, and helping users doing so is at the very heart of HCI [51, 82], where "the user has a task goal, and the application should support that goal as effectively as possible, both in terms of learning how to accomplish the goal and in actually accomplishing the goal" [51].

Therefore, I argue that, yes, there is a need to support user understanding also in systems whose effects might not be serious, but that are a part of daily life. The question is to find out *when* this support is needed, and *how* to address it to provide the "right" information, at the "right" time, in the "right" way [26] and best assist users in reaching their goal.

To this end, I have derived a research agenda consisting of five challenges and subchallenges that has emerged from my work. Some of these challenges have been addressed by publications presented in this thesis, others are left to be explored in the future. While intelligent everyday systems are in focus of this research agenda, several of the challenges and presented contributions might be applicable beyond this scope to intelligent systems in general or to similarly complex systems like everyday Internet of Things devices. For example, the method presented in [P8, P9] might be helpful for participatory approaches in these areas, given an adequate dataset. The framework for user support [P3] has been derived from a general survey of the literature about intelligent systems, and can therefore be used to advance conceptual clarity in a larger perspective on the field. Moreover, overarching challenges, such as the reflection about user empowerment, are of general concern in a world increasingly mediated by intelligent systems.

I hope that my work may inspire and guide future research to helping people live in such a world in better and in more informed ways, and to better ends and outcomes.

LIST OF FIGURES

1.1 Summary of the theoretical foundations of my work. I define interaction with intelligent everyday systems as *complex problem solving* in situations where users do not know how to reach their goal (2) (in contrast to (1), where the goal can be reached). Users then engage in *knowledge acquisition* and *knowledge application* – supporting this process is the overarching objective of this thesis. 4

3.1 Our method for assessing user need for support in daily interaction with intelligent everyday systems. It consists of two phases with three steps each to (1) elicit practical user problems, and (2) inform adequate solutions for support. 18

3.2 Our user-centric framework for defining and distinguishing different notions of user support as emerged through the lens of (implied) user questions. The three categories *user mindsets*, *user involvement*, and *knowledge outcomes* describe what users seek to know, how they acquire knowledge during interaction, and what kind of knowledge they acquire. 21

3.3 Our stage-based participatory design process for designing user support in real-world settings. The first three stages focus on *what to explain* in the system (information), the last two on *how to explain* (presentation). 24

List of Tables

2.1 Identified research problems and related guiding research questions of this thesis. . . 10

3.1 Overview of the primary contributions presented in this thesis. 13
3.2 Overview of the research methods used in the publications contributing to this thesis along with a short summary and the dataset size. 26

4.1 Clarification of my own and others' contribution on the publications included in this thesis. 58

GLOSSARY

Conceptual Research

Research "aimed at explaining previously unconnected phenomena occurring in interaction" [86].

Constructive Research

Research "aimed at producing understanding about the construction of an interactive artefact for some purpose in human use of computing" [86].

Complex Problem Solving

Complex problem solving concerns getting from a current state to a goal state in complex systems. It comprises two phases, *knowledge acquisition* and *knowledge application*. Together, these phases are passed through to reach the goal state, *see* complex system, knowledge acquisition & knowledge application.

Complex System

A complex system is defined in terms of five distinct characteristics [25]: (1) *Complexity* – the system entails a large number of variables), (2) *Interconnectedness* – the variables are interrelated and impact one another, (3) *Opacity* – how the variables are interrelated is not transparent, (4) *Dynamics* – the system state changes over time, dependent on or independent from an action taken, and (5) *Polytely* – the system serves multiple, ill-defined goals that might possibly conflict.

Empirical Research

Research "aimed at creating or elaborating descriptions of real-world phenomena related to human use of computing" [86].

Goal

The purpose of any interaction with an intelligent everyday systems. User goals may consists of subgoals and have different levels of abstraction, such as writing an email, getting updates from friends, or getting fit.

Goal State

see goal.

Intelligent Everyday System

An intelligent system embedded in a user's habitus that mediates everyday tasks and practices, *see* intelligent system.

Intelligent System

A complex system, *see* complex system.

Knowledge

Knowledge is required to get from a current state to a goal state in complex problem solving. Knowledge is thus always linked to a goal and can be implicit (not verbalisable) or explicit (verbalisable). Knowledge is acquired during *knowledge acquisition*, and applied during *knowledge application*, *see* knowledge acquisition & knowledge application.

Knowledge Acquisition

Knowledge acquisition is the first phase of complex problem solving, in which people acquire knowledge about the problem, *see* complex problem solving.

Knowledge Application

Knowledge application is the second phase of complex problem solving, in which people apply the knowledge acquired in the first phase to get to their goal state, and monitor the results, *see* complex problem solving.

Knowledge Outcomes

Knowledge outcomes describe *what kind of knowledge* users acquire through being supported in understanding the system. It can refer to system *outcome* (i.e., a specific system output), *process* (i.e., the system model and reasoning), *interaction* (i.e., how to control the system), and to *meta-knowledge* (i.e., information about the system beyond interaction). Knowledge can further have two qualities, *rigour* (i.e., completeness of information), or *relevance* (i.e., information required to reach a goal).

Normative Perspective

A normative perspective roots user support in ethical and legal reasoning, such as claiming a "user right to explanation".

Pragmatic Perspective

A pragmatic perspective focuses on user need for support and aims to incorporate real-world constraints and trade-offs.

Problem

A situation in which a person has a goal, but does not know how to reach it.

Problem Solving

The process of finding a way to get from a current state to a goal state, *see* complex problem solving.

Support, User Support

Assisting users in gaining the knowledge required to reach their goal based on the transfer of information between system and user.

Understanding

The knowledge users gain during knowledge acquisition and knowledge application, the two phases of complex problem solving, *see* complex problem solving, knowledge acquisition & knowledge application.

User

An *end-user* and a *layperson* who is familiar with using technology but not with technical details, and who does not operate within a particular domain.

User Involvement

User involvement describes *how users acquire knowledge* in interaction with a system. The nature of user involvement can be *active* (e.g., through system manipulation) or *passive* (e.g., through information presentation).

User Mindsets

User mindsets describe *what users seek to know* in interaction with a system, or the "cognitive orientation" of people [36], based on which they form a goal and plan successive actions towards reaching that goal. This mindset can be *utilitarian*, *interpretive*, and *critical*. A *utilitarian* mindset is directed towards usability and utility, an *interpretive* mindset carries a notion of user perceptions and experience, and a *critical* mindset stresses ethical and legal reflection about intelligent systems.

REFERENCES

[1] Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y. and Kankanhalli, M. 'Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda'. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. ACM, 2018, 582:1–582:18. DOI: 10.1145/3173574.3174156 (cited on pp. 11, 19, 32, 35).

[2] Alvarado, O. and Waern, A. 'Towards Algorithmic Experience: Initial Efforts for Social Media Contexts'. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. ACM, 2018, 286:1–286:12. DOI: 10.1145/3173574.3173860 (cited on p. 31).

[3] Amershi, S., Cakmak, M., Knox, W. B. and Kulesza, T. 'Power to the People: The Role of Humans in Interactive Machine Learning'. In: *AI Magazine* 35.4 (2014), p. 105. DOI: 10.1609/aimag.v35i4.2513 (cited on pp. 1, 3, 6, 9, 11, 21, 36).

[4] Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R. and Horvitz, E. 'Guidelines for Human-AI Interaction'. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. ACM, 2019, 3:1–3:13. DOI: 10.1145/3290605.3300233 (cited on pp. 9, 12).

[5] Ananny, M. and Crawford, K. 'Seeing without Knowing: Limitations of the Transparency Ideal and its Application to Algorithmic Accountability'. In: *New Media & Society* 20.3 (2018), pp. 973–989. DOI: 10.1177/1461444816676645 (cited on p. 32).

[6] Brignull, H. Dark Patterns. 2010. URL: https://www.darkpatterns.org/ (visited on 02/09/2019) (cited on p. 23).

[7] Bucher, T. 'The Algorithmic Imaginary: Exploring the Ordinary Affects of Facebook Algorithms'. In: *Information, Communication & Society* 20.1 (2017), pp. 30–44. DOI: 10.1080/1369118X.2016.1154086 (cited on pp. 1, 10, 20, 30, 31).

[8] Bunt, A., Lount, M. and Lauzon, C. 'Are Explanations Always Important?: A Study of Deployed, Low-cost Intelligent Interactive Systems'. In: *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces*. IUI '12. ACM, 2012, pp. 169–178. DOI: 10.1145/2166966.2166996 (cited on pp. 6, 10, 12, 16, 25, 31, 34).

[9] Burrell, J. 'How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms'. In: *Big Data & Society* 3.1 (2016), p. 2053951715622512. DOI: 10.1177/2053951715622512. eprint: https://doi.org/10.1177/2053951715622512 (cited on pp. 4, 12, 22, 33).

[10] Buschek, D. 'Behaviour-aware Mobile Touch Interfaces'. PhD thesis. 2018 (cited on p. 31).

[11] Carr, N. 'The Glass Cage: Where Automation is Taking Us'. Random House, 2015 (cited on p. 37).

[12] Chen, G. and Kotz, D. 'A Survey of Context-Aware Mobile Computing Research'. Tech. rep. 2000 (cited on p. 6).

[13] Cramer, H., Evers, V., Ramlal, S., Someren, M. van, Rutledge, L., Stash, N., Aroyo, L. and Wielinga, B. 'The Effects of Transparency on Trust in and Acceptance of a Content-based Art Recommender'. In: *User Modeling and User-adapted Interaction* 18.5 (2008), p. 455. DOI: 10.1007/s11257-008-9051-3 (cited on pp. 6, 9, 25, 34).

[14] Cramer, H., Wielinga, B., Ramlal, S., Evers, V., Rutledge, L. and Stash, N. 'The Effects of Transparency on Perceived and Actual Competence of a Content-based Recommender'. In: *CEUR Workshop Proceedings* 543 (2009), pp. 1–10. DOI: 10.1007/s11257-008-9051-3 (cited on p. 32).

[15] Diakopoulos, N. 'Accountability in Algorithmic Decision Making'. In: *Communications of the ACM* 59.2 (2016), pp. 56–62. DOI: 10.1145/2844110 (cited on pp. 11, 22, 36).

[16] Diakopoulos, N. 'Algorithmic Accountability Reporting: On the Investigation of Black Boxes'. In: *Tow Center for Digital Journalism Publications* (2014), pp. 1–33. DOI: 10.1002/ejoc.201200111. arXiv: arXiv:1011.1669v3 (cited on pp. 22, 36).

[17] Dinet, J. and Kitajima, M. ''Draw Me the Web': Impact of Mental Model of the Web on Information Search Performance of Young Users'. In: *Proceedings of the 23rd Conference on L'Interaction Homme-Machine*. IHM '11. ACM, 2011, 3:1–3:7. DOI: 10.1145/2044354.2044358 (cited on p. 16).

[18] Doshi-Velez, F. and Kim, B. 'Towards A Rigorous Science of Interpretable Machine Learning'. In: (2017). arXiv: 1702.08608 (cited on pp. 4, 11, 32, 37, 38).

[19] Dourish, P., Adler, A. and Smith, B. C. 'Organising User Interfaces around Reflective Accounts'. Tech. rep. 1996 (cited on p. 12).

[20] Dudley, J. J. and Kristensson, P. O. 'A Review of User Interface Design for Interactive Machine Learning'. In: *ACM Transactions on Interactive Intelligent Systems* 8.2 (2018), 8:1–8:37. DOI: 10.1145/3185517 (cited on pp. 1, 9, 12).

[21] Duncker, K. 'On Problem-solving'. In: *Psychological Monographs* 58.5 (1945), p. i (cited on p. 5).

[22] Edwards, L. and Veale, M. 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For'. In: *Duke Law & Technology Review* 16.1 (2017), pp. 18–84 (cited on pp. 11, 23).

[23] Eslami, M., Karahalios, K., Sandvig, C., Vaccaro, K., Rickman, A., Hamilton, K. and Kirlik, A. 'First I 'Like' It, then I Hide It: Folk Theories of Social Feeds'. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. ACM, 2016, pp. 2371–2382. DOI: 10.1145/2858036.2858494 (cited on p. 10).

[24] Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., Hamilton, K. and Sandvig, C. ''I Always Assumed that I Wasn't Really that Close to [Her]': Reasoning About Invisible Algorithms in News Feeds'. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15. ACM, 2015, pp. 153–162. DOI: 10.1145/2702123.2702556 (cited on p. 10).

[25] Fischer, A., Greiff, S. and Funke, J. 'The Process of Solving Complex Problems'. In: *The Journal of Problem Solving* 4.1 (2012). DOI: 10.7771/1932-6246.1118. eprint: https://docs.lib.purdue.edu/jps/vol4/iss1/3 (cited on pp. 1–3, 5, 43).

[26] Fischer, G. 'Context-aware Systems: The 'Right' Information, at the 'Right' Time, in the 'Right' Place, in the 'Right' Way, to the 'Right' Person'. In: *Proceedings of the International Working Conference on Advanced Visual Interfaces*. AVI '12. ACM, 2012, pp. 287–294. DOI: 10.1145/2254556.2254611 (cited on pp. 30, 38).

[27] Fischer, G. 'User Modeling in Human–Computer Interaction'. In: *User Modeling and User-Adapted Interaction* 11.1 (2001), pp. 65–86. DOI: 10.1023/A:1011145532042 (cited on p. 4).

[28] Fischer, G. and Lemke, A. C. 'Construction Kits and Design Environments: Steps Toward Human Problem-Domain Communication'. In: *Human–Computer Interaction* 3.3 (1987), pp. 179–222. DOI: 10.1207/s15327051hci0303_1. eprint: https://doi.org/10.1207/s15327051hci0303_1 (cited on p. 29).

[29] Fischer, G., Lemke, A. C., Mastaglio, T. and Morch, A. I. 'Using Critics to Empower Users'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '90. ACM, 1990, pp. 337–347. DOI: 10.1145/97243.97305 (cited on p. 36).

[30] Fischer, G. and Nakakoji, K. 'Beyond the Macho Approach of Artificial Intelligence: Empower Human Designers – Do Not Replace Them'. In: *Knowledge-Based Systems* 5.1 (1992), pp. 15–30. DOI: 10.1016/0950-7051(92)90021-7 (cited on p. 36).

[31] Fischer, G. and Stevens, C. 'Volunteering Information – Enhancing the Communication Capabilities of Knowledge-Based Systems'. In: *Human–Computer Interaction–INTERACT '87*. Ed. by H.-J. Bullinger and B. Shackel. North-Holland, 1987, pp. 965–971. DOI: https://doi.org/10.1016/B978-0-444-70304-0.50152-7 (cited on p. 31).

[32] Flanagan, J. C. 'The Critical Incident Technique'. In: *Psychological Bulletin* 51.4 (1954), pp. 327–360 (cited on p. 16).

[33] Frensch, P. A. and Funke, J. 'Definitions, Traditions, and a General Framework for Understanding Complex Problem Solving'. In: *Complex Problem Solving – The European Perspective*. 1995. DOI: 10.11588/HEIDOK.00008205 (cited on p. 3).

[34] Funke, J., Fischer, A. and Holt, D. V. 'Competencies for Complexity: Problem Solving in the Twenty-First Century'. In: *Assessment and Teaching of 21st Century Skills: Research and Applications*. Ed. by E. Care, P. Griffin and M. Wilson. Springer International Publishing, 2018, pp. 41–53. DOI: 10.1007/978-3-319-65368-6_3 (cited on p. 36).

[35] Gedikli, F., Jannach, D. and Ge, M. 'How Should I Explain? A Comparison of Different Explanation Types for Recommender Systems'. In: *International Journal of Human Computer Studies* (2014). DOI: 10.1016/j.ijhcs.2013.12.007 (cited on pp. 11, 21).

[36] Gollwitzer, P. M. 'Goal Achievement: The Role of Intentions'. In: 4.1 (1993), pp. 141–185. DOI: 10.1080/14792779343000059 (cited on pp. 20, 45).

[37] Goodman, B. and Flaxman, S. 'European Union Regulations on Algorithmic Decision-making and a 'Right to Explanation''. In: *AI Magazine* 38.3 (2017), pp. 50–57 (cited on p. 22).

[38] Gray, C. M., Kou, Y., Battles, B., Hoggatt, J. and Toombs, A. L. 'The Dark (Patterns) Side of UX Design'. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. ACM, 2018, 534:1–534:14. DOI: 10.1145/3173574.3174108 (cited on p. 23).

[39] Güss, D. C., Burger, M. L. and Dörner, D. 'The Role of Motivation in Complex Problem Solving'. In: *Frontiers in Psychology* 8 (2017), p. 851. DOI: 10.3389/fpsyg.2017.00851 (cited on p. 5).

[40] Guzman, E. and Maalej, W. 'How Do Users Like This Feature? A Fine Grained Sentiment Analysis of App Reviews'. In: *2014 IEEE 22nd International Requirements Engineering Conference*. RE '14. IEEE, 2014, pp. 153–162. DOI: 10.1109/RE.2014.6912257 (cited on p. 30).

[41] Harbach, M., Von Zezschwitz, E., Fichtner, A., De Luca, A. and Smith, M. 'It's a Hard Lock Life: A Field Study of Smartphone (Un)Locking Behavior and Risk Perception'. In: *Proceedings of the Tenth USENIX Conference on Usable Privacy and Security*. SOUPS '14. USENIX Association, 2014, pp. 213–230 (cited on p. 16).

[42] Harding, X. 'Google Has 7 Products With 1 Billion Users'. 2016. URL: https://www.popsci.com/google-has-7-products-with-1-billion-users/ (visited on 29/07/2019) (cited on p. 1).

[43] Harrison, S., Sengers, P. and Tatar, D. 'The Three Paradigms of HCI'. In: *Proceedings of alt.CHI*. 2007 (cited on p. 36).

[44] Herlocker, J. L., Konstan, J. A. and Riedl, J. 'Explaining Collaborative Filtering Recommendations'. In: *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*. CSCW '00. ACM, 2000, pp. 241–250. DOI: 10.1145/358916.358995 (cited on pp. 6, 9).

[45] Hildebrandt, M. 'The New Imbroglio – Living with Machine Algorithms'. In: *The Art of Ethics in the Information Society* (2016) (cited on p. 22).

[46] Holliday, D., Wilson, S. and Stumpf, S. 'The Effect of Explanations on Perceived Control and Behaviors in Intelligent Systems'. In: *CHI '13 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '13. ACM, 2013, pp. 181–186. DOI: 10.1145/2468356.2468389 (cited on pp. 9, 37).

[47] Höök, K. 'Steps to Take Before Intelligent User Interfaces Become Real'. In: *Interacting with Computers* 12.4 (2000), pp. 409–426. DOI: https://doi.org/10.1016/S0953-5438(99)00006-5 (cited on pp. 1, 9).

[48] Hornbæk, K. and Oulasvirta, A. 'What Is Interaction?' In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI '17. ACM, 2017, pp. 5040–5052. DOI: 10.1145/3025453.3025765 (cited on p. 6).

[49] Horvitz, E. 'Principles of Mixed-initiative User Interfaces'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '99. ACM, 1999, pp. 159–166. DOI: 10.1145/302979.303030 (cited on pp. 9, 31).

[50] Jhaver, S., Karpfen, Y. and Antin, J. 'Algorithmic Anxiety and Coping Strategies of AirBnB Hosts'. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. ACM, 2018, 421:1–421:12. DOI: 10.1145/3173574.3173995 (cited on p. 10).

[51] John, B. E. and Kieras, D. E. 'Using GOMS for User Interface Design and Evaluation: Which Technique?' In: *ACM Transactions on Computer-Human Interaction* 3.4 (1996), pp. 287–319. DOI: 10.1145/235833.236050 (cited on p. 38).

[52] Johnson, H. and Johnson, P. 'Explanation Facilities and Interactive Systems'. In: *Proceedings of the 1st International Conference on Intelligent User Interfaces*. IUI '93. ACM, 1993, pp. 159–166. DOI: 10.1145/169891.169951 (cited on p. 19).

[53] Kang, R., Dabbish, L., Fruchter, N. and Kiesler, S. ''My Data Just Goes Everywhere': User Mental Models of the Internet and Implications for Privacy and Security'. In: *Proceedings of the Eleventh USENIX Conference on Usable Privacy and Security*. SOUPS '15. USENIX Association, 2015, pp. 39–52 (cited on p. 16).

[54] Kapoor, A., Lee, B., Tan, D. and Horvitz, E. 'Interactive Optimization for Steering Machine Classification'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '10. ACM, 2010, pp. 1343–1352. DOI: 10.1145/1753326.1753529 (cited on p. 19).

[55] Kay, J. and Kummerfeld, B. 'Creating Personalized Systems That People Can Scrutinize and Control: Drivers, Principles and Experience'. In: *ACM Transactions on Interactive Intelligent Systems* 2.4 (2013), 24:1–24:42. DOI: 10.1145/2395123.2395129 (cited on pp. 11, 20).

[56] Kerstholt, J. H., Passenier, P. O., Houttuin, K. and Schuffel, H. 'The Effect of a Priori Probability and Complexity on Decision Making in a Supervisory Control Task'. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* (1996). DOI: 10.1518/001872096778940831 (cited on p. 33).

[57] Koch, J., Lucero, A., Hegemann, L. and Oulasvirta, A. 'May AI?: Design Ideation with Cooperative Contextual Bandits'. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. ACM, 2019, 633:1–633:12. DOI: 10.1145/3290605.3300863 (cited on pp. 6, 9, 29).

[58] Kocielnik, R., Amershi, S. and Bennett, P. N. 'Will You Accept an Imperfect AI?: Exploring Designs for Adjusting End-user Expectations of AI Systems'. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. ACM, 2019, 411:1–411:14. DOI: 10.1145/3290605.3300641 (cited on pp. 6, 9, 12, 29).

[59] Kulesza, T., Burnett, M., Wong, W.-K. and Stumpf, S. 'Principles of Explanatory Debugging to Personalize Interactive Machine Learning'. In: *Proceedings of the 20th International Conference on Intelligent User Interfaces*. IUI '15. ACM, 2015, pp. 126–137. DOI: 10.1145/2678025.2701399 (cited on pp. 6, 9, 10, 12, 19–22, 32).

[60] Kulesza, T., Stumpf, S., Burnett, M. and Kwan, I. 'Tell Me More?: The Effects of Mental Model Soundness on Personalizing an Intelligent Agent'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '12. ACM, 2012, pp. 1–10. DOI: 10.1145/2207676.2207678 (cited on p. 35).

[61] Kulesza, T., Stumpf, S., Burnett, M., Wong, W.-K., Riche, Y., Moore, T., Oberst, I., Shinsel, A. and McIntosh, K. 'Explanatory Debugging: Supporting End-User Debugging of Machine-Learned Programs'. In: *2010 IEEE Symposium on Visual Languages and Human-Centric Computing*. 2010, pp. 41–48. DOI: 10.1109/VLHCC.2010.15 (cited on p. 21).

[62] Kulesza, T., Stumpf, S., Burnett, M., Yang, S., Kwan, I. and Wong, W.-K. 'Too Much, Too Little, or Just Right? Ways Explanations Impact End Users' Mental Models'. In: *2013 IEEE Symposium on Visual Languages and Human Centric Computing*. IEEE, 2013, pp. 3–10. DOI: 10.1109/VLHCC.2013.6645235 (cited on pp. 6, 35).

[63] Kulesza, T., Stumpf, S., Wong, W.-K., Burnett, M. M., Perona, S., Ko, A. and Oberst, I. 'Why-oriented End-user Debugging of Naive Bayes Text Classification'. In: *ACM Transactions on Interactive Intelligent Systems* 1.1 (2011), 2:1–2:31. DOI: 10.1145/2030365.2030367 (cited on pp. 6, 10, 11).

[64] Langer, E. J., Blank, A. and Chanowitz, B. 'The Mindlessness of Ostensibly Thoughtful Action: The Role of 'Placebic' Information in Interpersonal Interaction.' In: *Journal of Personality and Social Psychology* 36.6 (1978), pp. 635–642. DOI: 10.1037/0022-3514.36.6.635 (cited on p. 23).

[65] Laudan, L. 'Progress and its Problems: Towards a Theory of Scientific Growth'. Vol. 282. University of California Press, 1978 (cited on p. 9).

[66] Lazar, A., Edasis, C. and Piper, A. 'Technology to Empower or Disempower: Design Tensions within the Context of Art Therapy'. In: *The Gerontologist* 56.Suppl_3 (2016), p. 660. DOI: 10.1093/geront/gnw162.2683. eprint: http://oup.prod.sis.lan/gerontologist/article-pdf/56/Suppl_3/660/7934583/gnw162.2683.pdf (cited on p. 36).

[67] Lee, J. D. and See, K. A. 'Trust in Automation: Designing for Appropriate Reliance'. In: *Human Factors* 46.1 (2004). PMID: 15151155, pp. 50–80. DOI: 10.1518/hfes.46.1.50_30392. eprint: https://doi.org/10.1518/hfes.46.1.50_30392 (cited on p. 16).

[68] Lim, B. Y. and Dey, A. K. 'Assessing Demand for Intelligibility in Context-aware Applications'. In: *Proceedings of the 11th International Conference on Ubiquitous Computing*. UbiComp '09. ACM, 2009, pp. 195–204. DOI: 10.1145/1620545.1620576 (cited on pp. 11, 20, 35).

[69] Lim, B. Y. and Dey, A. K. 'Design of an Intelligible Mobile Context-aware Application'. In: *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*. MobileHCI '11. ACM, 2011, pp. 157–166. DOI: 10.1145/2037373.2037399 (cited on pp. 6, 9, 12, 22, 25, 29).

[70] Lim, B. Y., Dey, A. K. and Avrahami, D. 'Why and Why Not Explanations Improve the Intelligibility of Context-aware Intelligent Systems'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '09. ACM, 2009, pp. 2119–2128. DOI: 10.1145/1518701.1519023 (cited on pp. 10, 11, 19, 29).

[71] Lipton, Z. C. 'The Mythos of Model Interpretability'. In: *Queue* 16.3 (2018), 30:31–30:57. DOI: 10.1145/3236386.3241340 (cited on p. 11).

[72] Lynham, S. A. 'The General Method of Theory-Building Research in Applied Disciplines'. In: *Advances in Developing Human Resources* 4.3 (2002), pp. 221–241. DOI: 10.1177/1523422302043002. arXiv: arXiv:1011.1669v3 (cited on p. 38).

[73] Maalej, W., Kurtanović, Z., Nabil, H. and Stanik, C. 'On the Automatic Classification of App Reviews'. In: *Requirements Engineering* 21.3 (2016), pp. 311–331. DOI: 10.1007/s00766-016-0251-9 (cited on p. 30).

[74] Mayer, R. E. and Wittrock, M. C. 'Problem Solving'. In: *Handbook of Educational Psychology*. 2006. DOI: 10.4324/9780203874790.ch13 (cited on p. 5).

[75] Moray, N. 'Intelligent Aids, Mental Models, and the Theory of Machines'. In: *International Journal of Man-Machine Studies* (1987). DOI: 10.1016/S0020-7373(87)80020-2 (cited on p. 33).

[76] Mulligan, D. K. and Bamberger, K. A. 'What Regulators Can Do to Advance Privacy Through Design'. In: *Communications of the ACM* 56.11 (2013), pp. 20–22. DOI: 10.1145/2527185 (cited on p. 35).

[77] Myers, B., Weitzman, D. A., Ko, A. J. and Chau, D. H. 'Answering Why and Why Not Questions in User Interfaces'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)* (2006). DOI: 10.1145/1124772.1124832 (cited on p. 34).

[78] Nguyen, A. T., Lease, M. and Wallace, B. C. 'Mash: Software Tools for Developing Interactive and Transparent Machine Learning Systems'. In: *Explainable Smart Systems Workshop at the 24th International Conference on Intelligent User Interfaces (IUI '19)*. 2019 (cited on p. 6).

[79] Nielsen, J. Mental Models. 2010. URL: www.nngroup.com/articles/mental-models/ (visited on 27/08/2019) (cited on pp. 16, 35).

[80] Nielsen, J. and Molich, R. 'Heuristic Evaluation of User Interfaces'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '90. ACM, 1990, pp. 249–256. DOI: 10.1145/97243.97281 (cited on p. 9).

[81] Niu, W. T. and Kay, J. 'PERSONAF: Framework for Personalised Ontological Reasoning in Pervasive Computing'. In: *User Modeling and User-Adapted Interaction* 20.1 (2010), pp. 1–40. DOI: 10.1007/s11257-009-9068-2 (cited on p. 31).

[82] Norman, D. A. 'The Psychology of Everyday Things'. Basic Books, 1988 (cited on pp. 6, 7, 9, 38).

[83] Norman, D. A. 'Some Observations on Mental Models'. In: *Mental Models* 7.112 (1983), pp. 7–14 (cited on pp. 5, 32, 35).

[84] Nunes, I. and Jannach, D. 'A Systematic Review and Taxonomy of Explanations in Decision Support and Recommender Systems'. In: *User Modeling and User-Adapted Interaction* 27.3 (2017), pp. 393–444. DOI: 10.1007/s11257-017-9195-0 (cited on pp. 11, 12, 32).

[85] Osman, M. 'Problem Solving: Understanding Complexity as Uncertainty'. In: *The Nature of Problem Solving: Using Research to Inspire 21st Century Learning*. OECD Publishing, 2017, pp. 47–57. DOI: 10.1787/9789264273955-5-en (cited on pp. 3, 5).

[86] Oulasvirta, A. and Hornbæk, K. 'HCI Research As Problem-Solving'. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. ACM, 2016, pp. 4956–4967. DOI: 10.1145/2858036.2858283 (cited on pp. 9, 10, 43).

[87] Picard, R. W. 'Affective Computing: Challenges'. In: *International Journal of Human-Computer Studies* 59.1 (2003). Applications of Affective Computing in Human-Computer Interaction, pp. 55–64. DOI: https://doi.org/10.1016/S1071-5819(03)00052-1 (cited on p. 31).

[88] Polanyi, M. 'Personal Knowledge: Towards a Post-critical Philosophy'. Routledge & Kegan Paul, 1958 (cited on p. 5).

[89] PrivacyPlan. Article 13 EU GDPR “Information to be Provided where Personal Data are Collected from the Data Subject”. 2017. URL: `http://www.privacy-regulation.eu/en/13.htm` (visited on 31/08/2019) (cited on pp. 12, 22, 33).

[90] PrivacyPlan. Article 22 EU GDPR “Automated Individual Decision Making, Including Profiling”. 2017. URL: `http://www.privacy-regulation.eu/en/22.htm` (visited on 31/08/2019) (cited on pp. 12, 33).

[91] Pu, P. and Chen, L. ‘Trust Building with Explanation Interfaces’. In: *Proceedings of the 11th International Conference on Intelligent User Interfaces*. IUI ’06. ACM, 2006, pp. 93–100. DOI: `10.1145/1111449.1111475` (cited on pp. 6, 9, 20, 29).

[92] Rader, E., Cotter, K. and Cho, J. ‘Explanations As Mechanisms for Supporting Algorithmic Transparency’. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI ’18. ACM, 2018, 103:1–103:13. DOI: `10.1145/3173574.3173677` (cited on pp. 6, 9–11, 20, 29, 32, 36).

[93] Rader, E. and Gray, R. ‘Understanding User Beliefs About Algorithmic Curation in the Facebook News Feed’. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI ’15. ACM, 2015, pp. 173–182. DOI: `10.1145/2702123.2702174` (cited on p. 10).

[94] Ribeiro, M. T., Singh, S. and Guestrin, C. ‘‘Why Should I Trust You?’: Explaining the Predictions of Any Classifier’. In: *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD ’16. ACM, 2016, pp. 1135–1144. DOI: `10.1145/2939672.2939778` (cited on pp. 11, 32, 33).

[95] Ribeiro, M. T., Singh, S. and Guestrin, C. ‘Anchors: High-precision Model-agnostic Explanations’. In: *AAAI Conference on Artificial Intelligence*. 2018 (cited on p. 33).

[96] Russell, S. J. and Norvig, P. ‘Artificial Intelligence: A Modern Approach’. Pearson Education Limited, 2016 (cited on p. 3).

[97] Sarkar, A. ‘Confidence, Command, Complexity: Metamodels for Structured Interaction with Machine Intelligence’. In: *Proceedings of the 26th Annual Conference of the Psychology of Programming Interest Group (PPIG 2015)*. 2015, pp. 23–36 (cited on pp. 1, 11).

[98] Sarkar, A. ‘Constructivist Design for Interactive Machine Learning’. In: *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. CHI EA ’16. ACM, 2016, pp. 1467–1475. DOI: `10.1145/2851581.2892547` (cited on p. 35).

[99] Schaub, F., Balebako, R., Durity, A. L. and Cranor, L. F. ‘A Design Space for Effective Privacy Notices’. In: *Proceedings of the Eleventh USENIX Conference on Usable Privacy and Security*. SOUPS ’15. USENIX Association, 2015, pp. 1–17 (cited on p. 35).

[100] Schoppek, W. ‘Examples, Rules, and Strategies in the Control of Dynamic Systems.’ In: *Cognitive Science Quarterly* 2.1 (2002), pp. 63–92 (cited on p. 6).

[101] Simon, H. A. ‘A Behavioral Model of Rational Choice’. In: *The Quarterly Journal of Economics* 69.1 (1955), pp. 99–118. DOI: `10.2307/1884852`. eprint: `/oup/backfile/content\_public/journal/qje/69/1/10.2307/1884852/2/69-1-99.pdf` (cited on p. 34).

[102] Simon, H. A. and Laird, J. E. ‘The Sciences of the Artificial’. The MIT Press. MIT Press, 2019 (cited on pp. 1, 3).

[103] Tintarev, N. and Masthoff, J. ‘A Survey of Explanations in Recommender Systems’. In: *2007 IEEE 23rd International Conference on Data Engineering Workshop*. IEEE, 2007, pp. 801–810. DOI: 10.1109/ICDEW.2007.4401070 (cited on p. 32).

[104] Tullio, J., Dey, A. K., Chalecki, J. and Fogarty, J. ‘How It Works: A Field Study of Non-technical Users Interacting with an Intelligent System’. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’07. ACM, 2007, pp. 31–40. DOI: 10.1145/1240624.1240630 (cited on pp. 9, 10, 35).

[105] Vaccaro, K., Huang, D., Eslami, M., Sandvig, C., Hamilton, K. and Karahalios, K. ‘The Illusion of Control: Placebo Effects of Control Settings’. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI ’18. ACM, 2018, 16:1–16:13. DOI: 10.1145/3173574.3173590 (cited on p. 37).

[106] Veinot, T. ‘Power to the Patient? A Critical Examination of Patient Empowerment Discourses’. In: *Configuring Health Consumers: Health Work and the Imperative of Personal Responsibility*. Ed. by R. Harris, N. Wathen and S. Wyatt. Palgrave Macmillan UK, 2010, pp. 30–41. DOI: 10.1057/9780230292543_3 (cited on p. 36).

[107] Wang, D., Yang, Q., Abdul, A. and Lim, B. Y. ‘Designing Theory-Driven User-Centric Explainable AI’. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI ’19. ACM, 2019, 601:1–601:15. DOI: 10.1145/3290605.3300831 (cited on p. 33).

[108] Weller, A. ‘Challenges for Transparency’. In: *CoRR* abs/1708.01870 (2017). arXiv: 1708.01870 (cited on p. 23).

[109] Wenke, D., Frensch, P. A. and Funke, J. ‘Complex Problem Solving and Intelligence: Empirical Relation and Causal Direction.’ In: *Cognition and Intelligence: Identifying the Mechanisms of the Mind.* Cambridge University Press, 2005, pp. 160–187 (cited on p. 5).

[110] Wiegand, G., Schmidmaier, M., Weber, T., Liu, Y. and Hussmann, H. ‘I Drive – You Trust: Explaining Driving Behavior Of Autonomous Cars’. In: *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI EA ’19. ACM, 2019, LBW0163:1–LBW0163:6. DOI: 10.1145/3290607.3312817 (cited on pp. 24, 34).

[111] Wobbrock, J. O. and Kientz, J. A. ‘Research Contributions in Human-computer Interaction’. In: *Interactions* 23.3 (2016), pp. 38–44. DOI: 10.1145/2907069 (cited on p. 13).

[112] Yee, K.-P. ‘Aligning Security and Usability’. In: *IEEE Security Privacy* 2.5 (2004), pp. 48–55. DOI: 10.1109/MSP.2004.64 (cited on p. 35).

[113] Zerilli, J., Knott, A., Maclaurin, J. and Gavaghan, C. ‘Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?’ In: *Philosophy & Technology* (2018). DOI: 10.1007/s13347-018-0330-6 (cited on p. 23).

Original Contributing Publications

The research that builds up this thesis would not have been possible without the excellent support of my supervisor, colleagues and students. Table 4.1 clarifies my own and others' contributions to each project and publication.

	My contribution	Others' contribution
[P1]	I came up with the concept idea together with Sarah Völkel and Daniel Buschek. I contributed substantially to the background part on *dark design patterns* and to the development and description of the actual patterns and visualisations.	Michael Chromik contributed substantially to the overall position paper as the leading author. Sarah Völkel and Daniel Buschek contributed substantially to the concept idea and the development and description of the presented patterns. In addition, Daniel Buschek contributed to the visualisations in the publication.
[P2]	I came up with the original research idea. I supervised the development of the prototype and the implementation of the user study. I wrote the whole text for the published version of the paper and contributed substantially to the quantitative data analysis.	Charlotte Anlauff, Tim Ordenewitz and Martin Zürn developed the prototype and conducted the user study. They contributed substantially to a first draft of the paper and the qualitative and quantitative data analysis. Heinrich Hußmann provided feedback on the work and the publication.
[P3]	I developed the research idea together with Daniel Buschek. I contributed substantially to the assembling of the paper set, the paper coding and analysis, and to the actual submission as the leading author.	Daniel Buschek contributed equally to the research idea, the assembling of the paper set, the paper coding and analysis, and the resulting submission. Heinrich Hußmann provided feedback on the work and the publication.
[P4]	I came up with the original research idea. I supervised the development of the prototype and implementation of the user study. I was the leading author of the resulting publication.	Alexander Kremer finalised the actual design of the prototype and contributed substantially to the scenario used in the user study. Daniel Buschek contributed to the discussion part of the resulting publication and provided overall feedback on the work. Heinrich Hußmann provided feedback on the work and the publication.
[P5]	I developed the research idea in collaboration with Mohamed Khamis, Emanuel von Zezschwitz and Florian Alt. I worked out and pre-tested the study method together with Mohamed Khamis, with whom I also did the qualitative data analysis, and was the leading author of the resulting publication.	Mohamed Khamis contributed equally to the development and pre-testing of the study method, the qualitative data analysis and the resulting publication. Emanuel von Zezschwitz contributed the background section of the publication and provided overall feedback. Florian Alt contributed substantially to the introduction and provided overall feedback. Heinrich Hußmann provided feedback on the work and the publication.
[P6]	I supervised the development and implementation of the concept and prototype and the study design during Hanna Schneider's three-month research stay abroad. Hanna Schneider and I contributed equally to the formalisation of the design process presented in the paper. I was the leading author of the resulting publication.	Hanna Schneider came up with the original research idea together with Mark Bilandzic. Mareike Haug and Julian Fazekas-Con designed and implemented the prototype and did the user studies involved in the project. Mark Bilandzic, Peter Just, Natalia Patiño Pérez and Renato Pereira of the industry partner Freelectics contributed in numerous discussions and brainstormings with their domain knowledge and expertise. Mark Bilandzic contributed a section about Freeletics to the paper. Hanna Schneider contributed substantially to the resulting publication. Heinrich Hußmann provided feedback on the work and the publication.
[P7]	I developed the idea and the concept of this publication together with Hanna Schneider and Daniel Buschek. Each of us contributed a section to the position paper.	
[P8, P9]	[P8] resulted from a journal invitation based on [P9]. I came up with the original research idea. Together with my colleague Sarah Völkel, I supervised Sophia Cook during the research project. Sarah Völkel and I equally contributed to the data analysis, namely the topic modelling and all manual coding steps. We developed and implemented the online survey and analysed the survey data. I was the leading author of both publications.	Sarah Völkel contributed equally to the data analysis and to the resulting publications. Daniel Buschek contributed substantially to the development of the presented method and to the written publications. Sophia Cook provided the data set on which our analysis is based and did the sentiment analysis. Heinrich Hußmann provided feedback on the work and the publications.
[P10]	I contributed to the development of the framework categories. I helped to code papers according to these categories and contributed substantially to the resulting publication.	Hanna Schneider came up with the original research idea. She also provided the data set and did the first coding iteration as well as the systematic coding after the framework categories had been defined. She was the leading author of this publication. Daniel Ullrich contributed to the development of the framework categories and to the discussion in the paper. Andreas Butz supervised the project and edited the paper for clarity and readability.
[P11]	I contributed to the user studies conducted in Germany and provided a section about my perspective and learnings in the resulting publication.	My colleague Hanna Schneider planned and conducted the field trip together with Florian Lachner. Hanna Schneider was also the leading author of the resulting publication. Florian Lachner contributed significantly to the planning and implementation of the field trip. All field trip participants contributed to the data collection. All co-authors contributed to the publication with writing about their perspective on the field trip or their expertise.

Table 4.1: Clarification of my own and others' contribution on the publications included in this thesis.

Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12.07.11, § 8, Abs. 2 Pkt. 5)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig und ohne unerlaubte Beihilfe angefertigt wurde.

München, den 8. Juli 2019

Malin Eiband