Progress in IS

Jan Ole Berndt
Marcus Grum
Kyra Göbel
Annette Kluge *Editors*

Intentional Forgetting with Intelligent Systems

Developing Dynamic Capabilities for Work and Organizations

OPEN ACCESS

Springer

Progress in IS

Progress in IS encompasses the various areas of Information Systems in theory and practice, presenting cutting-edge advances in the field. It is aimed especially at researchers, doctoral students, and advanced practitioners. The series features both research monographs, edited volumes, and conference proceedings that make substantial contributions to our state of knowledge and handbooks and other edited volumes, in which a team of experts is organized by one or more leading authorities to write individual chapters on various aspects of the topic. Individual volumes in this series are supported by a minimum of two external reviews.

The Series is SCOPUS-indexed.

Jan Ole Berndt · Marcus Grum · Kyra Göbel ·
Annette Kluge
Editors

Intentional Forgetting with Intelligent Systems

Developing Dynamic Capabilities for Work and Organizations

Springer

Editors
Jan Ole Berndt
German Research Centre for Artificial Intelligence (DFKI)
Trier, Germany

Marcus Grum
University of Potsdam
Potsdam, Germany

Kyra Göbel
Friedrich-Alexander-University Erlangen-Nürnberg
Erlangen, Germany

Annette Kluge
Ruhr University Bochum
Bochum, Germany

ISSN 2196-8705 ISSN 2196-8713 (electronic)
Progress in IS
ISBN 978-3-032-17620-2 ISBN 978-3-032-17621-9 (eBook)
https://doi.org/10.1007/978-3-032-17621-9

This work was supported by Deutsche Forschungsgemeinschaft (KL2207/5-2).

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The edited volume explores the concept of intentional forgetting within the context of organizations and intelligent systems. To begin, it is instructive to reflect briefly on the etymology of the key terms. The verb “to forget” originates from the Old English “forgietan,” combining the prefix “for-” (in the sense of “away” or “amiss”) with “gietan” (“to grasp”). The literal meaning is thus “to un-grasp,” indicating a loss of mental hold or access. Cognates in other Germanic languages, including Old High German (“firgezzan”) and modern German (“vergessen”), support this way of interpretation.[1]

But this book is not about forgetting – it is in particular about intentional forgetting. “Intent” derives from the Latin “intentio,” referring to “a stretching out” or “effort,” rooted in the verb “intendere”—“to direct one’s attention.” In Middle English, it represents purpose, aim, or mental effort.[1] In its contemporary usage, “intentional” implies a directed and controlled cognitive or behavioral act.

These origins point to an important distinction: forgetting is typically associated with loss or decay, while intent refers to deliberate control. The concept of intentional forgetting thus denotes a deliberate reduction of memory accessibility, rather than passive decay or malfunction. The chapters in this book contribute to the empirical research on intentional forgetting- in a special context, namely in organizations. In an 8-year research project, funded by the DFG (SPP 1921 “Intentional forgetting in organizations”), researchers form several discipline in psychology and computer science worked on answers to the question, how organizations can be supported in their effort to intentionally forget.

[1] https://www.etymonline.com/word/forget

The Cognitive View: Cognitive Foundations of Intentional Forgetting

Recent research in cognitive neuroscience (Anderson & Hulbert, 2021) provides empirical evidence for the active regulation of memory accessibility. This line of research frames forgetting not as a failure, but as a goal-directed process that supports adaptive behavior. Central to this idea is inhibitory control—the capacity to suppress or modulate internal processes such as memory retrieval. This control serves to resolve interference, prevent the intrusion of irrelevant or unwanted content, and support goal-oriented cognition.

Three forms of such inhibition have been identified:

1. Memory Inhibition: Targeted suppression of specific memory traces, e.g., during selective retrieval (leading to retrieval-induced forgetting) or intentional stopping of intrusive thoughts (suppression-induced forgetting).
2. Process Inhibition: Disruption of encoding or consolidation processes, which can lead to more generalized effects, including temporary amnesic states (e.g., amnesic shadow phenomena).
3. Context Inhibition: Suppression of mental contexts that serve as retrieval cues, reducing access to context-dependent memories without deleting them (e.g., in list-method directed forgetting paradigms).

These inhibitory mechanisms can operate dynamically and flexibly, often engaging in interplay depending on task demands and cognitive goals. They are not mutually exclusive but can be combined to optimize adaptive forgetting in complex environments. Moreover, the efficacy of inhibitory control varies across individuals, situations, and can be influenced by factors such as attention or motivation, highlighting the intricate regulation of forgetting as an active and adaptive cognitive function. The prefrontal cortex, particularly the lateral regions, has been identified as a key neural substrate for these inhibitory functions.

The Organizational Vie: Organizational Relevance of Intentional Forgetting

In organizational contexts, forgetting has received comparatively little systematic attention (Mull et al., 2023). Most theories and practices focus on organizational learning, knowledge acquisition, and memory-building. However, contemporary organizations are increasingly confronted with a continuous accumulation of information—supported by digital systems and data infrastructures—without corresponding mechanisms for information reduction. In this respect, intentional forgetting becomes relevant as a complementary organizational process. It can contribute to reducing complexity, freeing up cognitive and technical capacities, and supporting adaptation under dynamic conditions. This volume proposes that forgetting, when intentionally designed and operationalized, can support the development of dynamic capabilities—that is, the ability of an organization to adapt and reconfigure its resources and routines in response to environmental change.

Unlike in humans, forgetting does not occur automatically in organizations. Technological infrastructures preserve information indefinitely, unless active deletion or abstraction strategies are implemented. The design of forgetting mechanisms in organizations therefore requires conceptual, technical, and practical frameworks that go beyond existing theories. In this volume, Kluge and Roling (2026) explore organizational unlearning and intentional forgetting as crucial for adaptation and innovation. They propose that unlearning is a radical, strategic process linked to double-loop learning, focused on questioning mindsets and enabling radical innovation during the ideation phase. It involves discarding outdated routines and knowledge. Intentional forgetting, in contrast, is an incremental, operational process tied to single-loop learning, refining processes and behaviors during the evaluation phase.

Psychology & Computer Science Collaboration to Enhance Extraordinary Research in Intentional Forgetting in Socio-Digital Systems

Psychology and computer science—particularly research in artificial intelligence—offer numerous points of collaboration on the topic of forgetting. Psychological models of forgetting provide a valuable conceptual foundation that can be transferred to AI systems (Ellwart et al., 2018; Ellwart & Kluge, 2019). This interdisciplinary exchange holds potential for the implementation of functional mechanisms of human forgetting into digital systems, opening new opportunities for the development and application of intelligent technologies.

A central area of collaboration is the transfer of psychological concepts to AI systems. psychological models of individual forgetting offer insights into how learning systems can adapt to environmental change—for example, when parts of a trained knowledge base become outdated or uncertain. Such transfer can inform non-monotonic reasoning and belief revision in AI systems. Applications include robotics, navigation, or the coordination and self-organization of logistic networks, where forgetting curves can be implemented. Additionally, these psychological models offer approaches to information filtering and data reduction, which can help address challenges related to algorithmic complexity in planning and reasoning. For instance, Beierle et al. (2026) identify kinds of forgetting, axiomatically formalize belief change operators involving forgetting, and elaborate their interrelationships. Their developed general abstract framework is instantiated with conditional beliefs and realizations of forgetting operators employing ranking functions.

Team-level models such as transactive memory systems (TMS) can also be applied to multi-agent systems. These models emphasize the distribution and specialization of knowledge within teams and can enhance cognitive efficiency in agent-based architectures. In the context of future work, where humans increasingly interact with artificial agents, there is growing support for integrating human-like cognitive processes—such as memory decay—into digital systems. This is intended to increase user acceptance of long-term partnerships between humans and machines. Practical examples include systems that mimic human forgetting curves to remove unused desktop icons, identify and explain when digital information becomes irrelevant, and support product development by reducing unnecessary constraints. From a technical perspective, such applications require advanced AI methods in knowledge representation, pattern recognition, belief revision, and reasoning.

For example, Berndt et al. (2026) argue that organizationally successful forgetting means striking a balance between these opposing poles: whether a team assigns its tasks to predominantly specialist members or whether all team members are generalists having the same responsibilities. Their chapter introduces an agent-based social simulation approach for designing and evaluating the distribution of responsibilities in teams. Their approach helps identify which amount of forgetting is required and feasible to reduce cognitive workload and avoid mistakes. Berndt et al.'s chapter integrates a psychological measure for workload as well as heuristic decision-making into the Beliefs-Desires-Intentions agent architecture as a technological foundation.

An example of such interdisciplinary implementation is the development of assistive systems that support intentional deletion in everyday digital work. These systems operationalize psychological models of forgetting and inhibitory control to help users manage redundant or outdated information. By combining explainable AI with adaptive classification, they aim to reduce cognitive effort and foster deliberate disengagement from irrelevant content—particularly in contexts where deletion decisions are ambivalent or cognitively demanding. As an interdisciplinary collaboration between artificial intelligence (AI) research and work psychology, Schmid et al. in this volume developed and evaluated an intelligent assistant system to support the identification of irrelevant files. The AI-assistant Dare2Del combines explicit knowledge representation and machine learning. To allow for human control and oversight, Dare2Del relies on explanatory interactive methods. Julek et al., in this volume, present a compact overview of a decade of research on self-organizing personal knowledge assistants in evolving corporate memories inspired by real-world problems and resulting in Managed Forgetting and Self-organizing Context Spaces (cSpaces) as a novel approach to Personal Information Management (PIM) and Knowledge Work (KW). Finally, Reiner et al., in this volume, offer a scientifically grounded definition of trust relationships, providing a foundation for classifying current and future research on trust. The proposed model distinguishes three actors and five trust directions between them.

Another important area is the configuration of cyber-physical-social systems. A key task for both psychologists and computer scientists is to design these systems in alignment with principles of human-centered work design. Knowledge of human information processing and its relevance for enabling engaging, meaningful, and autonomy-supportive work must be integrated into the design of socio-digital teams. For instance, first attempts have implemented these cognitive foundations into artificial neural networks (Grum, 2023). Here, neural structures are prepared to take ownership of tasks of a higher cognitive level. Their output is used to inhibit memories, traces, processes, and artificial mental contexts of subordinate structures that would have served as retrieval cues. Additionally, Grum (2026) developed an AI Interaction Instrument (A3I) that enables interaction modelling of learning and forgetting human workers with learning and forgetting AIs. He illustrates in initial experiments with a focus on AI trust, AI frustration, and variations in AI reliability in an Industry 4.0 context. In a design science-oriented manner, the A3I was developed as a research tool that enables the quantification of the influence of forgetful AIs on forgetful humans in experiments using empirical data analysis tools.

In the chapter by Schüffler and Kluge (2026), the nature of switching costs in forgetting affordances in cyber-physical systems is investigated, highlighting their impact on efficiency and psychological well-being. A newly developed Subjective Switching Cost Scale is introduced for measuring workers adaptation in a controlled production environment. They showed that subjective switching costs correlate with objective measures, particularly in tasks that remain unchanged, suggesting that awareness of change influences performance outcomes.

Methodological synergies also arise from a interdisciplinary collaboration. Psychological research methods—particularly empirical and experimental approaches—can be supported, replicated, or extended through simulations with AI models. This makes it possible to overcome common limitations of psychological studies (e.g., sample size, duration, complexity), contributing to model testing and theory development. In this respect, Weyers et al. (2026) introduces a user-centred approach to designing persuasive systems that help people change their work-related habits. The proposed development pipeline transforms user insights into formal models, making it possible to create personalized and adaptive interventions. The process starts with collecting data through interviews, where users describe their habits and desired changes. These insights are turned into semi-formal models using "Business Process Model and Notation" (BPMN) models and then refined into formal reference nets, a special type of Petri nets.

Objectives and Contributions of the Volume

With this book, we emphasize the need for interdisciplinary collaboration, arguing that the development and implementation of intentional forgetting in socio-digital work systems requires expertise from cognitive psychology, team and organizational research, human-automation interaction, and computer science.

This edited volume addresses intentional forgetting from an interdisciplinary perspective, integrating insights from psychology, cognitive science, information systems, computer science, and organizational theory. The contributing chapters pursue three main objectives:

- To analyze mechanisms of forgetting in humans and explore their relevance for organizational and technical systems.

- To connect organizational theory—particularly perspectives on the alignment of processes, structures, people, and technologies—with the design of information systems capable of intentional forgetting.
- To present methodological approaches and transfer strategies for managing the balance between information accumulation and reduction in organizational contexts.

The book is situated within a broader research agenda that treats human forgetting not as a cognitive deficiency, but as an adaptive function valuable in terms of modern knowledge management. This perspective is transferred to socio-technical systems, where forgetting is conceived as a deliberate and functional capability, rather than a loss or failure. It respects humans as well as technical systems as knowledge carriers. In doing so, the book proposes a shift in perspective: away from viewing human cognition as inferior to machine memory and toward viewing human forgetting as a model for designing organizational systems that are capable of intentional information reduction and adaptive change. It contributes to considering the team of human and AI-based systems in a symbiotic relationship that needs to be monitored, controlled, managed, and brought into action.

The contributions in this volume underscore the relevance of intentional forgetting as a functional principle for adaptive behavior in individuals, organizations, and intelligent systems. Forgetting is not merely the absence of remembering—it should be treated as an actively shaped process that helps reduce complexity, maintain focus, and support reorientation in dynamic environments.

As socio-digital systems continue to evolve, the challenge will be not only to accumulate and preserve knowledge, but also to implement selective and purposeful forgetting. This volume lays a foundation for further interdisciplinary work in this emerging field and aims to stimulate new approaches to designing systems that are capable of both remembering wisely—and forgetting well.

Trier, Germany	Jan Ole Berndt
Potsdam, Germany	Marcus Grum
Erlangen, Germany	Kyra Göbel
Bochum, Germany	Annette Kluge

References

Anderson, M. C., & Hulbert, J. C. (2021). Active forgetting: Adaptation of memory by prefrontal control. *Annual Review of Psychology*, *72*, 1–36. https://doi.org/10.1146/annurev-psych-072720-094140

Ellwart, T. & Kluge, A. (2019). Psychological Perspectives on Intentional Forgetting. An Overview of Concepts and Literature. *KI — Künstliche Intelligenz* (German Journal on Artificial Intelligence., March 2019, Volume 33, Issue 1, pp 79–84 https://doi.org/10.1007/s13218-018-00571-0

Ellwart, T., Ulfert, A.-S., Antoni, C., Becker, J., Frings, C., Göbel, K., Hertel, G., Kluge, A., Meeßen, S., Niessen, C., Nohe, C., Riehle, D., Runge, Y., Schmid, U., Schüffler, A., Siebers, M., Sonnentag, S., Tempel, T., Thielsch, M., Wehrt, W. (2020) Forgetting in future work systems: System characteristics and user-related psychological consequences on emotion, cognition, and behaviors. *AES Transactions on Enterprise Systems*, Vol 4 No 1 (2019): Special Issue: Intentional Forgetting in Organizations and Work Settings, https://www.aes-journal.com/index.php/ais-tes/article/view/16/14, https://doi.org/10.30844/ais-tes.v4i1.16

Grum, M. (2023). Learning representations by crystallized back-propagating errors. In International Conference on Artificial Intelligence and Soft Computing (pp. 78–100). Cham: Springer Nature Switzerland.

Mull, M., Duffy, C. and Silberman, D. (2023), "Forgetting to learn and learning to forget: the call for organizational unlearning", *European Journal of Training and Development*, Vol. 47 No. 5/6, pp. 586–598. https://doi.org/10.1108/EJTD-10-2021-0162

Contents

Contributors

Conny H. Antoni Business, Work and Organizational Psychology, Trier University, Trier, Germany

Jörg Becker European Research Center for Information Systems, University of Münster, Münster, Germany

Christoph Beierle Faculty of Mathematics and Computer Science, FernUniversität in Hagen, Hagen, Germany

Jan Ole Berndt German Research Center for Artificial Intelligence (DFKI), Trier, Germany

Daniel Brand Predictive Analytics, TU Chemnitz, Chemnitz, Germany

Andreas Dengel Smart Data and Knowledge Services Department, German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany; Department of Computer Science, RPTU Kaiserslautern-Landau, Kaiserslautern, Germany

Thomas Ellwart Business, Work and Organizational Psychology, Trier University, Trier, Germany

Nico Feld Computer Science, Human-Computer Interaction, Trier University Universitätsring 15, Trier, Germany

Christian Frings Department of Cognitive Psychology, University of Trier, Trier, Germany

Benedikt Graf Work and Organizational Psychology, Göttingen University, Göttingen, Germany

Marcus Grum University of Potsdam, Potsdam, Germany

Kyra Göbel Friedrich-Alexander-University Erlangen, Erlangen, Germany

Guido Hertel Organizational and Business Psychology, University of Münster, Münster, Germany

Diana Howey Department of Computer Science, TU Dortmund University, Dortmund, Germany

Christian Jilek Smart Data and Knowledge Services Department, German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany

Gabriele Kern-Isberner Department of Computer Science, TU Dortmund University, Dortmund, Germany

Annette Kluge Chair of Industrial, Organizational and Business Psychology, Ruhr University Bochum, Bochum, Germany

Heiko Maus Smart Data and Knowledge Services Department, German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany

Lea Sophie Müller Organizational and Business Psychology, University of Münster, Münster, Germany

Claudia Niederée L3S Research Center, Hannover, Germany

Cornelia Niessen Friedrich-Alexander-University Erlangen, Erlangen, Germany

Marco Ragni Predictive Analytics, TU Chemnitz, Chemnitz, Germany

Sebastian Reiners European Research Center for Information Systems, University of Münster, Münster, Germany

Lukas Reuter Center for Informatics Research and Technology, Trier University, Trier, Germany

Wiebke M. Roling Ruhr University Bochum, Bochum, Germany

Kai Sauerwald Faculty of Mathematics and Computer Science, FernUniversität in Hagen, Hagen, Germany

Ute Schmid University of Bamberg, Bamberg, Germany

Arnulf Schueffler Chair of Industrial, Organizational and Business Psychology, Ruhr University Bochum, Bochum, Germany

Tobias Tempel Ludwigsburg University of Education, Ludwigsburg, Germany

Ingo J. Timm German Research Center for Artificial Intelligence (DFKI), Trier, Germany;
Center for Informatics Research and Technology, Trier University, Trier, Germany

Sara Todorovikj Predictive Analytics, TU Chemnitz, Chemnitz, Germany

Anna-Sophie Ulfert-Blank Industrial Engineering and Innovation Sciences, Human Performance Management Group, Eindhoven University of Technology, Eindhoven, The Netherlands

Benjamin Weyers Computer Science, Human-Computer Interaction, Trier University Universitätsring 15, Trier, Germany

Olivia Wilczewski University of Bamberg, Bamberg, Germany

Perspectives on the Relationship Between Unlearning and Intentional Forgetting in Organizations

Annette Kluge and Wiebke M. Roling

1 Introduction: Organizational Adaptation Requirement

> You cannot dig a hole in a different place by digging the same hole deeper. (de Bono, 2014)

The literature on organizational theory and science demonstrates that organizations can evolve, develop, and learn over time. They can grow, shrink, rise, decline, or even cease to exist. Understood as open systems, organizations interact with their environment and ideally strive to align with environmental requirements and affordances. These factors influence aspects such as the organization's purpose, its strategy, the technology it employs, and the structures and processes used to enhance efficiency.

Efficiency is a very relevant and inward-looking goal for organizations. In organizational learning, efficiency is enhanced by single-loop learning ("Doing the things right," Argyris & Schön, 1978). This describes, for example, in the sense of total quality management or the continuous improvement process (e.g., the Deming Cycle), perfecting one's own processes along the value chain more and more. In what you do and what you offer, you become better and better, more perfect, faster, and more error-free. The aim of single-loop learning is to leverage efficiency potential. But what do organizations do if they offer perfect products or services, but the market and customers no longer accept them? Or when the global community or parts of it realize that the way these products are produced is exploiting or polluting resources irresponsibly in the long term? Then, organizations have to ask whether they are doing the right things (double-loop learning: "Doing the right things," Argyris & Schön,

A. Kluge (✉) · W. M. Roling
Ruhr University Bochum, Bochum, Germany
e-mail: annette.kluge@rub.de

W. M. Roling
e-mail: wiebke.roling@rub.de

J. O. Berndt et al. (eds.), *Intentional Forgetting with Intelligent Systems*, Progress in IS,
https://doi.org/10.1007/978-3-032-17621-9_1

1978). Double-loop learning adds the questioning of action reasons and underlying mindsets (Argyris & Schön, 1978). As a result, organizations may realize that digging the hole deeper will not be successful in the medium and long term, and instead decide to change their own products, services, and processes (digging the hole elsewhere). This transition phase between single- and double-loop learning is addressed by the concepts of unlearning and intentional forgetting, as organizations have to move from the idea to dig the hole elsewhere to an implementable solution and the realignment of the organization.

Over the past decades, scholars from various perspectives in organizational theory and science, such as Tsang and Zahra (2008), de Holan and Phillips (2004), and others, have emphasized the importance of organizational unlearning and forgetting as necessary processes for successful development and achieving alignment between external demands and internal capabilities (Kluge & Gronau, 2018). For example, markets with competitive intensity and turbulence (Huang et al., 2018), as well as newly introduced EU regulations, available new materials and technologies force organizations to unlearn and forget (Kluge, 2023; Kluge et al., 2019). Unlearning results in cross-boundary innovation, positive organizational development and performance, and develops knowledge creation capability in innovative and innovation performance (Kluge, 2023).

2 Theoretical Reflections on the Relationship Between Unlearning and Intentional Forgetting

It is increasingly recognized that, for organizations to successfully learn and adapt, unlearning and forgetting are essential alongside knowledge acquisition and dissemination (Fiol & O'Connor, 2017b, 2017a; Grisold et al., 2017; Klammer & Gueldenberg, 2019; Morais-Storz & Nguyen, 2017). The term "organizational unlearning" was introduced almost simultaneously with the concept of "organizational learning" in management research (Hedberg, 1981; Howells & Scholderer, 2016; Nystrom & Starbuck, 1984). Since then, unlearning changed its meaning going beyond replacing outdated routines to a radical form of change that is particularly relevant when organizations reflect their business models and core processes in terms of asking "are we doing the right things" (double-loop learning). Unlearning is in terms of creativity and transformation of an organization's core aspects a more radical form (Mumford et al., 2012). It can be understood as an antecedence for radical innovation (Leal-Rodríguez et al., 2015; Sharma & Lenka, 2022b, 2014; Yang et al., 2022a).

In the early days, unlearning has been defined as the process of discarding and replacing outdated routines (Huber, 1991; Tsang & Zahra, 2008).

The concept of "organizational forgetting" entered the business and management literature somewhat later, around the turn of the century (e.g., Argote, 2013; Easterby-Smith & Lyles, 2003, 2011; de Holan & Phillips, 2004; de Holan et al., 2004; Martin de Holan, 2011). In analogy to the individual human brain, forgetting is not a malfunction of human information processing (Roediger III et al., 2010; Wixted, 2004), but rather a crucial adaptive function that enables overwriting, suppressing, and sorting out outdated information (Bjork et al., 1998). Human memory does not erase obsolete knowledge but can suppress it and make it inaccessible. When the environment changes, adaptability becomes essential, requiring the abandonment of previous goals to focus on those currently relevant (Altmann & Gray, 2002; Roediger III et al., 2010). Forgetting impedes the retrieval of obsolete knowledge in individuals (Schooler & Hertwig, 2005). The adaptive function of human information processing supports decision making and evaluation for the future (Klein et al., 2010).

The concept of "intentional forgetting" refers to the processes through which humans regulate and deliberately control their memories (Bjork et al., 1998; Lehman & Malmberg, 2011; Payne & Corrigan, 2007). Forgetting is equally relevant as learning for organizations in dynamic environments. Organizations adjust their goals and strategies in response to changes in markets, customer needs, technology, and regulations and must subsequently forget prior objectives and solutions to prioritize current ones. We propose that intentional forgetting is especially relevant in phases of incremental innovation, describing improvements in existing products, core processes and services (Mumford, 2012), and in terms of process and routine adaptation when these are highly overlearned and executed mindlessly (Kluge et al., 2024; Roling et al., 2024a). To achieve currently relevant goals, previous organizational goals need to be dropped. Organizations use intentional forgetting to support decision making and to regulate and control organizational memory.

Organizational Intentional Forgetting is defined as the regulation and deliberate control of memory items and includes deliberately reducing the influence of outdated knowledge on cognitive and behavioral processes (Grisold et al., 2017; Kluge & Gronau, 2018).

Common characteristic of unlearning and intentional forgetting is the goal of deliberately reducing the influence of organizational memory elements or items that hinder adaptation to a changing environment. This can be knowledge about former failure and success and mindsets of decision makers and new product development teams or behavioral patterns and highly overlearned routines. However, we also see relevant differences between these two concepts:

Unlearning: We conceptualize unlearning as part of a double-loop learning process, that is radical, with a strategic focus, and of episodic nature. Unlearning addresses the mental set of decision makers, aims at improving mid- and long-term results, and can be seen as part of an ideation process being the first phase of innovation. An example is an organization's transformation toward servitization as a new business model.

Intentional forgetting: In contrast, intentional forgetting addresses single-loop learning processes, that are incremental, with an operational focus and of continuous nature. Intentional forgetting addresses the behavioral set of employees with a short-term focus on improving implementation processes and is part of the evaluation process being the second phase of innovation. An example is the implementation of a new routine as a result of an innovation-related decision based on unlearning.

Commonalities and differences between unlearning and intentional forgetting are illustrated in Fig. 1.

3 Why Is Unlearning and Intentional Forgetting Easier Said Than Done?

Although the need for an organization to change and adapt to changes in the environment for the bystander seems obvious, organizations do not adapt and change as expected. A possible explanation for not changing is borrowed from cognitive and learning psychology. An organization's history and the reinforcement of its actions or products through success often lead to a "mental set" (Schultz & Searleman, 2002). Reinforcement learning—learning based on (past) success—results in being rewarded and reassured that the organization is doing the right things which have brought market success. According to self-regulation theory (Hellervik et al., 1992; Kanfer, 2005; Kanfer & Hagerman, 1987), signals from the environment could provide cues relevant to change to avoid unwanted and undesired outcomes. However, early and initially weak signals indicating market changes are not strong enough to prompt reflection and questioning of what is happening in the market. From an organizational psychology perspective, a long tradition and established identity as well as a strong culture in an industry can lead to clinging too long to practices or

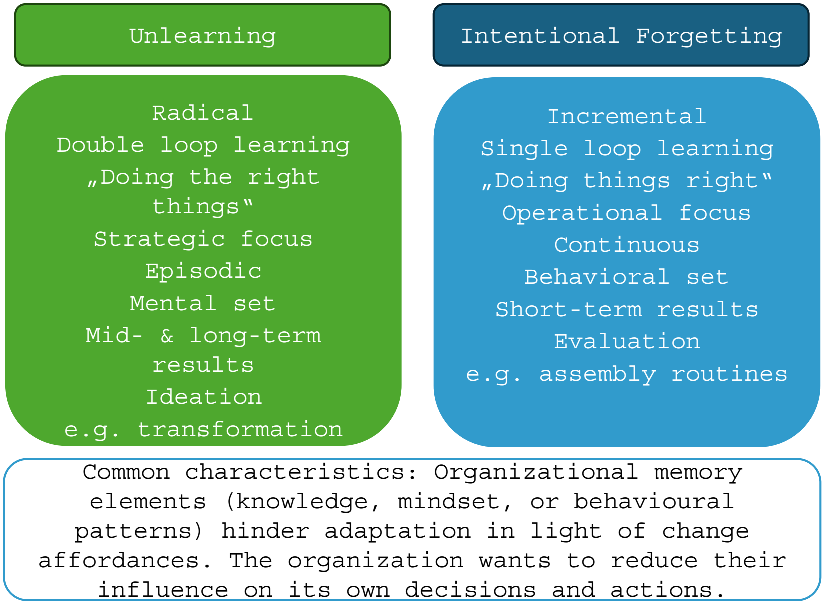


Fig. 1 Unlearning and intentional forgetting—Commonalities and differences

strategies that are no longer successful. A strong organizational culture, in this sense, can unintentionally suppress fresh perspectives (Buchanan et al., 2005; Choi, 2011; Kuipers et al., 2014).

Generally speaking, organizations show "inertia," even in cases of substantive adaptation affordances (Hannan & Freeman, 1984; Omidvar et al., 2023). Inertia is described as the inability to enact internal change in the face of significant external change (Gilbert, 2005). Such inertia is a result of path dependency due to organizational success, based on a formula that shapes processes, competencies, relationships, values, and resource investment patterns (Omidvar et al., 2023). Inertia is usually conceptualized as reduced rate of change (i.e., slow and/or insufficient response) relative to the occurrence of opportunities and threats in the task environment (Yi et al., 2016).

Scholars also use the term rigidity to explain, why organizations don't change. On an individual level, rigidity is the tendency of an individual not to change (Schultz & Searleman, 2002). For example, organizational decision makers can be constrained by existing cognitive frames that prevent them from either observing or effectively responding to a changing environment (Omidvar et al., 2023). Rigidity involves the formation of a mental or behavioral set, which is a learned mental or behavioral pattern being formed through repeated organizational experience situations (Luchins, 1942). Mental sets are expectations about future events (including attitudes, beliefs, expectancies, and schemas), whereas behavioral sets are patterns of observable responses. Rigidity involves the preservation of these sets. Preservation means the continuation of the set in the face of pressure to change. Rigidity is the tendency to develop and perseverate in the use of mental or behavioral sets.

A mental set (also called "Einstellung" or fixation; Blech et al., 2020; Luchins, 1942) describes the phenomenon where established routines or evaluating the usefulness of a situation can prevent finding other, possibly more efficient or more creative solutions. An organizational member or team is simply not aware of a new and promising option due to strong mental sets. If someone prefers a well-known routine although a simpler, more economic approach to a task or problem would be applicable, this might be a "blindness" toward alternative solutions (Blech et al., 2020; Luchins, 1942). It describes the idea that organizational members develop procedures for dealing with standard organizational situations while becoming blind to alternatives and changes. Thinking procedures and collections of rules to be applied in certain situations are a common way of dealing with affordances that become stronger with every application, resulting in an increased fixation of teams or team members on that particular way of dealing with a problem (Bilalić et al., 2008). The old procedure is readily available and, in many cases, works satisfactorily, thus giving no feedback that something else could be a more valuable solution.

Several cognitive theories assume that memory activation is behind the mental set-effect as the usually used solution is the first one that comes to mind (Bilalić & McLeod, 2014; Bilalić et al., 2010). In that respect, a subsequent search for new solutions (the ideation process) is influenced by the initial idea. Memory activation is just the beginning of a pernicious cycle where subsequent attention leads to further perceptions of the elements related to the initial idea (Blech et al., 2020). This in turn feeds back the initial memory activation. Although people believe that they are looking for new solutions, they will inevitably be trying to refine the initial way of dealing with the situation. Additionally, empirical evidence shows that individual intentional forgetting is facilitated by person-related variables such as the ability for critical reflection, the use of cognitive control strategies, retentivity as a facet of intelligence, the ability to suppress thoughts, and goal orientation, while it is impeded by defensive routines (Kluge, 2023).

At the organizational level, rigidity comes in two distinct forms: (a) resource rigidity, which is the failure to change resource investment patterns (e.g., following established organizational resourcing patterns that fail to prioritize investments in new technologies or capabilities (Omidvar et al., 2023)), and (b) routine rigidity, which is the failure to change the organizational processes that use those resource investments (Gilbert, 2005). Routines are defined as repeated patterns of response involving interdependent activities that become reinforced through structural embeddedness and repeated use (Gilbert, 2005).

Habits and so-called overlearned routines lead to the automatic execution of actions (mindlessness), a consequence of single-loop learning. While these processes may have once driven efficiency, they can now act as obstacles, preventing the organization from redirecting efforts to new opportunities. Long-standing processes and routines create a robust organizational memory and a routine rigidity. Routines are associated with a strong drive toward organizational stability and inertia (Yi et al., 2016). Organizational processes that are tightly aligned with one environment can be difficult to change, because they are self-reinforcing and are not built to adapt to discontinuities (Volberda et al., 2021).

4 Ideation-Exploration and Evaluation-Exploitation Processes in Innovation

As introduced above, we propose processes of unlearning as relevant for radical innovation and strategic change, and intentional forgetting as relevant for incremental innovation and the implementation of adapted routines. In the following section, we elaborate on these propositions focusing on unlearning and intentional forgetting as well as on two subsequent phases in an organization's innovation process: the ideation and evaluation process (Puccio & Cabra, 2012). For an overview, please see Fig. 2.

Generally, creativity and innovation need the process of idea generation and idea evaluation (Puccio & Cabra, 2012). Idea generation describes the development of new ideas while idea evaluation refers to the judgment of those (Basadur, 1995). At the organizational level, unlearning is required in the ideation phase to adapt strategically to a changing and turbulent environment and exploring new ideas. It takes place in top management teams, new product development teams, and other forms of team settings. Intentional forgetting focuses on adapting operational processes and exploiting the results of the ideation process to achieve the new strategic focus in the evaluation phase. Therefore, it takes place on the operational shopfloor level (Ellwart & Kluge, 2019; Kluge and Gronau, 2018).

The goal of unlearning for double-loop learning is to impede the retrieval and remembrance of beliefs about the company's failure and success and to detach itself from the old solutions to become open to the combination of novel and useful ideas (Plucker et al., 2004). At the organizational level, heterogeneous top management teams, guidelines on what should be unlearned, and impeding the reinforcement of outdated knowledge and routines are supportive for unlearning. Empirical evidence shows that team-level unlearning needs, e.g., supporting leadership exploration activities, encouragement for breaking patterns, creating space for experimentation, providing time, encouraging failure, and risk aversion (Kluge, 2023). These facilitating factors are important in the phase of ideation.

The ideation process in organizations might take place in creativity workshops, Design Thinking sessions, meetings, or new product development teams, in which individual members combine their knowledge about the organizational affordances, requirements, and possible solutions, which is called the problem space. From a psychological perspective, unlearning affects a team's cognitive representation of the problem space (Simon & Newell, 1971; Ward, 2012). The problem space contains states of knowledge (problem states) as well as means of moving from one state to another (operators). The way how organizational members construct the problem space is an important aspect of problem solving. Ward (2012) shows that, e.g., the recent exposure and frames in which organizational knowledge and operators are presented prohibit or facilitate problem solving and unlearning, respectively.

Unlearning in the ideation phase can be supported by techniques such as divergent thinking. Divergent thinking allows original, varied, and unconventional ideas and solutions. It is a process of merging disparate elements to develop a new and useful combination. Divergent thinking is the ability to combine and reorganize memories related to the aspired solution (Kilgour, 2006). The creative process of ideation and evaluation is the process of merging thought categories or mental images across and within domains in ways that have not been done before, to develop an original and useful solution to a situation or problem.

The ideation process requires the creation of new memory structures through the combination of distinct concepts or the new combination of elements of existing concepts. We propose that divergent thinking is a means of unlearning as learned combination of concepts and mental images need to be unlearned, e.g., the use of material, techniques, technical parts.

Subsequently, novelty generation must be followed by the evaluation and exploitation of novelty. The evaluative component of creativity is important because only ideas that are applicable, useful, and feasible as well as novel are likely to lead to a concrete contribution in a market or organization. Evaluation is required as not all novel ideas in organizations can be accepted. Overabundance of novel ideas may become distracters, and evaluation helps to identify options and solutions with noteworthy potential (Acar & Runco, 2012). The evaluation process is also viewed as a process of convergent thinking or conceptual foresight (Guilford & Hoepfner, 1971),

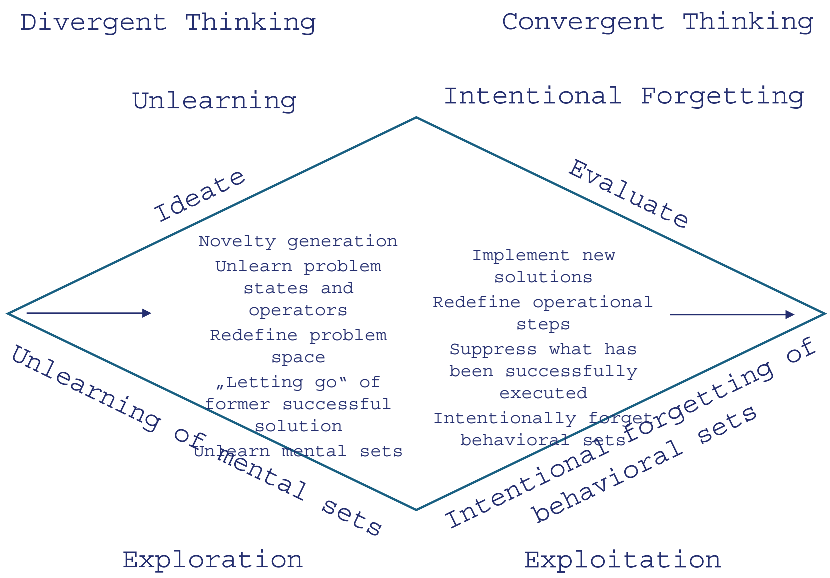


Fig. 2 Unlearning content in the ideation-exploration phase and intentional forgetting in the evaluation-exploitation phase

which includes to forecast the possible implications of a new idea as possible outcomes are projected. Regarding the forecast, research showed that in the evaluation process, teams might need to question former success beliefs (what makes us successful?) as well as failure beliefs (what was not successful in the past?) (Guilford & Hoepfner, 2023).

5 Unlearning for Radical Innovation: The Case of Servitization

An illustrative unlearning example is the process of servitization. Industries like mechanical engineering face significant challenges in servitization due to deeply ingrained product-centric practices (Pöppelbuss & Kluge, 2024). Servitization has been defined as a "transformational process of shifting from a product-centric business model and logic to a service-centric approach" (Kowalkowski et al., 2017, p. 7) that requires a company to redefine its business model and strategy, reconfigure its resource base, and transform its routines as well as shared norms and values (Kowalkowski et al., 2017). The transformation toward a service-oriented and increasingly digital business model leads to challenges in industries like mechanical engineering that have previously followed a predominantly product-centric approach to innovation (Ebel et al., 2022; Paiola & Gebauer, 2020). With customer-centricity as a key attribute of new (digital) service offerings, companies from such industries need to adapt their previous product-centric and technology-driven ways of thinking (Töytäri et al., 2018), as well as their capabilities (Castka et al., 2024; Kanninen et al., 2017).

Poeppelbuss and Kluge (2024) argue that it is overlooked that most organizations have well-functioning, highly overlearned routines and operational capabilities that possibly impede the organizational transformation and need to be unlearned and forgotten parallel to the servitization process. The need to cope with such product-oriented legacies is particularly apparent from paradoxes that companies typically face during servitization, e.g., "building a customer orientation vs. maintaining an engineering mindset" (Kohtamäki et al., 2020). In the transition phase, organizations try to explore the "new" while unlearning the "old," while exploiting new learnings and disciplining the "new" knowledge and routines.

The pace and the smoothness of servitization in companies depend on their ability to unlearn previously successful ways of doing business and to unlearn highly overlearned routines, e.g., those that resemble a strong orientation toward products and their technical features. Poeppelbuss and Kluge (2024) combine the concepts from servitization and unlearning research into an integrated framework (see Fig. 3). Servitization is assumed as a transformation that is achieved through interventions that help both build capabilities relevant to service-based business models and abandon obsolete knowledge.

At the organizational level, unlearning for servitization relates to the emphasis on product features and traditional product services (former success beliefs), unlearning of failures to monetize new service offerings due to the lack of willingness to pay, unlearning of beliefs about customer satisfaction based on product quality, unlearning of core processes regarding the interpretation of market indicators, product development, production/assembly, quality control and delivery, and unlearning of beliefs about the product-engineering mindset.

At the team level, unlearning includes the mere attention on product performance and the installation of management metrics to facilitate customer-focused and data-driven solutions, unlearning of product-only-related knowledge toward service-based offerings, and unlearning of knowledge sharing activities that optimize products instead of understanding and discussing customer needs and services.

Finally, at the individual level, this requires unlearning of the engineering and product-oriented mindset toward a service and solution-oriented and problem-solving attitude, while unlearning a "selling a product" attitude. When new routines are established, employees need to intentionally forget the production handling and replace it by error-free handling of service and solution delivery process.

Understanding unlearning requirements might be promising for better understanding and potentially reconciling the paradoxes in the process toward servitization experienced in product-oriented manufacturing companies (Kohtamäki et al., 2020).

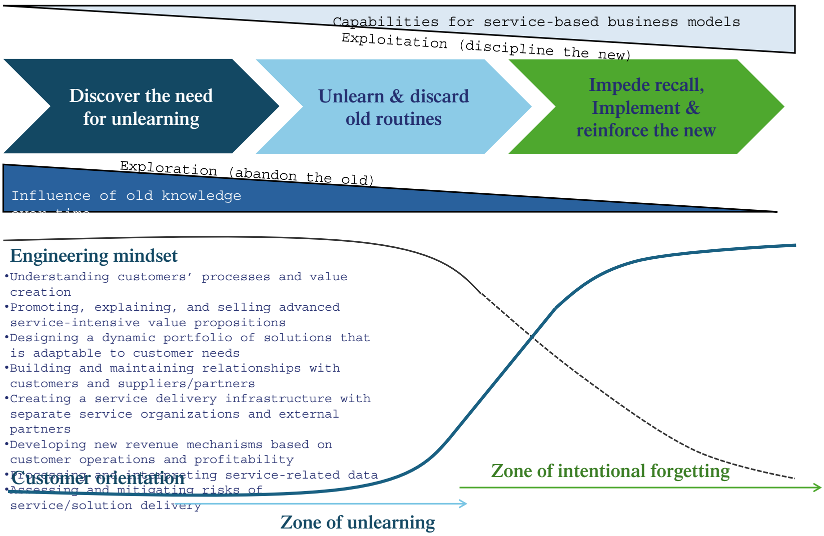


Fig. 3 Unlearning framework for servitization (Adapted from Pöppelbuss and Kluge [2024])

6 Intentional Forgetting for Incremental Innovation—Adapting to New Routines

As outlined earlier, intentional forgetting plays a key role in single-loop learning processes, focusing on gradual improvements aligned with operational goals. It shapes employee behaviors with a short-term emphasis on refining implementation processes and represents a critical part of the evaluation phase within the broader innovation cycle. Established processes and routines often build a strong organizational memory but can also lead to rigidity. At an organizational level, routines are closely tied to stability and can foster inertia (Yi et al., 2016).

Organizational routines differ in their content, structure, timing, degree of formalization, and the knowledge required for their execution. They not only streamline coordination and guide human actions but also optimize the use of cognitive resources. Parallels can be drawn between organizational routines and individual procedural knowledge, as well as group-level distributive knowledge (Cohen & Bacdayan, 1994), as they are defined as "multi-actor, interlocking, reciprocally triggered sequences of actions" (Cohen & Bacdayan, 1994, p. 554).

Procedural knowledge comprises condition-action units that link specific cues to corresponding cognitive actions (Anderson & Schunn, 2013). Routines serve to standardize transformation processes, converting inputs—such as materials, energy, or client orders—into outputs like products, services, or employee skills (Cyert & March, 1963). They are characterized by repetition, recognizable patterns of action (Becker, 2004), and the interdependent contributions of multiple actors (Becker, 2004; Pentland & Hærem, 2015). In manufacturing, routines transform workflows into efficient systems, offering clear instructions to ensure consistent and effective product creation (Mor et al., 2019).

The degree of routinization depends on the extent to which tasks lack variety (Perrow, 1967), involve minimal exceptions over time (Daft & Macintosh, 1981), and promote predictability and consistency (Ahuja & Carley, 1999; Kluge, 2014). Research on organizational routines (e.g., Becker, 2004; Gersick & Hackman, 1990; Helfat & Karim, 2014; Miller et al., 2012; Pentland & Hærem, 2015) consistently highlights their stabilizing influence but also their resistance to change.

Highly practiced routines, executed repetitively and without variation, become overlearned. This automaticity reduces cognitive effort, leading to smoother, goal-directed actions performed almost effortlessly (Proctor & Dutta, 1995). Such "mindless" execution, as Pentland and Hærem (2015) describe it, is a hallmark of well-rehearsed routines.

However, when organizations face change, deeply ingrained routines can hinder adaptability by narrowing focus and reinforcing past behaviors (Gilbert, 2005). Intentional forgetting becomes crucial in overcoming this inertia. As Miller et al. (2012) note, adaptation requires deliberately letting go of outdated routines (Miller et al., 2012). This process involves deliberate interventions, such as halting unnecessary steps, modifying existing ones, or incorporating new actions into established workflows.

Roling et al. (2023) and Schüffler et al. (2020) identify three types of routine adjustments: (1) executing actions differently (e.g., measuring in inches instead of centimeters), (2) omitting obsolete steps (e.g., delegating them to a robot), and (3) adding new steps (e.g., calibrating machinery). During such transitions, "mindful" execution becomes essential (Pentland & Hærem, 2015), as employees must actively monitor and adjust their actions to ensure smooth integration of changes.

To empirically investigate, what is needed to intentionally forget parts of routines, we conducted several studies at the Research and Application Center for Industry 4.0 (RACI) at the University of Potsdam and Ruhr University Bochum to explore the effects of intentional forgetting on change management within the context of Industry 4.0. The RACI provides a hybrid production simulation environment that integrates hardware components such as transport systems, robots, and QR scanners with advanced software tools (Gronau et al., 2023; Schüffler et al., 2020).

The research focused on the simulated production of highly individualized products by teams of two participants. The participants were assigned distinct worker roles on the assembly line and executed a manufacturing routine. At the first measurement point in the RACI, the participants learned and trained the specific work steps of the production process. After this, they further practiced their work steps using an online training app. At the second measurement in the RACI (which occurred either one or three weeks after the first), participants expected to execute the previously learned production process again. Unexpectedly, they were informed that specific steps of the production process would be subject to change. They had to adapt their routinized work steps, requiring learning and intentional forgetting (Fig. 4). Several contextual factors were varied to investigate their impact on intentional forgetting and adaptation (for details, see Kluge et al., 2024; Roling et al., 2024b).

- The impact of clarity on the need for intentional forgetting when changes to a previously learned production routine are required. Clarity was operationalized through job aids and posters displayed at the workplace, which either (a) showed and visualized the steps of the original routine, (b) showed and visualized the steps of the adapted routine, (c) showed both the original and the adapted routines, or (d) showed information unrelated to any routine.
- Time pressure experienced by the groups. During the production process, the production times for each participant, as well as for the team, were measured per workpiece. These times were displayed by timers on each workpiece. One group was not subjected to time pressure, and the timers mirrored those used during the initial laboratory session. The required production time did not influence outcomes, allowing teams to operate without time constraints. The other group faced time pressure, with the timers now displaying countdowns instead of counting upwards. When the timer expired, a signal tone sounded. Participants were informed that meeting the time requirements would result in an additional bonus.
- Convenience and effort reduction versus decreased convenience and increased effort. The newly introduced steps in the procedure were either noticeably more cumbersome and, therefore, less appealing or significantly more convenient and, thus, more attractive to execute. These changes either facilitated the routine's application or made it more challenging.
- Episodic versus continuous change. Changes were introduced either stepwise during the processing of multiple workpieces or all at once during the processing of a single workpiece. If participants did not adapt to the new requirements and still focused on the obsolete work steps, adaptation errors were counted.

To measure adaptation and forgetting, we utilized system log files to track participants' performance. Additionally, video data and participants' written documentation were analyzed in some of the sub-studies to gain insights into participants' observations and reflections.

The studies revealed several key findings. First, highly proceduralized actions—those that had been well rehearsed—were considerably harder to forget than less practiced actions (Schüffler et al., 2020).

Second, we found that all groups experienced an increase in production time after routine adaptations (Kluge et al., 2024). The mere modification of routines affected production time, with medium to large effect sizes. This effect persisted even when the number of actions was reduced, which theoretically should have saved time. The execution speed of the modified actions slowed down and became less smooth, independent of other experimental variables. Therefore, the most influential factor for increased production time was the introduction of the change itself. In contrast, the total number of general errors remained unaffected and stayed constant.

We also observed that explicitly highlighting the differences in how the adapted routine should be executed ("Do not do this—now do that") through posters, which indicated the steps no longer performed, facilitated the intentional forgetting process and supported the phase of mindful execution and re-learning (Kluge et al., 2024).

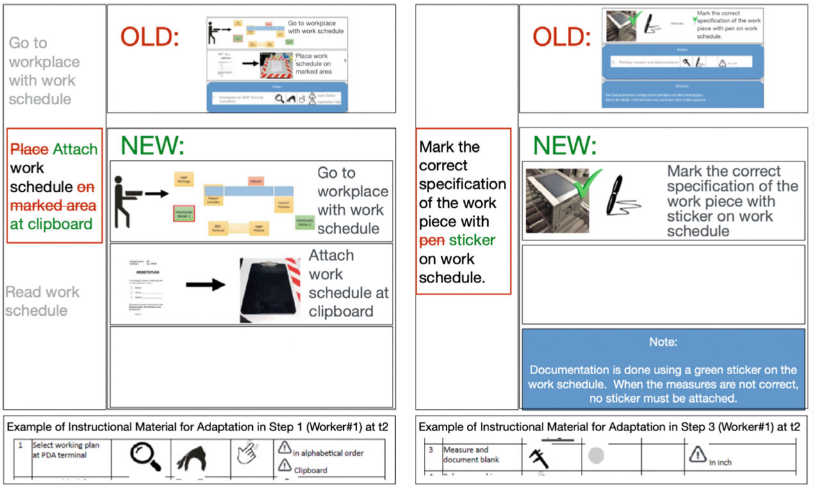


Fig. 4 Example of instructional material to support intentional forgetting (Kluge et al., 2024)

Additionally, regarding the impact of convenience changes, we observed a significant reduction in errors for both groups (Kluge et al., 2024). This result was unsurprising for the group experiencing increased convenience, as the actions became more intuitive and "natural." However, it was unexpected for the group with reduced convenience, as they had to perform more effortless and somewhat illogical actions. We suggest that the actions assigned to the reduced convenience group were so unconventional that they were memorized more easily despite their questionable rationale.

Change introduction in a stepwise manner led to significantly faster task execution time compared to change introduction all at once (Roling et al., 2024b). At the descriptive level, stepwise change introduction also indicated a lower number of adaptation errors. This is in line with our assumption that stepwise change introduction reduces cognitive load and facilitates behavioral adaption.

Furthermore, change introduction in a stepwise manner led to higher frustration at the descriptive level (Roling et al., 2024b). This result was surprising, as we assumed that stepwise change introduction would not only enhance adaptive performance but also decrease frustration. However, the results indicate that too frequent change implementation might have negative effects on frustration.

Lastly, Roling et al. (2024a) showed that the type of change content (addition of new actions, omission of old actions, changes in the manner of action execution) had an impact on adaptive performance.

We also identified effects of person-related variables, independent of other factors (Kluge et al., 2024). The level of intentional forgetting, as measured by errors, was influenced by retentivity. The higher the individual's retentivity, the fewer the errors (Kluge et al., 2024) and the faster the task execution time (Roling et al., 2024b).

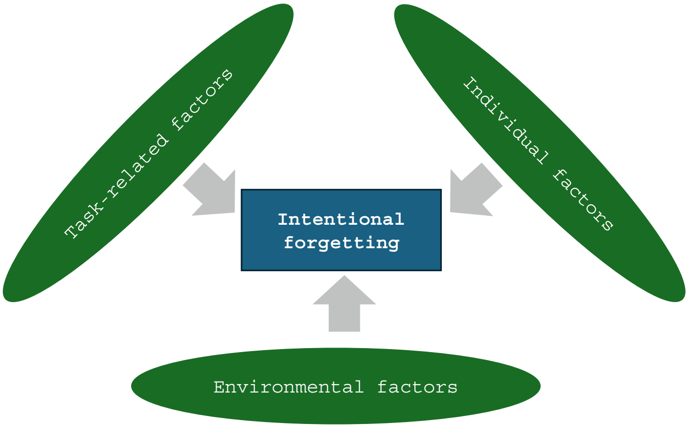


Fig. 5 Influence of individual, task-related, and environmental factors on intentional forgetting

This was also true for the perceived "switching costs," which refer to the perceived effort required for intentional forgetting and mindful execution of the adapted routine during its initial trials (Kluge et al., 2024). Individuals with higher retentivity experienced greater self-efficacy, lower perceived adaptation/switching costs, and made fewer errors overall (Kluge et al., 2024). Specific self-efficacy was significantly negative correlated with frustration (Roling et al., 2024b). In the condition of stepwise change introduction, the number of pre-change errors was significantly negative correlated with specific self-efficacy (Roling et al., 2024b).

The results from Kluge et al. (2024), Roling et al. (2024a; 2024b), and Schüffler et al. (2020) outline the importance of individual, environmental, and task-related factors for adaptive performance at the individual level. If organizations aim to change and adapt, they must consider employees' individual characteristics, design new tasks in a precise and detailed manner, and create a supportive environment for individual intentional forgetting (see Fig. 5). Intentional forgetting helps to detach from inefficient behavioral sets and facilitates adapting to new tasks. Supporting intentional forgetting at the individual level can be seen as precondition for successful forgetting at the organizational level (Zhao et al., 2013) and essential for overcoming inertia.

7 Intentional Forgetting as Incremental Change: The Case of New Communication Guidelines in Public Administration

A case for intentional forgetting outside the manufacturing context is illustrated by Roling et al. (2026). In public administration, work is characterized by routinized office work, e.g., when citizens' concerns are dealt with. Many administrative processes, such as the processing of identity card applications or issuing birth certificates, always contain the same steps that must be done. High levels of bureaucracy and standardization result in recurring tasks and repetitive task execution. Like manufacturing, public administration is another example for an organization strongly shaped by organizational routines.

Often, the public sector is considered less innovative than the private sector (Wilson & Mergel, 2022). In Germany, rigid hierarchical structures and low levels of revitalization, agility, and flexibility characterize many public sector organizations. However, public sector organizations must overcome those structures and processes to keep up with the times. Digital transformation is strongly needed to become a future-oriented organization that is able to offer the best possible services for citizens and is seen as attractive employer (Mergel et al., 2019). At the operational level, digital transformation comes along with specific task changes. Digital technologies automatically lead to specific work step changes, e.g., when all requests from citizens are processed using an online tool instead of in paper form.

The relationship between reflecting on previous work processes and relapsing into them was empirically examined through a longitudinal field study conducted in collaboration with a local public sector authority in Germany (see Roling et al., 2026). The study was framed around the implementation of new communication guidelines, using pre- and post-surveys. In the post-implementation survey, participants self-assessed their tendency to think about old work steps and their frequency of relapsing to them.

The findings revealed a significant positive correlation between these two variables: individuals who frequently thought about outdated work processes were more likely to relapse into them. This underscores a direct link between memory recall and behavior, suggesting that retaining outdated routines increases the likelihood of errors tied to those routines.

Additionally, both reflecting on former work steps and relapsing to them showed significant negative correlations with the perceived convenience of new tasks. This highlights the critical role of task design in supporting organizational change. When required adaptations are perceived as convenient and satisfying, they appear to promote intentional forgetting, which is essential for successful change. However, as the study is based on cross-sectional data, further research is needed to deepen our understanding of this assumption.

Another noteworthy finding was that intentional revitalization readiness significantly predicted reverting to old work steps, whereas cognitive and emotional change readiness did not explain additional relapsing variance. This suggests that the intention to embrace change is a key driver of behavioral adaptation.

This study specifically explored intentional forgetting at the individual level, emphasizing its importance for performance during organizational change. Individual perceptions and attitudes emerge as critical factors influencing intentional forgetting in such scenarios. To foster effective change, organizations should focus on the operational level, closely examining individual task execution. By doing so, they can better support single-loop learning, incremental improvements, and intentional forgetting as part of their change management strategies.

8 Discussion

Before discussing the implication of our proposed differentiation between unlearning and intentional forgetting, we acknowledge that understanding the phenomenon comprehensively, country-specific regulations and protective mechanisms, such as tariffs or subsidies, can enable a "business as usual" approach, even when the market has already shifted and the need for adaptation has been understood and realized. We understand that re-learning or transforming an entire organization as a social or socio-technical system is a significant challenge. Sometimes, we wish we could change an organization simply by installing a new "operating system." At other times, we dream of using the "neuralyzer" from Men in Black to make everyone forget what they've seen or done in the past. But we know: it's not that simple—not even for an operating system.

Unlearning and intentional forgetting have been considered as relevant constructs in the context of knowledge management during organizational change. However, past literature did not clearly distinguish between the two terms and often used them nearly interchangeably to describe the letting go of outdated memory elements (Kluge, 2023). This chapter elaborated the commonalities and differences of unlearning and intentional forgetting to shed light on these two concepts.

Both concepts have in common that they facilitate memory reorganization during a change. Unlearning and intentional forgetting support adaptation and help to overcome inertia which makes both concepts being crucial for organizational flexibility and future orientation (Klammer & Gueldenberg, 2019; Kluge & Gronau, 2018). However, unlearning and intentional forgetting differ regarding their foci and underlying principles.

We conceptualized unlearning as specifically addressing radical changes and strategic shifts while intentional forgetting is focused on incremental change at the operational level. Unlearning is seen as relevant when new ideas are generated and explored. Intentional forgetting specifically focuses on the evaluation of those ideas. However, based on the conceptual distinction outlined in this chapter the following questions arise:

Are there temporal and causal relationships between unlearning and intentional forgetting? Does intentional forgetting always follow unlearning?

When strategic changes are broken down to the operational level, modifications in work steps and routines are a natural consequence. Abernethy et al. (2021) stated that strategic change is followed by operational change, which in turn leads to performance change. Based on this assumption, one could argue that intentional forgetting must inevitably follow unlearning after a strategic shift. Changes in an organization's strategy require unlearning obsolete mindsets, while intentional forgetting subsequently becomes necessary to adapt behaviors to the new strategic orientation at the operational level. However, strategic change is influenced by multiple actors and antecedents, such as institutional or tech-related factors (Acciarini et al., 2024). The complexity of strategic change, shaped by numerous influencing factors, might also impact its temporal dynamics and sequential order.

Are unlearning and intentional forgetting distinct concepts, or do they form a continuum?

The considerations of temporality and causality suggest that organizations are not simply faced with either unlearning or intentional forgetting. Rather than being distinct categories between which organizations must choose, the two concepts should be viewed as a continuum. In many cases, both unlearning and intentional forgetting play a role in adapting to change. Both processes are essential for organizational flexibility, and play together during organizational change. Depending on the specific focus of change, either one or the other may be of greater importance. However, unlearning and intentional forgetting are not automatically mutually exclusive. Future research should further develop this perspective to better understand the interplay between unlearning and intentional forgetting.

Can we develop methods to actively support individuals and teams in the unlearning endeavor, e.g., equivalent to methods for divergent thinking and creative thinking?

De Bono stated that "It is not the ideas we do not have that block our thinking but the ideas that we do have" (Edward de Bono)—but how can organizational members use ideas, they did not have yet? We should elaborate on thinking techniques for individuals and teams, that can be trained and used in discussions about future directions and strategies of the organization. Methods designed to inhibit the automatic activation of frequently used memory patterns and solutions should be developed. For this purpose, identifying triggers for unlearning would be beneficial in designing supportive techniques. One example of such a technique could be the generation of new ideas through reverse brainstorming. Considering how to make a problem worse forces individuals to approach it from a different perspective, which may help detach from outdated solutions.

Which implications for change management can be derived? The results on routine adaptation showed that important insights arise from looking into the different ways routines may change. Regarding an organizational diagnosis, change teams in production management might be equipped with checklists, that help to identify the most critical steps facilitating intentional forgetting (e.g., designing new tasks more convenient) or hindering it (e.g., introducing too many changes at the same time). These must be integrated into the design of change processes to create supportive environmental conditions for adaptation.

Additionally, change managers should be supported in knowledge about the importance of time needed for intentional forgetting as a cognitive process in addition to commonly learned knowledge about resistance and emotional states in the first place. Furthermore, they should be aware of the individual differences people have regarding their ability to adapt to changes (e.g., the impact of retentivity). This emphasizes the need for well-tailored change management approaches rather than one-size-fits-all solutions.

Overall, organizations must consider knowledge management as an essential part of change management, as the two are strongly intertwined with each other. There is a need for an increased awareness of the critical role of unlearning and intentional forgetting during organizational change. This message must be transported into every organization to reach the goal of successful development and adaptation.

References

Abernethy, M. A., Dekker, H. C., & Grafton, J. (2021). The influence of performance measurement on the processual dynamics of strategic change. *Management Science, 67*(1), 640–659.

Acar, S., & Runco, M. A. (2012). Creative abilities. In M. D. Mumford (Ed.), *Handbook of organizational creativity* (pp. 115–139). Elsevier.

Acciarini, C., Boccardelli, P., & Peruffo, E. (2024). Strategic change: A systematic review, synthesis, and a future research agenda. *European Management Review, 21*(4), 782–802.

Ahuja, M. K., & Carley, K. M. (1999). Network structure in virtual organizations. *Organization Science, 10*(6), 741–757.

Altmann, E. M., & Gray, W. D. (2002). Forgetting to remember: The functional relationship of decay and interference. *Psychological Science, 13*(1), 27–33.

Anderson, J. R., & Schunn, C. D. (2013). Implications of the act-r learning theory. In R. Glaser (Ed.), *Advances in instructional Psychology* (Vol. 5, pp. 1–33). Routledge.

Argote, L. (2013). Organizational forgetting. In L. Argote (Ed.), *Organizational learning* (pp. 57–84). Springer, US.

Argyris, C., & Schön, D. A. (1978). *Organizational learning: A theory of action perspective. Organization development series*. Addison-Wesley Pub. Co.

Basadur, M. (1995). Optimal ideation-evaluation ratios. *Creativity Research Journal, 8*(1), 63–75.

Becker, M. C. (2004). Organizational routines: A review of the literature. *Industrial and Corporate Change, 13*(4), 643–678.

Bilalić, M., & McLeod, P. (2014). Why good thoughts block better ones. *Scientific American, 310*(3), 74–79.

Bilalić, M., McLeod, P., Gobet, F., & de Groot. (2008). Expert and "novice" problem solving strategies in chess: Sixty years of citing de Groot (1946). *Thinking & Reasoning,14*(4), 395–408.

Bilalić, M., McLeod, P., & Gobet, F. (2010). The mechanism of the einstellung (set) effect. *Current Directions in Psychological Science, 19*(2), 111–115.

Bjork, E. L., Bjork, R. A., & Anderson, M. C. (1998). Verieties of goal-directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting* (pp. 103–137). Lawrence Erlbaum.

Blech, C., Gaschler, R., & Bilalić, M. (2020). Why do people fail to see simple solutions? using think-aloud protocols to uncover the mechanism behind the einstellung (mental set) effect. *Thinking & Reasoning, 26*(4), 552–580.

de Bono, E. (2014) *Lateral thinking*. https://www.debono.com/quotes-1

Buchanan, D., Fitzgerald, L., Ketley, D., Gollop, R., Jones, J. L., Lamont, S. S., Neath, A., & Whitby, E. (2005). No going back: A review of the literature on sustaining organizational change. *International Journal of Management Reviews, 7*(3), 189–205.

Castka, P., Donovan, J., & Sousa, R. (2024). Pathways of development of dynamic capabilities for servitization transformation: A longitudinal multi-case study. *Industrial Marketing Management, 116*, 66–81.

Choi, M. (2011). Employees' attitudes toward organizational change: A literature review. *Human Resource Management, 50*(4), 479–500.
Cohen, M. D., & Bacdayan, P. (1994). Organizational routines are stored as procedural memory: Evidence from a laboratory study. *Organization Science, 5*(4), 554–568.
Cyert, R. M., & March, J. G. (1963). *A behavioral theory of the firm*. Prentice Hall.
Daft, R. L., & Macintosh, N. B. (1981). A tentative exploration into the amount and equivocality of information processing in organizational work units. *Administrative Science Quarterly, 26*(2), 207.
Easterby-Smith, M., & Lyles, M. (2003). Re-reading organizational learning: Selective memory, forgetting, and adaptation. *Academy of Management Perspectives, 17*(2), 51–55.
Easterby-Smith, M., & Lyles, M. A. (2011). In praise of organizational forgetting. *Journal of Management Inquiry, 20*(3), 311–316.
Ebel, M., Jaspert, D., & Poeppelbuss, J. (2022). Smart already at design time—Pattern-based smart service innovation in manufacturing. *Computers in Industry, 138*, 103625.
Ellwart, T., & Kluge, A. (2019). Psychological perspectives on intentional forgetting: An overview of concepts and literature. *KI - Künstliche Intelligenz, 33*(1), 79–84.
Fiol, C. M., & O'Connor, E. J. (2017). Unlearning established organizational routines—Part II. *The Learning Organization, 24*(2), 82–92.
Fiol, C. M., & O'Connor, E. J. (2017). Unlearning established organizational routines—Part I. *The Learning Organization,24*(1), 13–29.
Gersick, C. J., & Hackman, J. R. (1990). Habitual routines in task-performing groups. *Organizational behavior and human decision processes, 47*, 65–97.
Gilbert, C. G. (2005). Unbundling the structure of inertia: Resource versus routine rigidity. *Academy of Management Journal, 48*(5), 741–763.
Grisold, T., Kaiser, A., & Hafner, J. (2017). Unlearning before creating new knowledge: A cognitive process. In T. Bui (Ed.), *Proceedings of the 50th Hawaii International Conference on System Sciences* (pp. 1–10). AIS Electronic Library (AISeL).
Gronau, N., Kluge, A., Haase, J., & Thim, C. (2023). Experiential learning factories: Bridging the gap between lab and field experiments. *SSRN Electronic Journal*.
Guilford, J. P., & Hoepfner, R. (1971). *The analysis of intelligence*. McGraw-Hill.
Hannan, M. T., & Freeman, J. (1984). Structural inertia and organizational change. *American Sociological Review, 49*(2), 149.
Hedberg, B. (1981). How organizations learn and unlearn. In P. S. Nystrom & W. H. Starbuck (Eds.), *Handbook of organizational design* (pp. 3–27). Oxford: Oxford University Press.
Helfat, C. E., & Karim, S. (2014). Fit between organization design and organizational routines. *Journal of Organizational Design, 3*(2), 18–29. https://ssrn.com/abstract=2489886
Hellervik, L. W., Hazucha, J. F., & Schneider, R. J. (1992). Behavior change: Models, methods, and a review of evidence. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 823–870). Consulting Psychologists Press.
de Holan, M., Phillips, N., & Lawrence, T. B. (2004). *Managing organizational forgetting*. https://sloanreview.mit.edu/article/managing-organizational-forgetting/
de Holan, P. M., & Phillips, N. (2004). Remembrance of things past? The dynamics of organizational forgetting. *Management Science, 50*(11), 1603–1613.
Howells, J., & Scholderer, J. (2016). Forget unlearning? how an empirically unwarranted concept from psychology was imported to flourish in management and organisation studies. *Management Learning, 47*(4), 443–463.
Huang, D., Chen, S., Zhang, G., & Ye, J. (2018). Organizational forgetting, absorptive capacity, and innovation performance. *Management Decision, 56*(1), 87–104.
Huber, G. P. (1991). Organizational learning: The contributing processes and the literatures. *Organization Science, 2*(1), 88–115.
Kanfer, F. H., & Hagerman, S. (1987). A model of self-regulation. In F. Halisch & J. Kuhl (Eds.), *Motivation, intention, and volition* (pp. 293–307). Springer, Berlin Heidelberg.

Kanfer, R. (2005). Self-regulation research in work and i/o psychology. *Applied Psychology, 54*(2), 186–191.
Kanninen, T., Penttinen, E., Tinnilä, M., & Kaario, K. (2017). Exploring the dynamic capabilities required for servitization. *Business Process Management Journal, 23*(2), 226–247.
Kilgour, A. M. (2006). Improving the creative process: Analysis of the effects of divergent thinking techniques and domain specific knowledge on creativity. *International Journal of Business and Society, 7*(2), 79–107.
Klammer, A., & Gueldenberg, S. (2019). Unlearning and forgetting in organizations: A systematic review of literature. *Journal of Knowledge Management, 23*(5), 860–888.
Klein, S. B., Robertson, T. E., & Delton, A. W. (2010). Facing the future: Memory as an evolved system for planning future acts. *Memory & cognition, 38*(1), 13–22.
Kluge, A. (2014). *The acquisition of knowledge and skills for taskwork and teamwork to control complex technical systems*. Springer.
Kluge, A. (2023). Recent findings on organizational unlearning and intentional forgetting research (2019–2022). *Frontiers in Psychology, 14*, 1160173.
Kluge, A., & Gronau, N. (2018). Intentional forgetting in organizations: The importance of eliminating retrieval cues for implementing new routines. *Frontiers in Psychology, 9*, 51.
Kluge, A., Schüffler, A. S., Thim, C., Haase, J., & Gronau, N. (2019). Investigating unlearning and forgetting in organizations. *The Learning Organization, 26*(5), 518–533.
Kluge, A., Schüffler, A. S., Thim, C., & Gronau, N. (2024). Facilitating and hindering factors for routine adaptations in manufacturing and effects on human performance—Unexpected insights from three experimental studies in a special purpose setting. *Ergonomics, 67*(12), 2063–2081.
Kohtamäki, M., Einola, S., & Rabetino, R. (2020). Exploring servitization through the paradox lens: Coping practices in servitization. *International Journal of Production Economics, 226*, 107619.
Kowalkowski, C., Gebauer, H., Kamp, B., & Parry, G. (2017). Servitization and deservitization: Overview, concepts, and definitions. *Industrial Marketing Management, 60*, 4–10.
Kuipers, B. S., Higgs, M., Kickert, W., Tummers, L., Grandia, J., & van der Voet, J. (2014). The management of change in public organizations: A literature review. *Public Administration, 92*(1), 1–20.
Leal-Rodríguez, A. L., Eldridge, S., Roldán, J. L., Leal-Millán, A. G., & Ortega-Gutiérrez, J. (2015). Organizational unlearning, innovation outcomes, and performance: The moderating effect of firm size. *Journal of Business Research, 68*(4), 803–809.
Lehman, M., & Malmberg, K. J. (2011). Overcoming the effects of intentional forgetting. *Memory & Cognition, 39*(2), 335–347.
Luchins, A. S. (1942). Mechanization in problem solving: The effect of einstellung. *Psychological Monographs, 54*(6), i–95.
Martin de Holan, P. (2011). Agency in voluntary organizational forgetting. *Journal of Management Inquiry, 20*(3), 317–322.
Mergel, I., Edelmann, N., & Haug, N. (2019). Defining digital transformation: Results from expert interviews. *Government Information Quarterly, 36*(4), 101385.
Miller, K. D., Pentland, B. T., & Choi, S. (2012). Dynamics of performing and remembering organizational routines. *Journal of Management Studies, 49*(8), 1536–1558.
Mor, R. S., Bhardwaj, A., Singh, S., & Sachdeva, A. (2019). Productivity gains through standardization-of-work in a manufacturing company. *Journal of Manufacturing Technology Management, 30*(6), 899–919.
Morais-Storz, M., & Nguyen, N. (2017). The role of unlearning in metamorphosis and strategic resilience. *The Learning Organization, 24*(2), 93–106.
Mumford, M. D. (Ed.). (2012). *Handbook of organizational creativity*. Elsevier.
Mumford, M. D., Hester, K. S., & Robledo, I. C. (2012). Creativity in organizations: Importance and approaches. In M. D. Mumford (Ed.), *Handbook of organizational creativity* (pp. 3–16). Academic Press.
Nystrom, P. C., & Starbuck, W. H. (1984). To avoid organizational crises, unlearn. *Organizational Dynamics,12*(4), 53–65.

Omidvar, O., Safavi, M., & Glaser, V. L. (2023). Algorithmic routines and dynamic inertia: How organizations avoid adapting to changes in the environment. *Journal of Management Studies, 60*(2), 313–345.

Paiola, M., & Gebauer, H. (2020). Internet of things technologies, digital servitization and business model innovation in btob manufacturing firms. *Industrial Marketing Management, 89*, 245–264.

Payne, B. K., & Corrigan, E. (2007). Emotional constraints on intentional forgetting. *Journal of Experimental Social Psychology, 43*(5), 780–786.

Pentland, B. T., & Hærem, T. (2015). Organizational routines as patterns of action: Implications for organizational behavior. *Annual Review of Organizational Psychology and Organizational Behavior, 2*(1), 465–487.

Perrow, C. (1967). A framework for the comparative analysis of organizations. *American Sociological Review, 32*(2), 194.

Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational Psychologist, 39*(2), 83–96.

Pöppelbuss, J., & Kluge, A. (2024). Unlearning as a facilitator for servitization? Conceptualizing the role of abandoning the old on the way toward new service-based business models. In F. Vendrell-Herrero, E. Gomes, & L. Zejnilovic (Eds.), *11th International Conference on Business Servitization (ICBS 2024—Lisbon)* (pp. 70–77). OmniaScience.

Proctor, R. W., & Dutta, A. (1995). *Skill acquisition and humam performance*. Sage.

Puccio, G. J., & Cabra, J. F. (2012). Idea generation and idea evaluation. In M. D. Mumford (Ed.), *Handbook of organizational creativity* (pp. 189–215). Elsevier.

Roediger, H. L., III., Weinstein, Y., & Agarwal, P. K. (2010). Forgetting: Preliminary considerations. In S. Della Sala (Ed.), *Forgetting* (Vol. 1, pp. 1–22). Psychology Press.

Roling, W., Schüffler, A. S., Thim, C., Grum, M., Gronau, N., & Kluge, A. (2023). Human behavior in the context of continuous change—An exploratory analysis in a research and application center industry 4.0. In T. Bui (Ed.), *Proceedings of the 56th Hawaii International Conference on System Science* (pp. 4828–4837). AIS Electronic Library (AISeL). https://hdl.handle.net/10125/103221

Roling, W. M., Grum, M., Gronau, N., & Kluge, A. (2026). *The impact of change readiness and task convenience on adaptive performance—A field study in the public sector*. Zeitschrift für Arbeits- und Organisationspsychologie, 70(1), 29–38.

Roling, W. M., Grum, M., Gronau, N., & Kluge, A. (2024a). The roots of errors in adaptive performance: Clustering behavioral patterns after the introduction of a change. *Journal of Workplace Learning,36*(4), 267–281.

Roling, W. M., Grum, M., Gronau, N., & Kluge, A. (2024b). You cannot have both—Two different ways of change implementation and their effects on adaptive performance and frustration. *International Journal of Industrial Ergonomics,104*, 103669.

Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review, 112*(3), 610–628.

Schüffler, A. S., Thim, C., Haase, J., Gronau, N., & Kluge, A. (2020). Information processing in work environment 4.0 and the beneficial impact of intentional forgetting on change management. *Zeitschrift für Arbeits- und Organisationspsychologie A&O, 64*(1), 17–29.

Schultz, P. W., & Searleman, A. (2002). Rigidity of thought and behavior: 100 years of research. *Genetic, Social, and General Psychology Monographs, 128*(2), 165.

Sharma, S., & Lenka, U. (2022a). Counterintuitive, yet essential: Taking stock of organizational unlearning research through a scientometric analysis (1976–2019). *Knowledge Management Research & Practice,20*(1), 152–174.

Sharma, S., & Lenka, U. (2022b). On the shoulders of giants: Uncovering key themes of organizational unlearning research in mainstream management journals. *Review of Managerial Science,16*(6), 1599–1695.

Simon, H. A., & Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist, 26*(2), 145–159.

Töytäri, P., Turunen, T., Klein, M., Eloranta, V., Biehl, S., & Rajala, R. (2018). Aligning the mindset and capabilities within a business network for successful adoption of smart services. *Journal of Product Innovation Management, 35*(5), 763–779.
Tsang, E. W., & Zahra, S. A. (2008). Organizational unlearning. *Human Relations, 61*(10), 1435–1462.
Volberda, H. W., Khanagha, S., Baden-Fuller, C., Mihalache, O. R., & Birkinshaw, J. (2021). Strategizing in a digital world: Overcoming cognitive barriers, reconfiguring routines and introducing new organizational forms. *Long Range Planning, 54*(5), 102110.
Ward, T. (2012). Problem solving. In M. D. Mumford (Ed.), *Handbook of organizational creativity* (pp. 169–188). Elsevier.
Wilson, C., & Mergel, I. (2022). Overcoming barriers to digital government: Mapping the strategies of digital champions. *Government Information Quarterly, 39*(2), 101681.
Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology, 55*, 235–269.
Yang, K. P., Chou, C., & Chiu, Y. J. (2014). How unlearning affects radical innovation: The dynamics of social capital and slack resources. *Technological Forecasting and Social Change, 87*, 152–163.
Yi, S., Knudsen, T., & Becker, M. C. (2016). Inertia in routines: A hidden source of organizational variation. *Organization Science, 27*(3), 782–800.
Zhao, Y., Lu, Y., & Wang, X. (2013). Organizational unlearning and organizational relearning: a dynamic process of knowledge management. *Journal of Knowledge Management, 17*(6), 902–912.

Dare2Del—An Explanatory Interactive Learning Assistant to Support Individual Strategies to Identify Irrelevant Digital Objects

Ute Schmid, Kyra Göbel, Olivia Wilczewski, and Cornelia Niessen

1 Introduction

The growing number of digital objects—digital hoarding—results in rising costs for companies, negative environmental effects, and can also impair individual employees who often feel overwhelmed by the information overload due to the ever growing amount of files and emails (Arnold et al., 2023; George, 2024; Schmid & Niessen, 2022). In private as well as in working life, a large amount of files stored on personal desktops as well as in the cloud are outdated and are never again opened. A central challenge of digital hoarding on a personal level is distinguishing relevant from irrelevant information, filtering out irrelevant information, and developing efficient self-regulating work strategies. Cognitive inhibitory control, the ability to suppress irrelevant information, plays a crucial role in this (MacLeod, 2007).

Structuring and systematically deleting irrelevant information can reduce distractions and enable focused work (Dabbish & Kraut, 2006). However, such strategies are often associated with uncertainty and not systematically applied (Göbel & Niessen, 2025; Niessen et al., 2019). Especially in work contexts, employees fear that decisions to delete might be wrong and cannot be revoked. Research has further shown that individuals who struggle to inhibit distractions and manage their attention are more likely to experience reduced well-being, cognitive overload, and

U. Schmid (✉) · O. Wilczewski
University of Bamberg, Bamberg, Germany
e-mail: ute.schmid@uni-bamberg.de

O. Wilczewski
e-mail: olivia.wilczewski@uni-bamberg.de

K. Göbel · C. Niessen
Friedrich-Alexander-University Erlangen, Erlangen, Germany
e-mail: kyra.goebel@fau.de

C. Niessen
e-mail: cornelianiessen@fau.de

J. O. Berndt et al. (eds.), *Intentional Forgetting with Intelligent Systems*, Progress in IS,
https://doi.org/10.1007/978-3-032-17621-9_2

reduced performance (Feliu-Soler et al., 2019; Göbel & Niessen, 2021; Toh & Yang, 2024). For these individuals, assistive AI technologies can provide significant support, enhancing the person-environment fit by accommodating cognitive limitations and optimizing task demands (Vleugels et al., 2023).

A crucial requirement for an assistive AI tool which can support employees to identify irrelevant digital objects is that it is perceived as trustworthy (Schmid, 2024). Core requirements for trustworthy AI systems are that they are performant and robust, transparent and explainable and that they allow human oversight and control. To address these requirements, we designed and realized an intelligent assistant Dare2Del (Siebers et al., 2017) which combines knowledge representation and machine learning (Deng et al., 2020; Gromowski et al., 2020) to allow for compliance with regulations and at the same time can adapt to individual user preferences. Furthermore, the system classifications of irrelevance of specific files are made transparent to the user by explaining the features which have been used to classify a file as irrelevant (Atzmueller et al., 2024). Finally, the user can correct system classifications as well as explanations (Schmid, 2022). Both types of information are used for adaption of the machine learned model.

In the following, we first introduce the general architecture of Dare2Del. Afterwards, we present details of our approach to explainable interactive machine learning. We present results of an experiment, investigating how users can benefit from such an assistive system in terms of performance and mental load. We conclude with a summary of main results and an outlook.

2 Architecture of Dare2Del

Dare2Del is an intelligent assistant which supports users to identify irrelevant files. We assume that users profit from support of this data management task because it can reduce cognitive load as well as anxiety involved in the decision that files are irrelevant and therefore can be deleted. The core of Dare2Del is realized in Prolog (Sterling & Shapiro, 1994). Files are classified by rules such as

```
irrelvant(A) :- samePrefix(A,B), sameFolder(A,B), older(A,B).
protect(A)   :- isContract(A), endtime(A,T1), currenttime(T2), T1 > T2.
```

where the head of a rule is given by a predicate which can evaluate to true or false for a given instantiation of the variable. The predicate *irrelevant* is true—that is, classifies a file as irrelevant—if the first part of the file name is identical to the name of another file, both files are in the same folder, and the file under consideration is older than the other file. Typically, a predicate like *irrelevant* is defined by a set of rules and the predicate is evaluated to true if one of the rules can be evaluated to true. The rule set for predicate *irrelevant* can be modified by machine learning, specifically by Inductive Logic Programming (ILP).

ILP is a family of approaches to learn Prolog programs from examples (Cropper & Dumančić, 2022; Muggleton & De Raedt, 1994). Currently, we use the ILP system Aleph (Srinivasan, 2001). Alternatively to Prolog, answer set programming (ASP) could be considered (Brewka et al., 2011) where the *irrelevant* rule set could be modelled as a default with exceptions. However, there is only limited work on learning ASP programs (Law et al., 2014). An ILP algorithm which allows to learning default rules with exceptions has been proposed by Siebers and Schmid (2018).

For our domain of application, examples are files which are classified as being irrelevant (positive training example) or not irrelevant (negative training example). Information about these files is given as background knowledge. In Dare2Del we use meta-information of files, such as file name, name of the folder in which the file is located, size, creation time, last access time, and file type. This information can be used in the body of rules. Additionally, definitions of relations, such as *older*, are provided as background knowledge.

The predicate *protect* specifies characteristics of files which should be protected from deletion. For instance, it should be guaranteed that contracts which are still running are never be considered as possibly irrelevant. The *protect* rules are predefined, based on regulations and general policies of a specific company. In contrast to the irrelevancy predicate, the rule set characterizing which files are protected cannot be changed by learning.

A general overview of the Dare2Del system is given in Fig. 1. The core of the system is the ILP learned model for classifying files as irrelevant or not. The model is generalized from training data which are augmented with background knowledge. Initially, a first, plausible set of irrelevancy rules might be given. An arbitrary set of files can be classified according to these rules. Each irrelevancy classification is checked against the protection rules and files for which the *protected* predicate holds are filtered out. The remaining files are then presented to the user. In addition, for each file, an explanation why the file is considered to be irrelevant is generated (see Sect. 3). The user can now correct the label of each presented file according to their personal preferences.

For some files, the decision of the user whether this file is irrelevant or not can be obvious. For other files, the user might be uncertain. In these cases, Dare2Del supports decision making by (1) showing the file in the file system browser, allowing the user to see the file in context and to inspect its content, and (2) providing two forms of explanations. The first explanation is a verbal description of the instantiated body of that irrelevance rule which evaluated to true for this file (Siebers & Schmid, 2019). The second explanation is a verbal description of a file which is similar to the current file with respect to its meta-information but which is not classified as irrelevant with respect to the current rule set (Rabold et al., 2022). Such near miss explanations have been empirically shown as helpful in the domain of concept learning (Markman & Gentner, 1993) and have also been proposed as efficient strategy in the context of rule learning (Winston, 1970).

The number of files which are presented to a user at one time can be determined by the user. Furthermore, Dare2Del allows for different strategies to select which files are to be classified at one time. For instance, files in the same or neighbouring folders

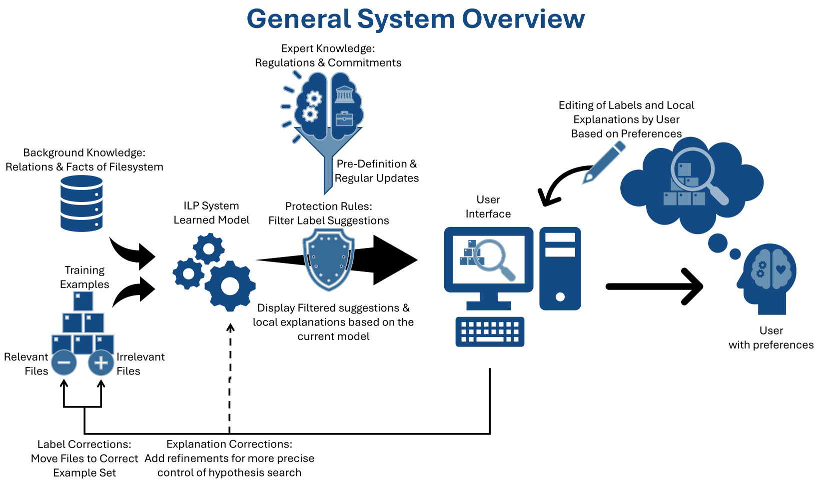


Fig. 1 Overview of the Dare2Del intelligent assistant

of the last used file might be prioritized. For more efficient learning, it is helpful to not only present files which are classified as irrelevant but also such which are classified as relevant. Currently, Dare2Del only relies on meta-information of files. In addition, similarity of content of text (Pradhan et al., 2015) or images (Shrivastava et al., 2011) might be considered. That is, a file might be classified as irrelevant if it has a similar content to another file but has an older creation time and probably smaller file size. However, incorporating similarity matching algorithms makes the approach more complex with respect to computational effort and also introduces another layer of uncertainty. Whether a text, table, or image is similar to another one and therefore might be irrelevant is a non-trivial decision problem which would require to extend the explainability methods currently realized in Dare2Del (Landthaler et al., 2018; Plummer et al., 2020).

The user interface of Dare2Del is shown in Fig. 2. DareDel is shown on the right hand as an extension to the file system browser. On top the user can select how many files they want to examine. In this case, three files which are classified as irrelevant and three files which are classified as relevant by the current ILP model are presented. The user can keep or change the classification. Currently, the first file is selected for closer inspection. It is highlighted in the file system browser and explanations for the current irrelevancy rating are shown.

3 Explainable Interactive Machine Learning with ILP

Research in explainable artificial intelligence (XAI) has resulted in a growing number of approaches to make machine learned models more transparent (Atzmueller et al., 2024; Schwalbe & Finzel, 2024). Most approaches address post hoc explanations for

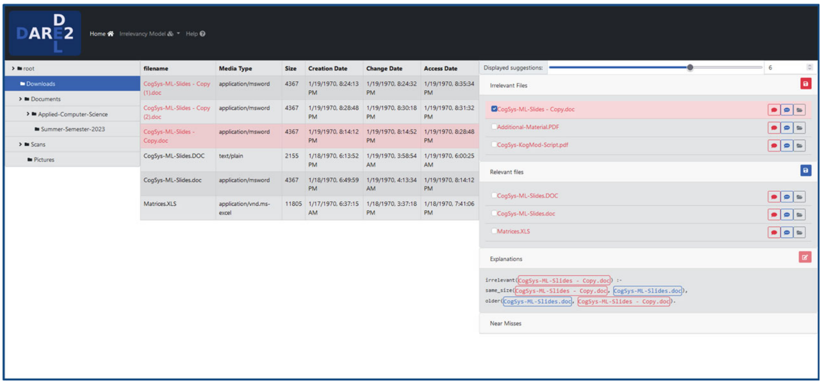


Fig. 2 Dare2Del user interface

black-box classifiers such as deep neural networks. In contrast, models learned with ILP are in principle interpretable by humans because they are represented in symbolic form (Muggleton et al., 2018; Rudin, 2019). However, in practice, interpretable models also might require additional explanation mechanisms—either because the model is too complex to make inspection easy or because the person interacting with the model is an non-expert user. For this reason, we incorporated two explanation methods in Dare2Del: a template-based method for translating a specific irrelevancy rule into a verbal description (Siebers & Schmid, 2019) and a method to create near miss explanations (Rabold et al., 2022).

A learned model for irrelevancy classification of files consists of a set of rules defining the predicate *irrelevant(A)*. To classify a file, the variables in the head and body of a rule are instantiated with the file specific information (substitution Θ of variables by constants). Such a rule is also called a ground clause (Muggleton & De Raedt, 1994). If the rule evaluates to true, the instantiated rule and possibly additional rules defined for predicates occurring in the body of this rule (the trace of the classification) constitute the explanation for the classification decision. Such an explanation of a specific instance is called a local explanation in XAI (Schwalbe & Finzel, 2024).

For ILP, a local explanation can be defined based on a given clausal theory T, which here is a set of Prolog rules where each rule corresponds to a clause (a disjunction of predicates). Each Prolog rule consists of a rule head and a rule body. The head of a rule for our application is the target predicate *irrelevant(A)*. Theory T can consist of several rules defining the target predicate where each rule body defines specific conditions for which the target predicate can be true. Rules are defined over variables as shown above. A rule can be instantiated by replacing variables with specific values which can be described by a substitution θ. For instance, in rule *irrelevant(A)* shown above, variable A can be replaced by the name of a specific file and the variable B used in the body predicate *samePrefix(A,B)* can be replaced by the name of another file. With these constituents, a local explanation can be defined as (Rabold et al., 2022):

Definition 1 (Local Explanation) A local explanation for a positive example P is a ground clause $C\theta$ where $C \in T$ such that $P = \text{head}(C\theta)$ and $T \models \text{body}(C\theta)$.

An example of such a local explanation is shown in Fig. 2. This rule can be rewritten into a natural language sentence (Siebers & Schmid, 2019). In the current version of Dare2Del, we present the Prolog rule with highlights of its specific components which the user can edit by interaction.

The local explanation should be helpful to the user to comprehend why the ILP model classified a specific file as irrelevant. For better understanding the scope of the learned classification rules, it also can be helpful for the user to see what kind of similar instance (in our case file) would not be classified as irrelevant by the model. Therefore, we also provide a near miss explanation (Rabold et al., 2022):

Definition 2 (Near Miss Explanation) Let $C\theta$ be a local explanation, C' a minimally changed clause, and θ' a minimal substitution. We call $C'\theta'$ a near miss explanation if $T \models \text{body}(C'\theta')$ and $T \not\models \text{head}(C'\theta')$. Using the operator Δ to mark near miss examples, ΔL is a near miss example if $L = \text{head}(C'\theta')$.

That is, we compute a minimal change of the local explanation for which an input instance will not be classified as belonging to the concept "irrelevant file." The algorithm for generating near miss explanations together with proofs of complexity, termination, and correctness is given in (Rabold et al., 2022). An example for a near miss explanation is given in Fig. 2.

XAI methods focus on explaining to better understand why a learned model gives a specific output. In the context of the AI assistant Dare2Del, explanations are additionally helpful to support the user to revise the model by interaction (Finzel et al., 2024). Accordingly, novel methods for explanatory interactive learning have been proposed (Slany et al., 2024; Teso & Kersting, 2019). These methods extend interactive machine learning for model revision by label correction (Berg et al., 2019; Fails & Olsen Jr., 2003) with the additional opportunity to correct the explanation. An explanation might uncover that the model has been overfitted to the training data (Lapuschkin et al., 2019) and correction of explanations can give guidance for model adaptation in cases where the model computes the right class but for the wrong reasons (Teso & Kersting, 2019). In the context of Dare2Del, corrections of explanations are used for cases where the model returns a classification which is contradictory to the user preferences. That is, explanation corrections are given together with label corrections. We also apply this strategy for text classification (Kiefer et al., 2022) and for image classification tasks in medical (Schmid & Finzel, 2020; Slany et al., 2022) as well as industrial (Herchenbach et al., 2022) applications.

There are two strategies for revision of ILP models with explanatory interactive machine learning: First, all data for which labels have been confirmed or corrected by the user can be used to train the model again from scratch making use of the corrected explanations to constrain rule induction. This approach has been realized for health-related classification tasks by Schmid and Finzel (2020) and also been explored for Dare2Del. The ILP system Aleph (Srinivasan, 2001) offers the possibility to include constraints to guide model induction. Another possibility, which we are currently investigating, is to incrementally modify the learned rules based on the corrections given by the user. This strategy has been already proposed for revision of propositional rules (Coste-Marquis, & Marquis, 2023). Here the main challenge is to identify possible inconsistencies in the model introduced by such modifications. A first approach for an interactive system for exploring and editing logic-based machine learning models has been proposed by Deane and Ray (2023).

The current version of Dare2Del incorporates the generation of local explanations and near miss explanations based on meta-data of files. Explainable interactive machine learning is realized by a preliminary approach to incremental model revision based on user input concerning label changes as well as modifications of the given local explanations.

4 Empirical Evaluation of Dare2Del

In the work context, the ability to control and inhibit distracting or unwanted thoughts is crucial for maintaining focus and meeting job demands. Employees need to manage both physical distractions (e.g. unstructured work environments) and mental distractions (e.g. off-task thoughts) in order to effectively focus on their goals and remain productive (Astle & Scerif, 2009; Ma et al., 2017). The assistance AI tool Dare2Del described above has been designed to support employees in managing physical distractions resulting from an ever growing number of irrelevant files. We assume that the explanatory interactive machine learning approach of Dare2Del has positive effects on the trust users have in file deletion recommendations. While Dare2Del is designed to support all users, it can be assumed that specially such users profit from Dare2Del which have low cognitive inhibitory control.

Cognitive inhibitory control, the ability to suppress activated, but outdated or inappropriate representations or processes (MacLeod, 2007; Wessel & Anderson, 2024), is a fundamental mechanism for adaptive behaviour and cognition that allows individuals to direct their attention to relevant tasks while filtering out distractions. In this sense, cognitive inhibitory control strategies complement physical activities such as sorting or filing documents by not only removing potentially distracting cues but also enabling the mental discarding or organizing of unnecessary, outdated, or false information (Niessen & Lang, 2021; Niessen et al., 2020; Nørby, 2015). However, there are individual differences in cognitive inhibitory control that can influence the effectiveness of these strategies in the workplace. While some individuals are successful in banishing distracting objects, others may struggle with controlling and navigating through them, which can hinder their focus on work tasks. These differences can be explained by variations in underlying executive control mechanisms, cognitive abilities such as working memory capacity, and prior experience in managing distractions (Anderson & Green, 2001; Anderson et al., 2004; Levy & Anderson, 2008). Relying on the idea that cognitive inhibitory control is an important and functional mechanism in navigating unstructured work environments, we derive the following hypotheses:

H1: Cognitive inhibitory control is positively related to task performance.
H2: Cognitive inhibitory control is negatively related to mental effort.

The concept of **Person-Environment fit** (PE fit) (Caplan, 1987; Caplan & Van Harrison, 1993; Kristof-Brown & Guay, 2011; Vleugels et al., 2023) describes the extent to which an individual's needs align with environmental conditions. Two key aspects are (1) supplementary fit, referring to the alignment between person and environment in terms of traits and values, and (2) complementary fit, where the environment complements an individual's needs and abilities (Muchinsky & Monahan, 1987). Research indicates that individuals with lower cognitive inhibitory control may find it more challenging to ignore irrelevant information and focus on core tasks (Hasher et al., 2007; Munakata et al., 2011; Van Moorselaar & Slagter, 2020; Wessel & Anderson, 2024). In unstructured environments, assistive systems that guide attention and actively help to regain focus—for example, by suggesting to delete specific irrelevant data like Dare2Del—can offer great benefits. Following the idea of complementary PE fit (Muchinsky & Monahan, 1987), these systems could improve the fit between individual abilities and environmental demands, thereby reducing cognitive load and enabling higher task performance. More specifically, we anticipate that individuals with high cognitive inhibitory control will find it easier to navigate unstructured environments and implement their own structuring and deletion strategies. In contrast, individuals with low cognitive inhibitory control will benefit more from assistive systems like Dare2Del, which are designed to guide them through unstructured digital environments. The system's deletion support should adapt the task demands to the user's cognitive abilities, rather than adding an additional burden:

H3: Compared to individuals with high cognitive inhibitory control, individuals with low cognitive inhibitory control benefit more from deletion support in terms of task performance.
H4: Compared to individuals with high cognitive inhibitory control, individuals with low cognitive inhibitory control benefit more from deletion support in terms of mental effort.

4.1 Method

Participants and Procedure. The sample comprised 94 students from different academic disciplines with a mean age of 22.13 years ($SD = 3.30$). Seventy-five participants reported being female, and 19 described themselves as male. They received course credits or 30 Euro for participation. The study consisted of two laboratory sessions with a one-week interval. In the first session, we assessed participants' individual levels of cognitive inhibitory control using the Think/No-Think task (TNT) (Anderson & Green, 2001). In the second session, participants had to complete a complex, digital medical data review task. Finally, they answered a short questionnaire containing demographic information and brief questions about their experiences in the experiment.

Think/No-Think Task (TNT). The TNT is a structured laboratory paradigm designed to assess retrieval suppression, a form of cognitive inhibition that helps control unwanted thoughts and memories (Anderson & Green, 2001; Nardo & Anderson, 2024). It is conducted in individual test sessions and lasts approximately two hours per participant. In the experiment's first phase, participants study 58 weakly associated cueâŁ“response word pairs (6000 ms). After assessing how well these have been learned, they enter the main phase of the experiment: In this phase, cue words are either presented in green (think) or red (no-think) for 3000 ms. Depending on the colour, participants are instructed rehearse (green) or suppress (red) the associated response word. Some of the words do not appear during this phase, serving as control items for passive forgetting (baseline). In the final recall phase, all word pairs are then tested again—regardless of prior instructions (for more details of the procedure see Göbel & Niessen, 2025).

The suppression effect then is calculated by comparing recall performance across conditions. Specifically, retrieval suppression is inferred from the difference in recall rates between the no-think and baseline conditions: Successful suppression is indicated by reduced recall for no-think items relative to baseline items. This suppression-induced forgetting provides a measure of cognitive inhibition, reflecting participants' ability to intentionally suppress unwanted memories.

Evaluation of Dare2Del: Medical Data Review Task. In their second laboratory appointment, participants completed a medical data review task that lasted about an hour. It was designed to simulate real-world challenges in a complex administrative setting, where the participants needed to identify relevant information, disregard irrelevant details, and make informed choices under time constraints. Thereby, the environment was intentionally unstructured to mirror the information overload often experienced in digital work settings. The task was designed as a three-level between-person experiment. Participants were therefore assigned to one of three experimental conditions: (1) A control condition where they received no instruction regarding deleting files ($n = 29$), (2) a self-initiated deletion condition where they were encouraged to structure the digital environment and delete irrelevant data by themselves ($n = 32$), or (3) an AI-assisted deletion condition where the deletion support system Dare2Del was introduced and provided specific deletion suggestions at four timepoints during the experiment ($n = 33$). Nevertheless, the final acceptance and implementation of these deletion suggestions was always up to the participants. However, to ensure that participants actually had to engage with the unstructured file system, the search function was disabled. The cover story involved reviewing medical data for a hospital. Participants had to process 24 trials that required making decisions based on specific patient information and medication rules—both of which they had to identify and evaluate in an unstructured digital environment. For example, they were asked to evaluate Patient A's medication according to rule set X—whereby each rule set comprised four distinct medication rules. Participants then had to search the medication and the rule set documents in the unstructured system, compare and cross-check them before entering their response. As each trial encompassed four rules, participants processed a total of $24 \times 4 = 96$ evaluations.

Table 1 Means, standard deviations, and correlations for all study variables

Variable	*M*	*SD*	1	2	3	4	5	6
1. Age	22.13	3.30						
2. Gender (f = 0, m = 1)	0.20	0.40	0.25*					
3. Cognitive inhibitory control	0.07	0.13	0.02	−0.12				
4. Number of deleted files	7.29	13.28	−0.01	−0.12	−0.05			
5. Task performance: accuracy	92.59	3.59	0.05	−0.06	0.23*	0.00		
6. Task performance: speed (in s)	2123.87	368.94	−0.10	−0.07	−0.02	0.12	−0.03	
7. Mental effort	3.00	0.58	−0.05	−0.02	0.08	−0.13	−0.08	0.15

Notes $N = 94$; $^*p < 0.05$ (two-sided)

To induce time pressure, participants were instructed that it was important to work quickly: The fastest 10% of participants were promised a performance bonus of 20 Euro.

The following **measures** have been assessed:

- **File Deletion.** We assessed the number of files participants had actually deleted during the task by analysing the digital trash bins.
- **Task Performance.** Task performance was assessed by evaluating (a) the accuracy and (b) the processing speed of task completion. Specifically, accuracy was measured by the number of correct vs. incorrect medication evaluation decisions across the 24 trials. As each trial encompassed four decisions, the maximum score was $4 \times 24 = 96$. Processing speed was measured by the total time participants took to complete the 24 trials. Any time required for restructuring and deleting activities was not included in the calculations.
- **Mental Effort.** Mental effort was assessed in the questionnaire administered after task completion. Participants evaluated the item "During the task, I had to put in a lot of effort" (adapted from Zijlstra et al., 1990) on a five-point scale ranging from 1 = *strongly disagree* to 5 = *strongly agree.*

4.2 Results

Table 1 shows the means, standard deviations, and correlations between all study variables.

Think/No-think Task. The final recall rates for previously learned word pairs were 0.98 (*SD* = 0.04) for think, 0.87 (*SD* = 0.14) for no-think, and 0.94 (*SD* = 0.08) for baseline words. To check whether suppression generally worked, we

Table 2 Descriptive statistics for file deletion, accuracy, processing speed, and mental effort across three conditions

	Control condition[1]		Self-initiated deletion[2]		AI-assisted deletion[3]	
	M	*SD*	*M*	*SD*	*M*	*SD*
File deletion	0.00	0.00	10.56	15.79	10.51	14.19
Accuracy	92.00	3.59	92.66	3.64	92.85	3.63
Processing speed (in s)	2044.75	340.12	2197.90	384.31	2121.60	374.34
Mental effort	2.72	1.12	2.91	0.96	2.91	0.98

Notes [1]$n = 29$; [2]$n = 32$; [3]$n = 33$

used a repeated-measure analysis of variance (ANOVA). We found significant differences between these conditions, which account for a general suppression-induced forgetting effect, $F(2, 145) = 45.97$, $p < 0.001$, $\eta_p^2 = 0.34$. Consecutive contrast analyses showed a significant difference between no-think and baseline items (mean difference $= 0.07$, $F(1, 91) = 25.50$, $p < 0.001$, $\eta_p^2 = 0.22$), which confirms that participants were generally able to actively suppress word associations for no-think items beyond the recall threshold. Moreover, think items were significantly better recalled than baseline items (mean difference $= 0.04$, $F(1, 91) = 33.50$, $p < 0.001$, $\eta_p^2 = 0.27$). This suggests that strengthening the association during the think/no-think phase resulted in additional learning. To obtain a measure for individual cognitive inhibitory control, we subtracted no-think from baseline recall rates ($M = 0.07$, $SD = 0.13$). Thus, higher values indicate higher inhibitory control.

Medical Data Review Task. Across conditions, participants deleted 7.29 files ($SD = 13.28$) on average, achieved a medium accuracy 92.59 ($SD = 3.59$), needed around 35 minutes to complete the evaluation of the medical data ($M = 2123.97$ seconds; $SD = 368.94$), and reported a mental effort of 2.85 ($SD = 1.04$). The values separated by experimental condition can be found in Table 2.

Hypothesis Testing to test H1 and H2, which assumed direct linear effects of cognitive inhibitory control on task performance and mental effort across experimental conditions, we calculated linear regression analyses. The predictor was z-transformed before being entered into the analyses. Results show a significant positive effect of cognitive inhibitory control on accuracy, $\beta = 0.23$, $t(92) = 2.28$, $p = 0.025$, $R^2 = 0.05$. However, there was no effect on processing speed, $\beta = -0.02$, $t(92) = -0.21$, $p = 0.834$, $R^2 = 0.00$. Moreover, cognitive inhibitory control was not able to predict mental effort, $\beta = 0.08$, $t(92) = 0.75$, $p = 0.454$, $R^2 = 0.01$. Thus, H1 could be confirmed partly, whereas H2 was not confirmed. To test H3 and H4, which assumed that individuals with low cognitive inhibitory control benefit more from deletion support in terms of task performance and mental effort than individuals with high cognitive inhibitory control, we again conducted hierarchical linear regression analyses. In the first step, we entered the z-transformed predictors cognitive inhibitory control, deleting behaviour (number of deleted files), and experimental condition (self-initiated deletion coded as 0 and AI-assisted deletion coded as 1), as well as their interactions (second step: three two-way interactions, third step: one three-way-interaction). Participants in the control condition were not included in the analysis, as they did not delete any files (thus, there was no variance in this variable). Therefore, the sample size for these analyses was 65. To examine how the combination of cognitive inhibitory control, deleting behaviour, and experimental condition influenced task performance and mental effort, we focused on the three-way interactions in the interpretation of results. The full regression models are depicted in Table 3 .

With regard to task performance, the three-way interaction showed no effect either for accuracy, $\beta = 0.19$, $t(63) = 0.97$, $p = 0.335$, $\Delta R^2 = 0.02$, or for processing speed, $\beta = 0.17$, $t(63) = 0.89$, $p = 0.378$, $\Delta R^2 = 0.01$, not confirming H3. However, the three-way interaction did support our considerations in terms of mental effort, $\beta = 0.44$, $t(65) = 2.43$, $p = 0.018$, $\Delta R^2 = 0.09$, indicating that individuals with low cognitive inhibitory control benefitted more from deleting with deletion support, whereas those with high cognitive inhibitory control benefitted more from self-initiated deletion (see Fig. 3). H4 was thus confirmed.

4.3 Discussion

Our findings demonstrate that individuals with high cognitive inhibitory control consistently perform better in disorganized environments, regardless of the degree or type of deletion support: They achieved higher accuracy than individuals with lower inhibitory control. This suggests that high-ability individuals can effectively

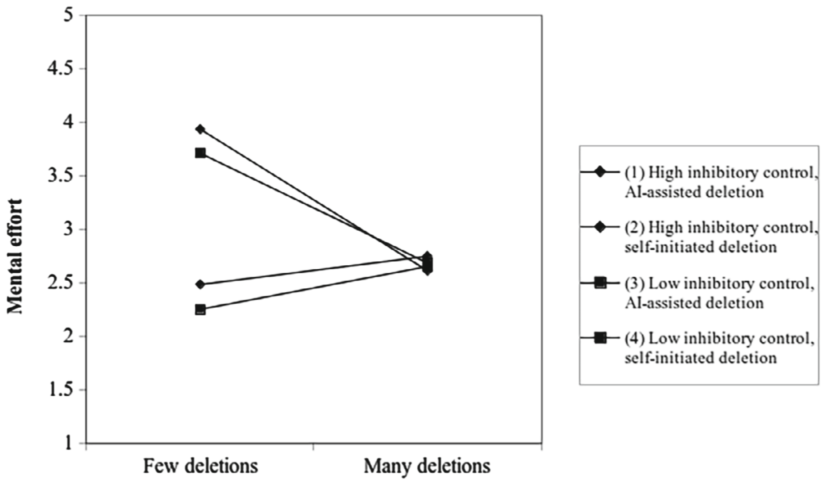


Fig. 3 Interaction plot showing the three-way interaction cognitive inhibitory control × number of deleted files × experimental condition on mental effort

Table 3 Results of hierarchical linear regression analyses testing H3 and H4

	Task performance: accuracy				Task performance: processing speed				Mental effort			
Predictor	*B*	*SE*	*β*	*t*	*Est.*	*SE*	*β*	*t*	*Est.*	*SE*	*β*	*t*
Constant	92.58	0.68		135.26***	2206.17	70.77		31.17***	2.86	0.17		16.85***
Cognitive inhibitory control	1.24	0.86	0.26	1.44	−30.70	89.20	−0.06	−0.34	0.41	0.21	0.32	1.91
Number of deleted files	−0.12	0.61	−0.04	−0.20	−9.23	63.00	−0.03	−0.15	−0.23	0.15	−0.26	−1.49
Experimental condition[1]	0.20	0.96	0.03	0.21	−107.42	99.40	−0.14	−1.08	0.04	0.24	0.02	0.19
Cognitive inhibitory control × number of deleted files	−0.38	0.86	−0.09	−0.45	27.42	88.46	0.06	0.31	−0.43	0.21	−0.36	−2.01
Cognitive inhibitory control × experimental condition[1]	−1.46	1.33	−0.20	−1.10	23.78	137.73	0.03	0.17	−0.70	0.33	−0.37	−2.13*
Number of deleted files × experimental condition[1]	0.16	0.87	0.03	0.18	79.71	89.76	0.16	0.89	0.04	0.22	0.03	0.18
Cognitive inhibitory control × number of deleted files × experimental condition[1]	1.21	1.25	0.19	0.97	114.30	128.78	0.17	0.89	0.75	0.31	0.33	2.43

Notes $N = 65$; [1]0 = self-initiated deleting condition, 1 = AI-assisted deleting condition; $^{***}p < .001$, $^{*}p < .05$

navigate disorganization, making them less dependent on external structuring mechanisms. However, regarding the different levels of external structuring, results align with the complementary fit (Muchinsky & Monahan, 1987) perspective: Individuals with lower cognitive inhibitory control experienced lower mental effort when they received AI-assisted deletion support, whereas those higher in cognitive inhibitory control experience lower mental effort when they self-structured their file systems. This suggests that external structuring can compensate for internal difficulties in navigating in and managing unstructured digital environments.

Interestingly, these interaction effects were found for mental effort but not for task performance, indicating that while structured assistance can help to reduce mental load, it does not necessarily translate into an immediate improvement of outcomes. The findings are particularly relevant for designing tailored cognitive support in work environments. They highlight that assistive systems should be adapted to individual needs rather than uniformly applied. By identifying who benefits from which type of support, employers can enhance efficiency without overloading employees cognitively.

There are, however, some limitations that need to be mentioned. One of them is study's rather small and specific student sample, which restricts both power and generalizability of the findings. Additionally, the lack of variance in the assistive system's support prevents conclusions about whether different forms or degrees of deletion support could have differential effects. Ceiling effects may have further influenced the results, as overall task performance was close to the maximum. This could have masked potential benefits of the support system.

Despite these limitations, the findings provide valuable insights for designing tailored cognitive support in work environments. They highlight that individual inhibitory control is a key predictor of task performance in unstructured environments, suggesting that targeted training of cognitive inhibition skills and strategies to minimize distractions could improve workplace efficiency (Anderson & Green, 2001; Soucek & Moser, 2010). When considering mental effort, the fit between the individual and the type of support emerges as a crucial factor: Individuals with high cognitive inhibitory control benefit more from self-initiated deletion—a finding which was also found by Göbel and Niessen (2025)—whereas those with lower inhibitory control experience reduced mental effort when using AI-based structuring support. This underscores the need for assistive systems that are flexible and adaptable to individual needs rather than applying a uniform solution. By identifying who benefits from which type of support, employers can optimize workplace efficiency without unnecessarily increasing cognitive strain. Future research should explore whether different structuring approaches influence not only mental effort but also long-term performance outcomes. Additionally, investigating how different levels of support complement not only the individual, but also different levels of environmental chaos could provide further insights into the conditions under which tailored support is most effective.

5 Conclusions and Outlook

We presented the intelligent assistive system Dare2Del which supports users to counteract digital hoarding. Dare2Del is designed as a human-AI partnership system. For many domains, hybrid intelligence (Dellermann et al., 2019) can result in more efficient problem solving by combining human reasoning and AI methods in such a way, that the resulting performance is higher than by human or AI system alone (Schmid, 2024; Schmid & Finzel, 2020). We could empirically show that the explanatory interactive approach of Dare2Del supports calibrated trust in file deletion recommendations (Göbel et al., 2022).

Dare2Del supports employees to identify irrelevant files which could be deleted. We could empirically show that especially persons with low cognitive inhibitory control can profit from DareDel by being relieved from mental load. Because deletion of files is a critical task, often associated with anxiety that a specific file might be needed in future, it is important that an assistive system supports humans in their decision-making process. Dare2Del realizes this by showing the files under consideration in context of the file system browser and by giving reasons why a file can be considered as irrelevant in form of rich explanations. It has been shown empirically that this kind of concept-level explanations can increase trust as well as performance (Thaler & Schmid, 2021).

Dare2Del has been designed with a focus on application in work contexts as an assistive tool to support intentional forgetting of digital objects. At the same time, it inspired the development of novel AI methods to realize human-AI partnership systems. The approach underlying Dare2Del combines knowledge representation and machine learning (Muggleton et al., 2018; Von Rueden et al., 2021). Furthermore, it provides rich explanations based on novel concept-based methods of explainable AI (Rabold et al., 2022; Ramaswamy et al., 2023). Finally, it realizes explainable interactive learning (Schmid, 2022; Teso & Kersting, 2019). All three aspects of Dare2Del contribute to the trustworthiness requirements addressed in the European AI Act (Laux et al., 2024; Schmid, 2024): Combining explicit knowledge representation with machine learning allows to incorporate safety criteria—in the case of Dare2Del a guarantee that specific files will never be classified as irrelevant. Explanations support transparency of the learned model and explanatory interactive machine learning is a way to realize human agency and oversight.

Acknowledgements We want to thank Michael Siebers for supporting our research during the first project period and Durgesh Nandini and Sebastian Seufert for their support during the second project period. Special thanks go to Sebastian Seufert for realizing the Dare2Del MockUp for the experiment.

References

Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature, 410*(6826), 366–369.

Anderson, M. C., Ochsner, K. N., Kuhl, B., Cooper, J., Robertson, E., Gabrieli, S. W., Glover, G. H., & Gabrieli, J. D. (2004). Neural systems underlying the suppression of unwanted memories. *Science, 303*(5655), 232–235.

Arnold, M., Goldschmitt, M., & Rigotti, T. (2023). Dealing with information overload: A comprehensive review. *Frontiers in Psychology, 14*, 1122200.

Astle, D. E., & Scerif, G. (2009). Using developmental cognitive neuroscience to study behavioral and attentional control. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology, 51*(2), 107–118.

Atzmueller, M., Fürnkranz, J., Kliegr, T., & Schmid, U. (2024). Explainable and interpretable machine learning and data mining. *Data Mining and Knowledge Discovery, 38*(5), 2571–2595.

Berg, S., Kutra, D., Kroeger, T., Straehle, C. N., Kausler, B. X., Haubold, C., Schiegg, M., Ales, J., Beier, T., Rudy, M., et al. (2019). Ilastik: Interactive machine learning for (bio) image analysis. *Nature Methods, 16*(12), 1226–1232.

Brewka, G., Eiter, T., & Truszczyński, M. (2011). Answer set programming at a glance. *Communications of the ACM, 54*(12), 92–103.

Caplan, R. D. (1987). Person-environment fit theory and organizations: Commensurate dimensions, time perspectives, and mechanisms. *Journal of Vocational Behavior, 31*(3), 248–267.

Caplan, R. D., & Van Harrison, R. (1993). Person-environment fit theory: Some history, recent developments, and future directions. *Journal of social Issues, 49*(4), 253–275.

Coste-Marquis, S., & Marquis, P. (2023). Rectifying binary classifiers. In K. Gal, A. Nowé, G. J. Nalepa, R. Fairstein, & R. Radulescu (Eds.), *The 26th European Conference on Artificial Intelligence (ECAI'23)* (pp. 484–492). IOS Press.

Cropper, A., & Dumančić, S. (2022). Inductive logic programming at 30: A new introduction. *Journal of Artificial Intelligence Research, 74*, 765–850.

Dabbish, L. A., & Kraut, R. E. (2006). *Email overload at work: An analysis of factors associated with email strain*. In Proceedings of the 20th Anniversary Conference on Computer Supported Cooperative Work (pp. 431–440).

Deane, O., & Ray, O. (2023). *Interactive model refinement in relational domains with inductive logic programming*. In Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23) (pp 127–129). ACM.

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. *Business & Information Systems Engineering, 61*(5), 637–643.

Deng, C., Ji, X., Rainey, C., Zhang, J., & Lu, W. (2020). Integrating machine learning with human knowledge. *Iscience, 23*(11), 101656.

Fails, J. A., & Olsen, D. R., Jr. (2003). *Interactive machine learning*. In Proceedings of the 8th International Conference on Intelligent User Interfaces (IUI 2003) (pp. 39–45). ACM.

Feliu-Soler, A., Pérez-Aranda, A., Montero-Marín, J., Herrera-Mercadal, P., Andrés-Rodríguez, L., Angarita-Osorio, N., Williams, A. D., & Luciano, J. V. (2019). Fifteen years controlling unwanted thoughts: A systematic review of the Thought Control Ability Questionnaire (TCAQ). *Frontiers in Psychology, 10*, 1446.

Finzel, B., Knoblach, J., Thaler, A., & Schmid, U. (2024). *Near hit and near miss example explanations for model revision in binary image classification*. In International Conference on Intelligent Data Engineering and Automated Learning (pp. 260–271). Springer.

Göbel, K., & Niessen, C. (2025). Deleting and sorting files is beneficial—But only when you fully let go. *Computers in Human Behavior Reports, 18*, 100646.

George, A. S. (2024). Digital hoarding: The rising environmental and personal costs of information overload. *Partners Universal Multidisciplinary Research Journal, 1*(2), 51–67.

Göbel, K., & Niessen, C. (2021). Thought control in daily working life: How the ability to stop thoughts protects self-esteem. *Applied Cognitive Psychology, 35*(4), 1011–1022.

Göbel, K., Niessen, C., Seufert, S., & Schmid, U. (2022). Explanatory machine learning for justified trust in human-AI collaboration: Experiments on file deletion recommendations. *Frontiers in Artificial Intelligence, 5*, 919534.

Gromowski, M., Siebers, M., & Schmid, U. (2020). A process framework for inducing and explaining datalog theories. *Advances in Data Analysis and Classification, 14*(4), 821–835.

Hasher, L., Lustig, C., Zacks, R. T., et al. (2007). Inhibitory mechanisms and the control of attention. *Variation in Working Memory, 19*, 227–249.

Herchenbach, M., Müller, D., Scheele, S., & Schmid, U. (2022). *Explaining image classifications with near misses, near hits and prototypes: Supporting domain experts in understanding decision boundaries*. In International Conference on Pattern Recognition and Artificial Intelligence (pp. 419–430). Springer.

Kiefer, S., Hoffmann, M., & Schmid, U. (2022). Semantic interactive learning for text classification: A constructive approach for contextual interactions. *Machine Learning and Knowledge Extraction, 4*(4), 994–1010.

Kristof-Brown, A., & Guay, R. P. (2011). Person–environment fit. In *APA handbook of industrial and organizational psychology, Vol. 3. Maintaining, expanding, and contracting the organization* (pp. 3–50). American Psychological Association.

Landthaler, J., Glaser, I., & Matthes, F. (2018). Towards explainable semantic text matching. In *Legal knowledge and information systems* (pp. 200–204). IOS Press.

Lapuschkin, S., Wäldchen, S., Binder, A., Montavon, G., Samek, W., & Müller, K. R. (2019). Unmasking Clever Hans predictors and assessing what machines really learn. *Nature Communications, 10*(1), 1096.

Laux, J., Wachter, S., & Mittelstadt, B. (2024). Trustworthy artificial intelligence and the European Union AI act: On the conflation of trustworthiness and acceptability of risk. *Regulation & Governance, 18*(1), 3–32.

Law, M., Russo, A., & Broda, K. (2014). *Inductive learning of answer set programs*. In 14th European Conference on Logics in Artificial Intelligence (JELIA 2014) (pp. 311–325). Springer.

Levy, B. J., & Anderson, M. C. (2008). Individual differences in the suppression of unwanted memories: The executive deficit hypothesis. *Acta Psychologica, 127*(3), 623–635.

Ma, A., Landau, M. J., Narayanan, J., & Kay, A. C. (2017). Thought-control difficulty motivates structure seeking. *Journal of Experimental Psychology: General, 146*(8), 1067.

MacLeod, C. M. (2007). The concept of inhibition in cognition. In C. M. MacLeod & D. S. Gorfein (Eds.), *Inhibition in cognition* (pp. 3–23). American Psychological Association.

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology, 25*(4), 431–467.

Muchinsky, P. M., & Monahan, C. J. (1987). What is person-environment congruence? Supplementary versus complementary models of fit. *Journal of Vocational Behavior, 31*(3), 268–277.

Muggleton, S., & De Raedt, L. (1994). Inductive logic programming: Theory and methods. *The Journal of Logic Programming, 19*, 629–679.

Muggleton, S. H., Schmid, U., Zeller, C., Tamaddoni-Nezhad, A., & Besold, T. (2018). Ultra-strong machine learning: Comprehensibility of programs learned with ILP. *Machine Learning, 107*, 1119–1140.

Munakata, Y., Herd, S. A., Chatham, C. H., Depue, B. E., Banich, M. T., & O'Reilly, R. C. (2011). A unified framework for inhibitory control. *Trends in Cognitive Sciences, 15*(10), 453–459.

Nardo, D., & Anderson, M. C. (2024). Everything you ever wanted to know about the think/no-think task, but forgot to ask. *Behavior Research Methods, 56*(4), 3831–3860.

Niessen, C., Göbel, K., Lang, J. W., & Schmid, U. (2020). Stop thinking: An experience sampling study on suppressing distractive thoughts at work. *Frontiers in Psychology, 11*, 1616.

Niessen, C., Göbel, K., Siebers, M., & Schmid, U. (2019). Time to forget. *Zeitschrift für Arbeits-und Organisationspsychologie (A&O), 64*(1), 30–45.

Niessen, C., & Lang, J. W. (2021). Cognitive control strategies and adaptive performance in a complex work task. *Journal of Applied Psychology, 106*(10), 1586.

Nørby, S. (2015). Why forget? On the adaptive value of memory loss. *Perspectives on Psychological Science, 10*(5), 551–578.

Plummer, B. A., Vasileva, M. I., Petsiuk, V., Saenko ,K., & Forsyth, D. (2020). *Why do these match? Explaining the behavior of image similarity models*. In 16th European Conference on Computer Vision (ECCV 2020) (pp. 652–669). Springer.

Pradhan, N., Gyanchandani, M., Wadhvani, R., et al. (2015). A review on text similarity technique used in IR and its application. *International Journal of Computer Applications, 120*(9), 29–34.

Rabold, J., Siebers, M., & Schmid, U. (2022). Generating contrastive explanations for inductive logic programming based on a near miss approach. *Machine Learning, 111*(5), 1799–1820.

Ramaswamy, V. V., Kim, S. S., Fong, R., & Russakovsky, O. (2023). *Overlooked factors in concept-based explanations: Dataset choice, concept learnability, and human capability*. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10932-10941).

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215.

Schmid, U. (2022). Interactive learning with mutual explanations in relational domains. In N. Chater & S. Muggleton (Eds.), *Human-like machine intelligence* (pp. 338–354). Oxford University Press.

Schmid, U. (2024). Trustworthy artificial intelligence—Comprehensible, transparent, correctable. In C. Ghezzi, J. Kramer, J. Nida-Rümelin, B. Nuseibeh, E. Prem, A. Stanger & H. Werthner (Eds.), *Introduction to digital humanism* (pp. 151–164). Springer.

Schmid, U., & Finzel, B. (2020). Mutual explanations for cooperative decision making in medicine. *KI-Künstliche Intelligenz, 34*(2), 227–233.

Schmid, U., & Niessen, C. (2022). Dare to delete more: An interactive-learning assistance system to remove irrelevant computer file. *German Research (DFG Magazine), 3*, 28–31.

Schwalbe, G., & Finzel, B. (2024). A comprehensive taxonomy for explainable artificial intelligence: A systematic survey of surveys on methods and concepts. *Data Mining and Knowledge Discovery, 38*(5), 3043–3101.

Shrivastava, A., Malisiewicz, T., Gupta, A., & Efros, A. A. (2011). Data-driven visual similarity for cross-domain image matching. *ACM Transactions on Graphics, 30*(6), 154.

Siebers, M., Gobel, K., Niessen, C., & Schmid, U. (2017). *Requirements for a companion system to support identifying irrelevancy*. In International Conference on Companion Technology (ICCT 2017) (pp. 1–2). IEEE.

Siebers, M., & Schmid, U. (2018). *Was the year 2000 a leap year? Step-wise narrowing theories with Metagol*. In 28th International Conference on Inductive Logic Programming (ILP 2018) (pp. 141–156). Springer.

Siebers, M., & Schmid, U. (2019). Please delete that! why should I? Explaining learned irrelevance classifications of digital objects. *KI-Künstliche Intelligenz, 33*(1), 35–44.

Slany, E., Ott, Y., Scheele, S., Paulus, J., & Schmid, U. (2022). *CAIPI in practice: Towards explainable interactive medical image classification*. In IFIP International Conference on Artificial Intelligence Applications and Innovations (pp. 389–400). Springer.

Slany, E., Scheele, S., & Schmid, U. (2024). *Explanatory interactive machine learning with counterexamples from constrained large language models*. In German Conference on Artificial Intelligence (Künstliche Intelligenz 2025) (pp. 324–331). Springer.

Soucek, R., & Moser, K. (2010). Coping with information overload in email communication: Evaluation of a training intervention. *Computers in Human Behavior, 26*(6), 1458–1466.

Srinivasan, A. (2001). *The Aleph manual*. University of Oxford.

Sterling, L., & Shapiro, E. Y. (1994). *The art of prolog: Advanced programming techniques*. MIT Press.

Teso, S., & Kersting, K. (2019). *Explanatory interactive machine learning*. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (pp. 239–245).

Thaler, A. M., & Schmid, U. (2021). *Explaining machine learned relational concepts in visual domains-effects of perceived accuracy on joint performance and trust*. In Proceedings of the 43rd Annual Meeting of the Cognitive Science Society (Vol. 43).

Toh, W. X., & Yang, H. (2024). Do cognitive resources matter for eudaimonia? The role of inhibitory control in psychological well-being. *Trends in Psychology, 32*(2), 525–540.

Van Moorselaar, D., & Slagter, H. A. (2020). Inhibition in selective attention. *Annals of the New York Academy of Sciences, 1464*(1), 204–221.

Vleugels, W., Verbruggen, M., De Cooman, R., & Billsberry, J. (2023). A systematic review of temporal person-environment fit research: Trends, developments, obstacles, and opportunities for future research. *Journal of Organizational Behavior, 44*(2), 376–398.

Von Rueden, L., Mayer, S., Beckh, K., Georgiev, B., Giesselbach, S., Heese, R., Kirsch, B., Pfrommer, J., Pick, A., Ramamurthy, R., et al. (2021). Informed machine learning—A taxonomy and survey of integrating prior knowledge into learning systems. *IEEE Transactions on Knowledge and Data Engineering, 35*(1), 614–633.

Wessel, J. R., & Anderson, M. C. (2024). Neural mechanisms of domain-general inhibitory control. *Trends in Cognitive Sciences, 28*(2), 124–143.

Winston, P. H. (1970). *Learning structural descriptions from examples* (Tech. rep., MIT AITR-231).

Zijlstra, F., Cavalini, P., Wiethoff, M., & Meijman, T. F. (1990). Mental effort: Strategies and efficiency. In *European perspectives in psychology* (Vol. 1, pp. 435–447). Wiley.

Objective and Subjective Switching Costs on Individual, Team, and Organizational Levels

Arnulf Schueffler and Annette Kluge

1 Introduction

Change means stopping practicing old behavior, especially routines, and quitting using structures, at least modifying or substituting them on individual, team, and organizational levels in a period of adaptation. At the end of this period, old behavior, routines, and structures should no longer be executed. They should be forgotten. Instruments of intentional forgetting can support those necessary processes of forgetting [24].

During this period of adaptation switching costs must be accepted. These costs arise due to uncertainty and conflicting new and old patterns [22]. Even though old patterns are no longer executed, they are still present in the minds of the people in charge. Here, old and new patterns cause conflicts until the old ones are forgotten. Solving those conflicts may lead to errors until the new, adapted patterns replace the old ones. This will lead to slower behavior during adaptation anyway.

Therefore, developing awareness and an understanding of such changing situations may enable the introduction of implications for reducing such switching costs on individual, team, and organizational levels, at least to save resources such as money and time but more importantly to reduce frustration and enhance feelings of success and achievement for the people undergoing a change. Instruments to reduce switching costs are methods of intentional forgetting. Measuring the switching costs can provide information about how successful such methods are, how far the adaptation process has already progressed, and where significant challenges remain.

A. Schueffler (✉) · A. Kluge
Chair of Industrial, Organizational and Business Psychology, Ruhr University Bochum, Bochum, Germany
e-mail: arnulf.schueffler@rub.de

A. Kluge
e-mail: annette.kluge@rub.de

J. O. Berndt et al. (eds.), *Intentional Forgetting with Intelligent Systems*, Progress in IS,
https://doi.org/10.1007/978-3-032-17621-9_3

We can differentiate between objective and subjective switching costs. Objective switching costs are the observable costs of errors and increasing working time [1, 12, 29]. They are the observable results. If we observe them, everything that led to those results has already happened.

Therefore, the protagonists causing those switching costs might be the key to coming much earlier into the loop and understanding what causes those costs and how to prevent them. The resulting question is how far and precisely the protagonists experience possible reasons for switching costs, and how far something like consciousness exists for them or their underlying reasons. This we call subjective switching costs.

Suppose we can understand which reasons for switching costs exist in consciousness and which do not. In that case, we can develop more potent interventions, i. e., in terms of intentional forgetting, to prevent protagonists from causing and experiencing them and to help them adapt more successfully to a given change.

Even a single person can be understood as the last acting instance, contributing to an organizational outcome through their behavior. Switching costs can occur not only on an individual level. They can occur differently on the team and organizational levels. Therefore, team cognition and organizational cognition, understood as interacting processes between teams and at least across the whole organization, must be analyzed and differentiated.

The chapter reviews what we already know about switching costs. Then, the development of the first questionnaire is described, and the first results are reported. This is followed by an outline of measuring switching costs at the team and organizational levels. Finally, a critical appraisal is conducted, followed by practical implications and a conclusion.

2 What We Already Know

In history, psychological research has investigated switching costs as a matter of costs that occur due to task switches or situations, where familiar tasks conflict with unusual ones, such as the Stroop task [8, 9, 37]. In the typical Stroop task, participants must name the font color of a word that itself names a different color. The task involves suppressing the urge to read the word's meaning, which is what humans are used to, and instead naming the font color, which is less natural. Results are compared with trials with identical font color and word meaning, leading participants to show a non-conflicting outcome.

From this research, we understand that suppressing (intentionally forgetting) one highly trained behavior (reading) and executing a non-trained behavior (naming the font color) results in costs, which can be measured in terms of increasing errors and work time. However, participants reported that naming the font color and suppressing reading was more effortful than just reading or naming the font color in a consistent condition [29].

The consciousness people show in this traditional and not too complicated research about their cognitive effort gives hope that, when asked, they can name cognitive effort even in other changing situations where adaptive behavior is needed [30, 32]. The most important question is how detailed and differentiating those results might be and how specific they are in measurable terms in more complex work contexts.

Research on adaptation to changing rules shows that participants who suppress (intentionally forget) a prior behavior in favor of a new behavior pay attention to understand the situation and prepare their cognition for retrieving the new correct behavior. Besides suppression and activation, awareness, clarity, orientation, and preparation are also necessary factors [3, 30].

Finally, subjective switching costs result from an increasing number of cognitive operations necessary to adapt to a given change [2, 17, 28]. Here, intentional forgetting can reduce the cognitive workload, preventing information regarding the old from being processed [22].

3 Questionnaire Development

In the working context, switching costs result in cognitive effort in managing adaptation to a given change of actions within a work routine or process.

Not every action will be changed, mainly within a work routine change. Some actions remain unchanged, whereas others are changed. Knowing that clarity concerning the condition under which to fulfill a task or action is relevant to preparing to fulfill the action correctly, we have to generally differentiate between the subjective effort needed to perform actions according to their changing condition, changed vs. unchanged.

Under changing conditions, effort increases even to fulfill unchanged, used actions because there must be clarity that this upcoming action is unchanged. Also, uncertainty may conflict with this knowledge about the action change state because every action had to be executed unchanged before the change. Dealing with an unchanged action is not unique but risky because deciding to deal with an action in the used manner might be the wrong decision, overseeing the necessary adaptation whereas knowing that an action has to be executed in an adapted manner provides the safety that it is a changed action because one has the idea of the specific change.

When designing the Subjective Switching Cost Scale, we first differentiated between dealing with changed and unchanged actions and addressed dealing with unchanged actions as problematic as dealing with changed actions in the same manner.

Next, we structured the occurrence of switching costs by time and by execution phase. We started with the preparation phase before the execution, then continued focusing on the execution phase itself, and finally, the subjective experience and judgment of the results.

In the preparation phase, we differentiated between general feelings of confidence and orientation and specific feelings of clarity and effort according to the changing state of an upcoming action.

Addressing the execution phase, we asked for perceived subjective and personal feelings of ease, effort, and confidence.

According to the results, we asked to recall the subjective judgment on work time and correctness in executing an action. This should explicitly address the components of objective switching costs on a subjective level to draw a comparison. Specifically, regarding changing actions, we asked for possible conflicts requiring cognitive effort to differentiate between the old and used manner of executing a changed action and the new adapted way, respectively, the awareness of the effort needed to suppress the old manner in favor of the new one.

When constructing the items, we also paid attention to the fact that we did not always ask for either a negative or positive experience, but mixed reporting both. Therefore, we want to measure costs, so we gave the negative perspective the way, so the polarity of all positive formulated items must be changed for analysis.

Finally, we decided to provide a 5-point Likert scale for answering (ranging from 1 = not at all true, to 5 = exactly true).

4 Experimental Review of the Subjective Switching Cost Scale on an Individual Level

The questionnaire was validated as part of a larger experimental series involving multiple experiments and surveys designed to replicate the necessary change-setting practically. Participants underwent an actual change in a production routine and were subsequently asked to report their experienced subjective switching costs.

4.1 Experimental Setup

To adequately validate the developed questionnaire, it was implemented in an experiment in which participants first learned a production routine during an initial session. This routine was then modified as part of a predefined change in a second session, conducted either after seven or 21 days.

In groups of 2 or 3 workers, the participants learned how to produce knee joint prostheses. The product selection to be manufactured was primarily based on the requirement that the production process should be unfamiliar to all participants (this was also controlled for), while maintaining a certain level of complexity and requiring meticulous and precise work.

The research was conducted at the Research and Application Centre Industry 4.0 (RACI) at the University of Potsdam and simultaneously in a twin laboratory at

Table 1 Subjective Switching Costs Scale—items and subscales with abbreviations for the subscales in brackets

Subscale	No	Polarity	Item
1 Changed actions (CA)	1	–	I was able to perform changed actions fast
	2	+	I found it difficult to remember how to perform the changed action
	3	–	I was able to perform changed actions confident
	4	–	Before beginning to perform a changed action, I was confident how to perform the changed action
	5	+	I found it difficult to perform a changed action
	6	+	I was able to perform changed actions without errors
	7	–	I found it easy to perform a changed action
2 Unchanged actions (UCA)	8	–	I was able to perform unchanged actions fast
	9	+	Before beginning to perform an unchanged action, I first had to think about what to do
	10	+	I found it difficult to remember how to perform the unchanged action
	11	–	I was able to perform unchanged actions confident
	12	–	Before beginning to perform an unchanged action, I was confident how to perform the unchanged action
	13	+	I found it difficult to perform an unchanged action
	14	–	I was able to perform unchanged actions without errors
	15	–	I found it easy to perform an unchanged action
3 Effort recalling action's change state (ECS)	16	+	Before beginning to perform an unchanged action, I had to think about whether it was a changed or unchanged action
	17	+	Before beginning to perform a changed action, I had to think about whether it was a changed or unchanged action
4 Effort solving conflicts (ESC)	18	+	Before beginning to perform a changed action, I first had to think about what to do
	19	+	Before beginning to perform a changed action, the old action came to my mind
	20	+	While performing a changed action, the old action came to my mind
	21	+	I found it difficult not to perform the old action when performing a changed action
5 Clarity recalling action's change state (CCS)	22	–	Before beginning to perform an unchanged action, it was clear to me whether it would be an unchanged or a changed action
	23	–	Before beginning to perform a changed action, it was clear to me whether it would be an unchanged or changed action

Ruhr University Bochum. The settings on both locations were established as equal as possible.

As a laboratory-based special-purpose setting [23, 36], the Research and Application Centre Industry 4.0 provides a hybrid production simulation. Industrial hardware equipment (e.g., transportation systems, manufacturing robots, QR scanners) is enriched with software components. Participants can experience their interaction's physical, visual, and auditory effects. Using industrial components and footage from a real production environment makes it possible to re-enact modified real-world processes [5, 13, 26].

The data were collected along different experimental conditions, which do not conflict with the evaluation of the Subjective Switching Costs Scale (for further information about the experimental settings see [21–24, 33, 34]).

The study was approved by the Ethics Committee of the Faculty of Psychology at Ruhr University Bochum (Ethics vote No. 243). The participants were informed about the general setup of the study and assured that they could discontinue their participation at any time (informed consent). All participants were novices in learning the multi-actor routine and received 60 EUR for their participation. They were recruited through online postings and flyers handed out on campuses.

4.2 Experimental Design

During their first visit, the participants were welcomed and informed. After completing questionnaires, each participant was randomly assigned to a specific worker position. They received a booklet that introduced the required actions for their position. Each participant was allowed to produce three products without time pressure, followed by 40 min of production time. They were instructed to create prostheses as accurately and quickly as possible. After the production phase, the participants answered additional questionnaires.

During their second stay at the RACI, they felt expected to demonstrate what they had learned and produce as many products as possible using their trained routine. However, the welcome message revealed that the production process had changed due to a merger between a German and an American company. As a result, nearly half of the actions in the routine had been modified (e.g., measurements shifted from centimeters to inches and from degrees Celsius to degrees Fahrenheit). After completing a short questionnaire, the participants received another booklet tailored to their worker position, outlining the updated actions. After producing one workpiece using the booklet, the participants had 40 min to carry out the revised production as accurately and quickly as possible. This 40-minute production phase, conducted without the booklet, was the basis for collecting performance data. Finally, the participants completed additional questionnaires.

In the routines, each action was designed and described to align with the definition that actions are the most minor individually detectable units of behavior that can be sensibly delimited. According to [15], actions are units of activities characterized by specific objectives. An action is a temporally structured unit of activity and represents the smallest psychologically significant unit of deliberately controlled activities. The delimitation of actions occurs through conscious goals. A goal embodies the mental anticipation of an event (anticipation) connected to the intention of realization (intention). The design of the actions was specific at a micro level and comparable at a macro level [34].

The specific definition of each action on a micro level enabled a clear interpretation of its execution, which is a precondition for measuring behavior-based performance variables, especially objective switching costs [23, 33].

4.3 Experimental Variables

Subjective Switching Costs

As the central focus of this contribution, we collected the Subjective Switching Cost Scale in the version presented in Table 1 after the completion of the production phase in the second laboratory session.

Objective Switching Costs

We measured the objective switching costs in terms of work time and errors, as defined by [25, 29, 38], and [30].

After understanding how the modified production routine should be executed during the second stay at RACI 4.0, we examined data from the production process of the first workpiece produced without the learning material.

Working time

The working time was logged through entries in log files generated during interactions with the machines for each workpiece produced. Each individual's working time for the first workpiece produced without a booklet was measured in seconds.

Errors

Errors were assessed in a preliminary qualitative evaluation that compared the observed behavior with the defined ideal outlined in the booklet. We could distinguish the mistakes made based on the actions change condition.

We measured errors by assessing the relationship between incorrect actions performed in each category and all actions within that same category (Table 2).

Table 2 Change conditions of actions, error descriptions, and designations

Change condition of an action	Description of possible error	Error designation
Changed action to be performed in a modified manner	General wrong execution of a changed action but not in the way as in its original version	Unspecific incorrect execution of an action that needs to be modified
	Unchanged execution of a changed action to be executed in a modified manner	Maintained execution of an action that needs to be modified
Action that needs to be omitted	Continuation of the execution of an action that is no longer to be performed	Maintained execution of an action that needs to be omitted
Added action to be performed	The execution of the new action was not learned correctly	Incorrect execution of an added action
Action to be performed unchanged	General wrong execution of an unchanged action	Incorrect execution of an unchanged action

Further Person-Related Variables

Retentivity and self-efficacy were assessed as further person-related variables. Previous studies have demonstrated that retentivity is an important individual-related variable [14]. Self-efficacy is particularly relevant in this context due to the construct's proximity to potential subjective switching costs.

Retentivity

At the end of the second production phase, retentivity was measured using a subscale of the Wilde-Intelligenz-Test 2 (WIT 2) [19].

Self-efficacy

Self-efficacy was measured in general [35] at the very beginning of the experiment, as well as domain specific [20], at the end of the first production phase, and again after the change was announced at the beginning of the second production phase.

4.4 Sample

A total of 390 participants rated the subjective switching costs. Due to participant dropouts, performance data were available for analysis from only 201 participants (overall age $M = 23.78$, $SD = 4.42$; 90 females, age $M = 24.3$, $SD = 4.71$; 111 males, age $M = 23.35$, $SD = 4.15$).

4.5 Results

First, the scale itself was validated. Subsequently, the relationships between the entire scale and its identified subscales with actual working time and errors made were examined. Additionally, the correlations with the control variables retentivity and self-efficacy were analyzed.

The Subjective Switching Costs Scale

An exploratory factor analysis was conducted on the items with varimax rotation. The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis (KMO = 0.9, which is recommended as marvelous. [10, 18]). Five factors had eigenvalues over Kaiser's criterion of 1, which combined explained 54.9% of the variance, which is recommended as acceptable [11].

The five dimensions found in the exploratory factor analysis could be confirmed by a confirmatory factor analysis with sufficient results (corrected model: Scaled χ^2 (215) = 399.169, $p < 0.01$; Robust CFI = 0.940; Robust TLI = 0.930; Robust RMSEA = 0.054; SRMR = 0.065 [4]).

The overall Subjective Switching Costs Scale reached a Cronbach's Alpha of 0.89, which is recommended as good [31]. The following sections address the identified subscales individually. Therefore, the study was conducted at German universities. The items were developed in German and are only translated for publication (the data refer to the German version of the questionnaire in this publication).

Changed Actions

The Changed Actions subscale consists of seven items that focus on perceived ease, effort, and confidence in managing changed actions, as well as the subjective sense of correctness and working time. Cronbach's Alpha for this subscale reached 0.9, which is recommended as excellent [31], and it correlates with the overall scale with $r = 0.81$ ($p < 0.01$), which is recommended as large [7].

Unchanged Actions

The subscale Unchanged Actions consists of eight items that focus on perceived ease, effort, and confidence in managing unchanged actions, as well as the subjective sense of correctness and working time. Cronbach's Alpha for this subscale reached 0.88, which is recommended as good [31] and it correlates with the overall scale with $r = 0.75$ ($p < 0.01$), which is recommended as large [7].

Effort Recalling Actions Change State

The subscale Effort Recalling Actions Change State consists of two items addressing the cognitive effort required to determine, before initiating action, whether it involves a changed or unchanged action. This subscale's Cronbach's Alpha reached 0.71, which is recommended as acceptable [31], and it correlates with the overall scale with $r = 0.55$ ($p < 0.01$), which is recommended as large [7].

Effort Solving Conflicts

The subscale Effort Solving Conflicts consists of four items that measure the effort required to solve cognitive conflicts between the previous and the modified execution when actions have changed. This subscale's Cronbach's Alpha reached 0.76, which is recommended as acceptable [31], and it correlates with the overall scale with $r = 0.48$ ($p < 0.01$), which is recommended as medium [7].

Clarity Recalling Actions Change State

The subscale Clarity Recalling Actions Change State consists of two items addressing the clarity before initiating action, whether it involves a changed or unchanged action. This subscale's Cronbach's Alpha reached 0.66, which is recommended as acceptable for research in an early state [31], and it correlates with the overall scale with $r = 0.53$ ($p < 0.01$), which is recommended as large [7].

Objective and Subjective Switching Costs

The following section examines the relationship between objective and subjective switching costs to address how accurately switching costs can be measured subjectively and how precisely employees can assess them themselves (Table 3).

Working time

Overall, there is a positive correlation between the total Subjective Switching Costs Scale and working time ($r = 0.24$, $p < 0.01$), which is recommended as small [7]. The highest significant positive correlation exists between working time and the subscale Unchanged Actions ($r = 0.31$, $p < 0.01$), which is recommended as medium [7]. Lower correlations exist as well between the subscale Effort Recalling Action's Change State ($r = 0.15$, $p < 0.05$), which is recommended as small [7], and Clarity Recalling Action's Change State ($r = 0.15$, $p < 0.05$), which is recommended as small as well [7].

Among the two specific items that capture the subjective perception of working time for both changed (item 1) and unchanged actions (item 8), only the item addressing working time for unchanged actions shows a significant positive correlation with actual working time ($r = 0.24$, $p < 0.01$), which is recommended as small [7].

Table 3 Means, standard deviations, and correlations Subjective Switching Costs Scale, subscales, and performance measure ($n = 201$). Cronbach's Alpha of the Subjective Switching Costs Scale and its subscales is presented on the diagonal in brackets

Variable	M (*SD*)	1	2	3	4	5	6	7	8	9
1 CA	2.67 (0.83)	(0.9)								
2 UCA	2.14 (0.76)	0.42**	(0.88)							
3 ECS	2.66 (1.12)	0.21**	0.32**	(0.71)						
4 ESC	3.18 (0.89)	0.37**	−0.02	0.34**	(0.76)					
5 CCS	2.22 (0.97)	0.30**	0.43**	0.35**	−0.01	(0.66)				
6 SSCS overall	2.53 (0.53)	0.81**	0.75**	0.55**	0.48**	0.53**	(0.89)			
7 Retentivity	13.44 (3.39)	−0.16*	−0.23**	−0.06	0.04	−0.20**	−0.21**			
8 Self-efficacy	3.01 (0.34)	−0.22**	−0.02	0.07	−0.15*	0.00	−0.14	−0.16		
9 Error's	0.18 (0.11)	0.12	0.29**	0.11	−0.01	0.23**	0.24**	−0.35**	0.11	
10 Time	417.82 (92.19)	0.09	0.31**	0.15*	0.03	0.15*	0.24**	0.03	−0.04	0.01

Note M and SD are used to represent mean and standard deviation, respectively. CA represents the subscale Changed Action, UCA represents the subscale Unchanged Action, ECS represents the subscale Effort Recalling Action's Change State, ESC represents the subscale Effort Solving Conflicts, CCS represents the subscale Clarity Recalling Action's Change State, and SSCS overall represents the overall Subjective Switching Costs Scale. *Indicates $p < 0.05$. ** indicates $p < 0.01$

The results indicate that, according to working time, the items related to subjective switching costs, which refer to unchanged actions, can be answered more accurately and yield more reliable results.

Errors

In total, the overall Subjective Switching Costs Scale correlates with the overall relation of measured errors significantly ($r = 0.24$, $p < 0.01$), which is recommended as small [7]. In detail, only the subscales Unchanged Actions ($r = 0.29$, $p < 0.01$) and Clarity Recalling Action's Change State ($r = 0.23$, $p < 0.01$) show small [7] significant correlations.

According to the specific items addressing the subjective feelings of errors made, the item addressing changed actions (item 6) correlates with $r = 0.14$ ($p < 0.05$). The item addressing unchanged actions (item 14) correlates with $r = 0.21$ ($p < 0.01$). Both correlations are recommended as small [7].

These results indicate that employees generally seem aware of their errors, suggesting that this awareness can be measured. However, the results appear more robust concerning unchanged actions than unfamiliar, modified actions, as observed already with the working time. They also emphasize the importance of clarity about the change condition before initiating an action.

Unspecific incorrect execution of an action that needs to be modified

In total, we find the smallest significant correlation with the overall Subjective Switching Costs Scale ($r = 0.17$, $p < 0.05$), which is recommended as small [7]. Further significant correlations with subscales exist with the subscale Clarity Recalling Action's Change State ($r = 0.20$, $p < 0.01$) and Unchanged Actions ($r = 0.16$, $p < 0.05$). Both correlations are recommended as small [7]. The specific items addressing the subjective feelings of errors made do not correlate significantly. But the specific items addressing the subjective feelings of work time do (according to changed actions (item 1) $r = 0.15$, $p < 0.05$, according to unchanged actions (item 8) $r = 0.17$, $p < 0.05$). Both correlations are recommended as small [7].

Maintained execution of an action that needs to be modified

Here, we find no significant correlation with the overall scale ($r = -0.07$, $p > 0.05$) and only a negative significant correlation with the subscale Changed Actions ($r = -0.16$, $p < 0.05$), which is recommended as small [7]. Also, the specific items do not have any significant correlation.

The negative correlation is striking. Individuals who report higher subjective switching costs according to changed actions tend to make fewer errors while maintaining the execution of an action that requires modification. This further suggests the significance of awareness of the modification and its clarity. The awareness of those switching costs might be decisive.

Maintained execution of an action that needs to be omitted

With this kind of error, we cannot find any significant correlation. The correlations we find are close to zero. This may suggest that there is almost no awareness of this type of error when it occurs.

Incorrect execution of an added action

Here, we identify only one significant correlation with the Unchanged Action subscale ($r = 0.15$, $p > 0.05$), which is recommended as small [7], and no significant correlation with the overall scale or the specific items that address the subjective feelings regarding errors made.

Incorrect execution of an unchanged action

As expected, we find that this error has one of the highest correlations with the Unchanged Actions subscale ($r = 0.33$, $p < 0.01$), which is recommended as medium [7]. Additionally, there are correlations with the Changed Actions subscale ($r = 0.21$, $p < 0.01$), the Clarity Recalling Action's Change State subscale ($r = 0.21$, $p < 0.01$), and the overall Subjective Switching Costs Scale ($r = 0.28$, $p < 0.01$). All three correlations are recommended as small [7]. The specific items that address the subjective feelings of errors made also correlate significantly (the item addressing Changed Actions (item 6) $r = 0.16$, $p < 0.05$, the item addressing Unchanged Actions (item 14) $r = 0.29$, $p < 0.01$). Both correlations are recommended as small [7].

Further Person-Related Variables

Since measuring subjective switching costs relies on assessing individual selfperception and evaluation, the relationship with already well-established person-related variables—retentivity and self-efficacy (both general and specific)—which have been shown to correlate with objective measures such as error rate and working time [14]—can serve as an additional validity criterion for the scale.

Retentivity

In the current sample, retentivity correlates significantly negatively with the general error rate ($r = -0.35$, $p < 0.01$), as expected, which is recommended as medium [7].

The overall Subjective Switching Costs Scale goes along with this finding with a correlation $r = -0.21$ ($p < 0.01$). The subscales, Changed Actions ($r = -0.16$, $p < 0.05$), Unchanged Actions ($r = -0.23$, $p < 0.01$), and Clarity Recalling Action's Change State ($r = -0.20$, $p < 0.01$) contribute to the confirmation of fitting relationships. Those four correlations are recommended as small [7].

Self-Efficacy

In this sample, the participants' overall self-efficacy does not show any significant relationship with the error rate or working time; however, production-specific self-efficacy, particularly measured immediately after the announcement of the change, does demonstrate such a relationship (error rate $r = -0.23$, $p < 0.01$, working time $r = -0.20$, $p < 0.01$). Both correlations are recommended as small [7]. This means that as a person's self-efficacy score increases, the number of errors made decreases, and less working time is required.

For general self-efficacy, we do not observe significant correlations with the overall Subjective Switching Costs Scale, but we do find significant negative correlations for the subscales Changed Action ($r = -0.22$, $p < 0.01$) and Effort Solving Conflicts ($r = -0.15$, $p < 0.05$). Both correlations are recommended as small [7].

For production-specific self-efficacy, we find negative correlations with the overall scale ($r = -0.35$, $p < 0.01$), which is recommended as medium [7], and with the subscales Changed Actions ($r = -0.21$, $p < 0.01$), which is recommended as small [7], Unchanged Actions ($r = -0.46$, $p < 0.01$), and Clarity Recalling Action's Change State ($r = -0.32$, $p < 0.01$). Both correlations are recommended as medium [7].

Those findings fit in with the overall context of the results and support the validity of the Subjective Switching Costs Scale, which measures subjective switching costs at the individual level.

In addition to the individual level, subjective switching costs can also arise at the team and organizational levels.

5 Perspectives on Team and Organizational Levels

Switching costs involve errors and working time spent at both the team and organizational levels, too. The smallest unit where switching costs can arise is the individual action of a single person. However, change-induced switching costs are not fully overcome in a team setting until the last team member has completed the necessary adaptation. Therefore, adaptation during a change process is a conjunctive task that all members must effectively manage [6].

Unlike individuals acting independently, switching costs at the team level can also emerge through parallel or sequential collaboration.

When considering the measurability of subjective switching costs at the team level, which extends beyond an individual's perception of their performance, the question arises as to whether team members develop a collective awareness of switching costs—one that may occur independently of their own contributions and outside their immediate perception.

Switching costs at the team level can easily go unnoticed because they do not necessarily need to be perceived equally by all team members. Therefore, switching costs are also a matter of the team's shared mental model, meaning that exchange and reflection are necessary to establish an everyday awareness of their existence within the team [27] provide an initial idea.

At the most fundamental level, this exchange can occur through mutual observation among team members. Still, it can only be reliably ensured if integrated into daily routines as a formal debriefing and reflection mechanism.

Well-established, high-performance teams are usually capable of managing this process autonomously. They are highly mutually performance sensitive and have established mechanisms for performance-related exchanges.

For less experienced teams, ensuring the necessary exchange becomes more of a leadership responsibility. Initiating discussions about prevalent switching costs during adaptation to change can be a role-specific task within the team.

In this context, teams do not require every member to be initially aware of switching costs. Instead, at the beginning of the exchange, it may be sufficient for only a few team members to identify and highlight existing switching costs. However, switching costs can also escalate within teams, particularly when no team member acknowledges them. This allows them to be unknowingly reinforced among several members or even the entire team.

Beyond team maturity and the direct influence of leadership, the prevailing organizational climate is another crucial factor. How an organization deals with mistakes shapes how individuals handle their own errors, how teams collectively address errors, and how openly and learning-oriented improvement potentials are discussed.

The organizational climate surrounding error management influences the development of a shared, subjective awareness of switching costs.

At the organizational level, the likelihood of switching costs going unnoticed increases further, as direct observation and interaction among all participants in a cross-team process are often impractical. Additionally, the direct influence of leadership diminishes because individual leadership spans no longer directly encompass all process participants.

To answer the question of what basis remains for a subjective awareness of existing switching costs, an analysis of individual subjective switching costs provides insight into aspects of uncertainty and situational clarity.

This analysis reveals that, alongside immediate awareness of errors and working time, perceptions of uncertainty and ambiguity are closely linked to objectively existing switching costs.

These change-related characteristics of subjective switching costs can also be measured across different departments and at the organizational level, providing at least an approximate assessment of the potential existence of switching costs.

This potential is further supported by the strong relationship between subjective and objective costs and their connection to particular self-efficacy expectations.

Therefore, focusing on perceived uncertainty during adaptation to change can serve as an early and subjectively measurable indicator of potentially hidden switching costs at the individual, team, and organizational levels.

Research into subjective and objective switching costs, especially at the team and organizational level, is still in its beginning.

6 Practical Implications and Possible Interventions

Results on an individual level show that subjective switching costs are in line with objective ones. Even so, there might still be some potential for improvement, especially in precision. This means that workers have a sense of their change-related effort, uncertainties, errors, and work time needed.

Reported subjective switching costs associated with unchanged actions correlate better with errors and work time than those tied to changed actions. This could be because workers find it easier to compare their experiences executing the same action before the change instead of during the change, which may give them a sense of security in evaluating their cognitive effort.

This may lead to two conclusions. First, even unchanged actions, which one might assume are unaffected during a change, incur switching costs. Second, workers who report higher rates of switching costs also tend to incur higher objective switching costs.

While we may not be as precise in measuring subjective switching costs at the level of specific actions, the general findings suggest that inquiring about subjective switching costs could lead to personal identification and help identify workers who require additional support and experience greater uncertainty.

The results indicate that felt uncertainty may be understood as the underlying factor influencing the entire scale and becomes evident in the subscale Clarity in Recalling Action's Change State. This subscale also demonstrates the importance of orientation and preparation before executing an action.

The cognitive preparation phase for executing an action [16] is an opportunity to introduce cues that alter the condition of an action and support adaptive behavior. This idea aligns with the principles of many management initiatives aimed at drawing attention to nearly automated behaviors performed with low awareness. Concepts known from workplace safety (e.g., stop before start) could also be beneficial in situations where it is not the surroundings but the actions that change.

The lack of correlation between subjective switching costs and errors related to maintaining actions that should be omitted also indicates a lack of awareness of certain changes. The first productive step must be identifying these actions in a specific change and increasing awareness of the errors made. Otherwise, altering a behavior whose misleading nature goes unnoticed may be nearly impossible.

7 Conclusion

Switching costs are an inherent challenge in organizational change, impacting individuals, teams, and entire organizations. They are a measure of success in forgetting the old and applying the new. This chapter highlights the distinction between objective switching costs, such as errors and increased working time, and subjective switching costs, encompassing cognitive effort and uncertainty. Validation of the Subjective Switching Cost Scale demonstrates that employees can perceive and report individual switching costs, particularly regarding unchanged actions. At the team and organizational levels, shared awareness and structured reflection processes are crucial for minimizing these costs. Practical interventions, such as leadership support, adaptive training, error management strategies, and applying methods of intentional forgetting, can facilitate smoother transitions, enhance change readiness, and success. Recognizing and addressing both objective and subjective switching costs is essential for improving efficiency, reducing frustration, and fostering long-term organizational adaptability. Measuring subjective and objective switching costs provides information about the success of the change, how methods of intentional forgetting were successfully applied, and the extent to which beneficial forgetting processes were successful on individual, team, and organizational levels.

Acknowledgements Special thanks go to our cooperation partners for their support in conducting the laboratory survey: Norbert Gronau, Christof Thim, and Jennifer Haase, as well as the numerous student assistants who supported the data collection and analysis to the best of their abilities.

Funding The project was funded from 2016 to 2019 by the Deutsche Forschungsgemeinschaft (DFG), project number KL2207/6-1/6-2, Project number 317987159.

Competing Interests No potential conflict of interest was reported by the authors.

Ethics Approval The study was approved by the local ethics committee of the faculty of psychology (No. 243, from Nov 20th 2015), when the funding proposal was submitted.

References

1. Adner, R. (2006). Match your innovation strategy to your innovation ecosystem. *Harvard Business Review, 84*(4), 98.
2. Allport, D. A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In Umiltà, C., & Moscovitch, M. (Eds.), *Attention and Performance XV: Conscious and Nonconscious Information Processing* (Vol. 15, pp. 421–452). Erlbaum.
3. Aufschnaiter, S., Kiesel, A., & Thomaschke, R. (2021). Time-based transition expectancy in task switching: Do we need to know the task to switch to? *Journal of Cognition, 4*(1), 19, 1–14. https://doi.org/10.5334/joc.145
4. Beauducel, A., & Wittmann, W. W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. *Structural Equation Modeling, 12*(1), 41–75.
5. Bender, B., Grum, M., Gronau, N., Alfa, A., & Maharaj, B. T. (2019, June). Design of a worldwide simulation system for distributed cyber-physical production networks. In *2019 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)* (pp. 1–7). IEEE.

6. Burke, C. S., Stagl, K. C., Salas, E., Pierce, L., & Kendall, D. (2006). Understanding team adaptation: A conceptual analysis and model. *Journal of Applied Psychology, 91*(6), 1189–1207.
7. Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.
8. Cohen, J. D., & Huston, T. A. (1994). *Progress in the use of interactive models for understanding attention and performance*. The MIT Press.
9. Chamberland, C., & Tremblay, S. (2011). Task switching and serial memory: Looking into the nature of switches and tasks. *Acta Psychologica, 136*(1), 137–147.
10. Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? *Some decision rules. Psychological Bulletin, 81*(6), 358–361.
11. Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis*. Oxford University Press.
12. Feldman, M. S., Pentland, B. T., D'Adderio, L., & Lazaric, N. (2016). Beyond routines as things: Introduction to the special issue on routine dynamics. *Organization Science, 27*(3), 505–513.
13. Gronau, N., Kluge, A., Haase, J., & Thim, C. (2023). Experiential learning factories: Bridging the gap between lab and field experiments. In *Proceedings of CLF 2023 13th Conference on Learning Factories, 9-11 May 2023, Reutlingen University, Germany*. https://doi.org/10.2139/ssrn.4469819
14. Haase, J., Matthiessen, J., Schüffler, A. S., & Kluge, A. (2020). Retentivity beats prior knowledge as predictor for the acquisition and adaptation of new production processes. In *HICSS 2020* https://doi.org/10.24251/HICSS.2020.589
15. Hacker, W. (2010). *Psychische Regulation von Arbeitstätigkeiten* (Vol. 1, pp. 3–38). Hogrefe.
16. Heckhausen, J., & Heckhausen, H. (Eds.). (2018). *Motivation und Handeln*. Springer-Verlag.
17. Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.
18. Kaiser, H. F. (1974). An index of factor simplicity. *Psychometrika, 39*(1), 31–36.
19. Kersting, M., Althoff, K., & Jäger, A. O. (2008). *Wilde-Intelligenz-Test 2: WIT-2*. Hogrefe.
20. Kluge, A. (2004). *Wissenserwerb für das Steuern komplexer Systeme*. Pabst.
21. Kluge, A., & Gronau, N. (2018). Intentional forgetting in organizations: The importance of eliminating retrieval cues for implementing new routines. *Frontiers in Psychology, 9*, 51.
22. Kluge, A., Schueffler, A. S., Thim, C., & Gronau, N. (2024). Facilitating and hindering factors for routine adaptations in manufacturing and effects on human performance - Unexpected insights from three experimental studies in a special purpose setting. *Ergonomics, 67*(12), 2063–2081.
23. Kluge, A., Schueffler, A. S., Thim, C., Haase, J., & Gronau, N. (2019). Investigating unlearning and forgetting in organizations: Research methods, designs and implications. *The Learning Organization, 26*(5), 518–533.
24. Kluge, A., Schueffler, A. S., Thim, C., Vladova, G., & Gronau, N. (2018). Putting intentional organisational forgetting to an empirical test: Using experimental designs to measure forgetting of organisational routines. In *Proceedings of the IFKAD Conference* (pp. 245–267).
25. Koch, I., Poljac, E., Müller, H., & Kiesel, A. (2018). Cognitive structure, flexibility, and plasticity in human multitasking - An integrative review of dual-task and task-switching research. *Psychological Bulletin, 144*(6), 557.
26. Lass, S., & Gronau, N. (2020). A factory operating system for extending existing factories to Industry 4.0. *Computers in Industry, 115*, 103128.
27. Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *The Academy of Management Review, 26*(3), 356–376.
28. Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobeâŁž tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100.
29. Monno, I., Spitzer, M., Miller, J., Dignath, D., & Kiesel, A. (2021). Scaling of the parameters for cost balancing in self-organized task switching. *Journal of Cognition, 4* (1), 8, 1–16. https://doi.org/10.5334/joc.137
30. Monsell, S. (2003). Task switching. *Trends in Cognitive Science, 7*, 134–140.
31. Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

32. Rubinstein, J. S., Meyer, D. E., & Evans, J. E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 763–797.
33. Schueffler, A., Thim, C., Haase, J., Gronau, N., & Kluge, A. (2019). Willentliches Vergessen – Voraussetzung für Flexibilität und Veränderungsfähigkeit in einer sich permanent verändernden Welt. *Gruppe. Interaktion. Organisation. Zeitschrift für Angewandte Organisationspsychologie (GIO), 50* (2), 197–209.
34. Schueffler, A., Thim, C., Haase, J., Gronau, N., & Kluge, A. (2020). Information processing in work environment 4.0 and the beneficial impact of intentional forgetting on change-management. *Zeitschrift für Arbeits- und Organisationspsychologie (ZAO), 64* (1), 17–29.
35. Schwarzer, R., & Jerusalem, M. (1999). Skalen zur Erfassung von Lehrer- und Schülermerkmalen. *Dokumentation der psychometrischen Verfahren im Rahmen der wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*. Freie Universität Berlin.
36. Stone-Romero, E. F. (2011). Research strategies in industrial and organizational psychology. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology* (Vol. 1, pp. 37–72). American Psychological Association.
37. Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.
38. Wylie, G., & Allport, A. (2000). Task switching and the measurement of "switch costs". *Psychological Research,63*(3), 212–233.

Trust Matters: The Role of Trust and Intentional Forgetting for Information Systems in Organizations

Sebastian Reiners, Lea Sophie Müller, Jörg Becker, and Guido Hertel

1 Introduction

In organizational settings, Information Systems (ISs) typically process central information for the execution of user tasks. Hence, ISs are seen as triggers of intentional forgetting in organizations. They provide users with relevant material so that less relevant information can fade into the background. However, engaging with ISs introduces a certain degree of vulnerability [18]. Users relinquish control and open themselves up to unexpected (consequences) sourced from (semi-) automated decision-making performed by these systems [40, 53]. To tolerate these circumstances and to forget non-relevant information safely, users need to trust technology. For ISs, trust describes a user's willingness to depend on and be vulnerable toward the system in uncertain and risky situations [18]. Users need to believe 'a specific technology has the attributes necessary to perform as expected in a given situation in which negative consequences are possible' [41].

Recognizing the importance of trust in this context, recent years [24, 47, 63] have seen a culmination of research building on the extensive history [12, 20, 37] of interpersonal and organizational trust studies. Notably, findings shed light on the intricate trust between individuals and IS within organizational settings [44]. However, the

S. Reiners (✉) · J. Becker
European Research Center for Information Systems, University of Münster, Leonardo-Campus 3, Münster, Germany
e-mail: sebastian.reiners@ercis.uni-muenster.de

J. Becker
e-mail: joerg.becker@ercis.uni-muenster.de

L. S. Müller · G. Hertel
Organizational and Business Psychology, University of Münster, Münster, Germany
e-mail: lea.mueller@uni-muenster.de

G. Hertel
e-mail: ghertel@uni-muenster.de

J. O. Berndt et al. (eds.), *Intentional Forgetting with Intelligent Systems*, Progress in IS,
https://doi.org/10.1007/978-3-032-17621-9_4

character of trust poses a big challenge for organizations trying to govern user trust in ISs. One must consider the factors that influence the development of trust in ISs (antecedents), effects that result from trust in technology (outcomes), and aspects that alter the strength of antecedents/outcomes on/of trust (moderators) simultaneously. At the same time, trust dynamics between individuals and ISs within organizational settings lack sufficient empirical validation for theoretical assumptions [33, 41, 44]. There is an opportunity to enhance our understanding by investigating the temporal dynamics of trust from IS adoption to routine operation [59]. This can help in making informed predictions and decisions regarding trust-building processes as well as governing the ISs adoption process [41, 44].

In this book chapter, we contrast two research papers [48, 49] featuring longitudinal case studies in five distinct organizations [68]. Doing so, we unveil the positive impacts of trust in IS relationships, offering valuable insights for trust in organizational environments. We pay special attention to the longitudinal perspective on the antecedents, outcomes, and moderators of trustful IS usage in organizations. Sourcing from the comparison, we discuss the idea of managing of trust in employee-technology relationships from an organizational perspective [8, 13, 55]. A contribution in this regard supplements the understanding of trust as a dynamic construct within the life cycle of an IS, further aggregating the knowledge of antecedents and outcomes of trustful IS usage.

2 Background

2.1 Trust

Trust touches on seemingly all aspects of life. It is the foundation upon which societal structures are built—including the physical and digital worlds [12]. Numerous studies and in-depth discussions have contributed to robust concepts of trust in specific domains [18, 42, 65]. Being 'somewhat mystical and intangible' [20], it brings many ambiguities and difficulties when describing its conception. Fellow researchers have illustrated this maze by describing it as 'a bewildering array of meanings' [61] or even 'a conceptual morass' [7].

In its essence, trust encompasses a specific mental state experienced by the trustor (the actor who trusts) concerning the trustee (the actor who is trusted) [56]. Historically, trust was primarily investigated between individuals. This type of trust is referred to as interpersonal trust. Key aspects among leading definitions of interpersonal trust [18, 40, 57] highlight pivotal characteristics that contribute to a general definition of trust: It is a psychological state of willingness to be vulnerable and positive expectations about another party's intentions or behavior.

One can identify three distinct actors that assume trustor and trustee roles in trust relationships: individuals, organizations, and technology. In this particular work, we focus on 'Trust of Individuals and Organizations in Technology'. Numerous studies have explored the development of trust between users (as trustors) and technological artifacts (as trustees), such as e-commerce systems [17] or websites [42]. As a result, trust has been established as the object of dependence in ISs research and psychology, although as a rather abstract construct. A special focus is given to the relationship between employees and ISs as a manifestation of an employed technology within an organization. In organizational contexts, the handling of technology by individuals is not always of a voluntary nature [63]. In such environments, ISs are frequently associated with diverse assignments and characterized by high complexity, uncertainty, and personal accountability [26, 63].

For the subsequent parts of this work, a technology is considered a trustee, and users, i.e., individuals using the technology, are considered trustors. Furthermore, a particular technology upon which trust can be developed is defined as an IS. Therefore, an individual's general propensity toward technology is distinct to a specific IS upon which trust is developed [41].

2.2 Trust Dynamics in Information System Adoption

The relationship of trust in technology is defined and conceptualized in several theories, each focusing on specific particularities and settings [33, 41, 44]. A widespread perspective identifies trust beliefs as specifically associated with a particular technology [41]. Furthermore, it is highlighted that trust in technology cannot be examined in isolation. Figure 1 illustrates the antecedents, outcomes, and moderators of trust in ISs as it is used in this work. Initially, the model was proposed to conceptualize trust in organizations [40] and later adapted to depict trust in management ISs [44].

We define trust in an IS as an employee's willingness to rely on and be vulnerable to an IS used to fulfill occupational tasks according to organizational demands [18, 40, 44]. Extending this definition with time-related dynamics, trust is the experienced state of an individual user shaped by that individual's disposition to technology, contextual factors, and the perceived trustworthiness of the IS [44]. As a result, the experienced state of trust encompasses both affective and cognitive dimensions [44, 63].

Perceived trustworthiness refers to an individual's subjective appraisal of an IS's characteristics (e.g., credibility) [44, 63]. It is distinct from trust, with the latter delineated as an emergent state in which an individual is willing to become vulnerable to an IS by depending on it [40].

The concept of trust disposition, sometimes propensity [41], originates from interpersonal trust in which it represents an individual's willingness to trust others in general [40]. For trust in technology, trust disposition refers to an individual's tendency to rely on technology in various situations and with different technologies [41]. The concept of intentional forgetting in organizations refers to the deliberate omission or suppression of information that is no longer relevant to a specific task, goal, or context [30]. This strategy is increasingly recognized as essential for maintaining organizational agility and individual cognitive efficiency in complex

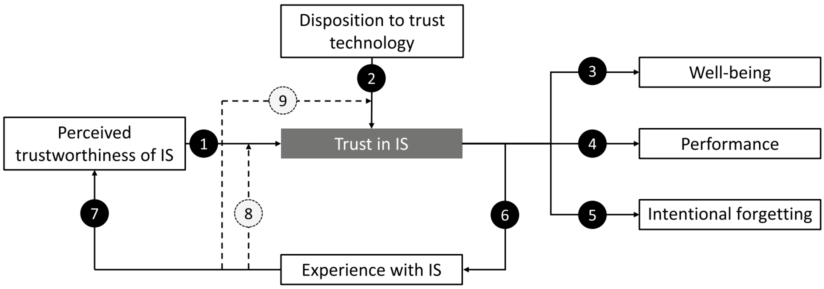


Fig. 1 Trust model of antecedents, moderators, and outcomes

and dynamic work environments. Individuals can engage in intentional forgetting when explicitly instructed or prompted by contextual cues, which enables them to discard obsolete knowledge and redirect their cognitive resources toward more pertinent and time-sensitive tasks [3, 24]. This selective unlearning supports adaptability, fosters innovation, and helps prevent cognitive overload. In contrast, the accumulation and retention of outdated or irrelevant information can contribute to mental clutter, reduced decision-making quality, and decreased task performance. Here, the concept of strain becomes relevant, as it describes the negative psychological and emotional consequences of prolonged exposure to stressors such as excessive workload, time pressure, and high accountability demands [46]. Strain can manifest in various forms, including fatigue, decreased motivation, and burnout, ultimately impairing employee well-being and organizational productivity. ISs when thoughtfully designed and implemented offer significant potential to reduce such work-related strain. By structuring, filtering, and prioritizing information, ISs can help employees focus on what is most relevant while minimizing distractions caused by outdated or excessive data [23, 64]. Moreover, ISs can support intentional forgetting through automated archiving, context-aware notifications, and adaptive workflows that align information presentation with the evolving needs of users. In this way, digital systems not only enhance task efficiency but also contribute to healthier, more sustainable work environments.

Within an organizational context, trustful ISs usage has been recognized for its potential to improve employee performance in recent empirical studies [63]. This is done by simplifying tasks and expediting task completion [5, 50]. However, higher levels of task performance are contingent upon the trust users place in the system. When employees lack trust in an IS, they may engage in inefficient behaviors such as double-checking information.

Considering trust as an experienced state of an individual user entails that early trust in an IS is influenced by factors other than information about specific system characteristics. Hence, the concerns and anxieties of users are more prevalent in the beginning stages of IS usage [21]. Limited knowledge can make trust disposition a crucial determinant of trust in the early stages [47, 57]. The user's experience utilizing an IS can be considered a moderator for the factors contributing to trust in

ISs [45]. Evaluating factors such as reliability will become more refined as users gain specific experience with a particular IS. If a system consistently exhibits unreliability, such as displaying incorrect calculations, the system's trustworthiness will likely be evaluated more negatively. As users gain more experience over time, they acquire the necessary information to make informed trust decisions based on their first-hand encounters [44, 47]. This makes for a case in which the strength of the relationship between the trustworthiness of an IS and user trust is also moderated by experience.

2.3 Intentional Forgetting

The concept of intentional forgetting is particularly evident in the role of ISs as external memory, providing relevant information for users' tasks that employees theoretically do not need to retain in their own memories. Hertel et al. (2019) [24] highlight the potential of IS to serve as cognitive aids, relieving employees of cognitive load and thereby releasing capacities for other tasks. Drawing on early research on individual-level memory [3], recent research suggests that trust in the IS is a prerequisite for experiencing the beneficial effects of intentional forgetting [24, 47]. In an information-rich environment, forgetting can be advantageous for organizations and employees facing increased volumes of data [11, 15]. ISs that process and analyze data, e.g., to support decision-makers, can be seen as structural triggers for intentional forgetting [24].

The significance of trust in this context is underscored by the nature of ISs processing central information for users' task execution [54]. To fully benefit from the cognitive offloading potential of information systems, employees must trust the system enough to entirely rely on it. Conversely, in the absence of trust, employees may be less inclined to forget information processed by the system, consequently releasing fewer cognitive capacities. Trust, therefore, emerges as a critical factor not only in optimizing task execution but also in unlocking the cognitive benefits associated with ISs reliance.

3 Method

Leveraging the strengths of diversity in IS research methods, we apply a concurrent mixed methods approach [34, 58, 68]. Mixed methods research involves utilizing the duality of qualitative insights with quantitative data analysis to simultaneously address exploratory and confirmatory research inquiries [60, 68]. In this work, we combined survey data and interview data collected in case studies with respect to the research objective of this work. Case studies are particularly valuable for understanding how ISs interact with organizational structures, processes, and people. They pose an excellent opportunity for studying socio-technical effects in a real-world context [27]. In this work, two case studies are presented. In case study #1 [48], we examined antecedents, outcomes, and moderators of trust in a newly introduced IS. Following a longitudinal design, data collection was done over four time points to compare short-, mid-, and long-term attitudes toward the IS. The examination included an individual's disposition to trust, the perceived trustworthiness of the IS, and the context (e.g., support structures). The survey questions were designed in close collaboration with the project team responsible for introducing the IS. Exemplary questions are presented in Table 1. Participants rated all items on a 7-point Likert scale, ranging from 1 (fully disagree) to 7 (fully agree), unless otherwise indicated. All items were presented in the German language. For the reliability of the scales, we calculated Cronbach's alpha for scales with three or more items and Spearman-Brown correlations for two-item scales.

A multiple case study approach was carried out in case study #2 [49]. Four cases were investigated, featuring four distinct organizations and three different ISs. Notably, two organizations introduced the same ISs. In all organizations, the surveys were done over four time points. The initial survey (T0) was issued shortly before the system's launch. Further data were collected approximately one month after the launch (T1), about four months after launch (T2), and about eight months after launch (T3) following the measurement schedule of case study #1. A flowchart of the schedule is presented in Fig. 2. Qualitative interviews with project managers from each organization delved into the organizational strategies behind the IS introductions after T3 to complete the data collection. Following the naming in the publication, we refer to the investigated ISs using 'system 1', 'system 2', and 'system 3'. Furthermore, to be consistent in this work, we continue the enumeration for the IS investigated in case study #1. Hence, the system is referred to as 'system 4' in the following. An overview on the case studies is given in Table 1.

4 Antecedents, Outcomes, and Moderators of Trust

4.1 Case Study Results

Trust relations presented in the theoretical model were subsequently investigated. In the following, the results are presented while paying special attention to the dynamics of trust. Building on the IS success model [9] and later research [63], the investigation of case study results revealed that different system properties are indeed (positively) related to trust. We found that the system's reliability significantly predicted trust at T1 and T2, credibility at T2 and T3, and usability at T1 and T3 [46]. Hence, we found a system's reliability, credibility, and usability to predict trust significantly [46, 49]. However, although all named properties significantly predicted trust, at least in one of our measurements, there are differences. That is, their influence was not enduring but shifted between time points (for time-related specifics, see relations (6)–(9)). Table 2 presents the results from different measurement points.

Table 1 Exemplary measures of trust, its antecedents, and outcomes

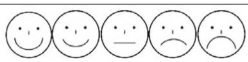
Target	Description / Item	Scale / Measure	Source (if applicable)
Trust	(1) I completely trust the IS (2) I heavily rely on the IS (3) I feel comfortable relying on the IS	5-/7-point Likert scale Calculate mean score	[49] (Thielsch et al., 2018)
Trust	How do you rate your current trust in the IS?	1 - 7 / 1 - 10 / 1 - 100	/
Trust	What are the reasons for your current level of trust in the IS?	open text	[49]
Trust disposition	(1) I usually trust computer technology until it gives me break a reason not to trust it. (2) I generally give computer technology the benefit of the doubt when I first use it. (3) My typical approach is to trust new computer technologies until they prove to me that I shouldn't trust them.	5-/7-point Likert scale Calculate mean score	[33] (Lankton et al., 2015)
Trustworthiness (reliability)	(1) The IS is a very reliable piece of software (2) The IS does not fail me (3) The IS does not malfunction for me	5-/7-point Likert scale Calculate mean score	[41] (Mcknight et al., 2011)
Trustworthiness (functionality)	(1) The IS has the functionality I need (2) The IS has the features required for my tasks (3) The IS has the ability to do what I want to do	5-/7-point Likert scale Calculate mean score	[41] (Mcknight et al., 2011)
Trustworthiness (helpfulness)	(1) I found the IS unnecessarily complex (2) I thought the IS was easy to use (3) I found the various functions in the IS were well integrated	5-/7-point Likert scale Calculate mean score	(J. R. Lewis et al., 2013)
Trustworthiness (credibility)	(1) I find the information provided on the IS to be authentic (2) The information provided on the IS is reliable (3) I can trust the information on the IS	5-/7-point Likert scale Calculate mean score	[62] (Thielsch & Hirschfeld, 2019)
Trustworthiness (support structures)	(1) If problems with the system occur, support is available. (2) The support is helpful.	5-/7-point Likert scale	[63] (Thielsch et al., 2018)
Trustworthiness (abilities)	(1) The qualification of persons involved is sufficient. (2) I trust in the professional competence of the people responsible for the system. (3) I think I can rely on the skills of persons involved.	yes/no	(Hertel et al., 2004)
Well-being	 How do you feel at this moment?	5-point smiley scale	(Jäger, 2004)
Well-being	(1) I find working with the IS demanding (2) I find working with the IS stressful (3) I find working with the IS onerous	5-/7-point Likert scale Calculate mean score	(Stanton et al., 2001)
Performance	(1) Using the IS increases the quality of my work (2) Using the system makes my job easier (3) Using the IS saves me time (4) Using the IS helps me fulfill the needs and requirements of my job	5-/7-point Likert scale Calculate mean score	(Etezadi-Amoli Farhoomand, 1996)

Table 2 Case study information

Research method	Case study #1	Case study #2
Data source	Survey	Survey / Interview
Sample	313	157 / 4
Measurement points	0: 07/2020; 1: 08/2020; 2: 01/2021; 3: 06/2021	0: before launch; 1: one month; 2: few months; 3: eight months Interview after '3'
Organization	University	Four private companies

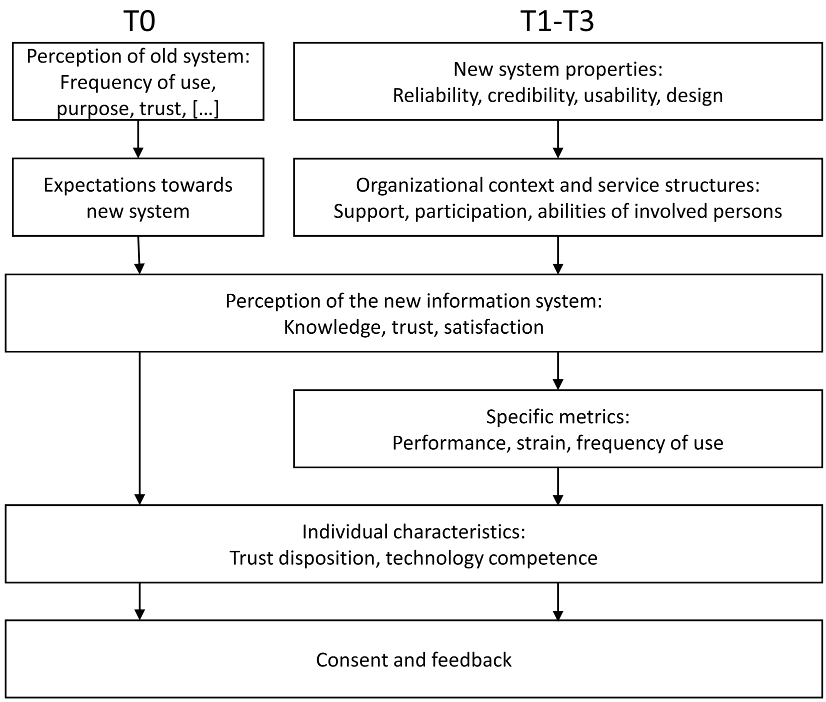


Fig. 2 Flowchart of surveys for longitudinal case studies

We measured the significance of various trust factors at different stages of system adoption as perceived by individuals. Given that our initial measurements took place prior to the system's launch (T0), we found that participants already had a certain level of trust in the system, even without any usage experiences. This aligns with earlier suggestions that a phase of initial trust building exists before the first interaction with an IS [59]. In case study #2, we observed similar results across three other systems. We found context structures to be evaluated positively if the adoption of the IS was perceived as successful by the employees [49]. Nonetheless, we never found

Table 3 Exemplary results for system 4 showing relationships of trust in information systems with antecedents and outcomes at T0-T3

	T0		T1		T2		T3	
	β	p	β	p	β	p	β	p
Trust disposition	0.216***	0.007	0.061	0.208	0.064	0.164	0.023	0.601
System properties								
Reliability			0.216*	0.015	0.304***	0.000	0.163	0.118
Credibility			0.084	0.318	0.184*	0.012	0.253**	0.008
Usability			0.640***	0.000	0.384***	0.000	0.388***	0.000
Design			–0.114	0.080	0.068	0.315	0.114	0.054
Context & service								
Structures								
Support			–0.063	0.412	0.093	0.160	–0.068	0.310
Participation			0.061	0.489	–0.028	0.720	0.061	0.412
Persons involved			0.134	0.096	0.110	0.166	0.185**	0.008
Outcomes								
Reliance			0.265**	0.002	0.195*	0.028	0.083	0.474
Performance			0.808***	0.000	0.784***	0.000	0.810***	0.000
Strain			–0.777***	0.000	–0.758***	0.000	–0.776***	0.000

Note. *p <0.05. **p <0.01. ***p <0.001

this relationship between context structures and trustworthiness to be significant. In addition to the trustworthiness per se, we also assessed the strength of the relationship between the properties of an IS and trust in the system. Our data showed that only the usability-trust association significantly decreased between T1 and T2 as well as T1 and T3 (Table 3). As the relationship between usability and trust diminishes over time, usability is particularly crucial for establishing trust during the early stages of interaction [48, 49]. Our findings can potentially be explained by considering the time users require to evaluate these system features. Through superficial interactions, users can quickly identify system malfunctions such as long loading times or errors influencing system features (e.g., reliability) [9, 25, 41]. Thus, reliability is significantly related to trust at earlier time points (T1 and T2) but not in the long term (T3). In contrast, credibility is significantly related at the later time points (T2 and T3) as evaluating the trustworthiness of its output [16, 52] may necessitate experiences with the system in various usage scenarios.

As a result, its impact on trust might only become significant after a sufficient number of interactions. Furthermore, system characteristics and organizational context features were consistently both rated more positively within companies with a linear increase of trust [49]. However, only 'Persons involved' were significantly associated with users' trust in the system. Although we initially hypothesized that trust in the IS following its launch (T1 to T3) would be influenced by both the system's properties and the contextual and service structures, our findings predominantly support the former. That is not to say contextual and service structures are not important. They do influence trust building in organizational settings. We simply highlight the inherent characteristics of the IS have a stronger influence. We also provide evidence that trust disposition holds particular importance during the initial stages of a user-IS relationship, especially when actual usage experiences are yet to occur [40, 43]. Consequently, the more specific experience a user has with an IS, the smaller the effect of trust disposition.

We investigated and proved intentional forgetting promoted by IS usage outside the laboratory as one of the first groups of researchers [48, 49]. We found that trust in an IS was positively associated with the user's forgetting of irrelevant work processes [48]. More specifically, users are more likely to forget those processes that are taken on by the system. Again, this research extends initial research, which found trust to promote intentional forgetting [24]. We, therefore, underline the capability of IS in a working environment to relieve the cognitive resources of workers in daily tasks. Furthermore, we also found a positive relationship between trust and forgetting of task-related information. Mirroring previous outcomes of trust in IS, we observe an increase/decrease in forgetting scores with increasing/decreasing trust (c.f., Fig. 3). Similarly, trust was consistently positively related to performance in all measurement points in the investigated case studies [48, 49]. Similar to the relationship between trust and well-being, we also observe an increase/decrease in performance scores with increasing/decreasing trust.

Our data showed that trust was negatively related to strain, aligning with earlier research claiming that trust in an IS is positively related to well-being [24, 45, 62]. A similar relationship was found in [49]. Studying the introduction of ISs in the

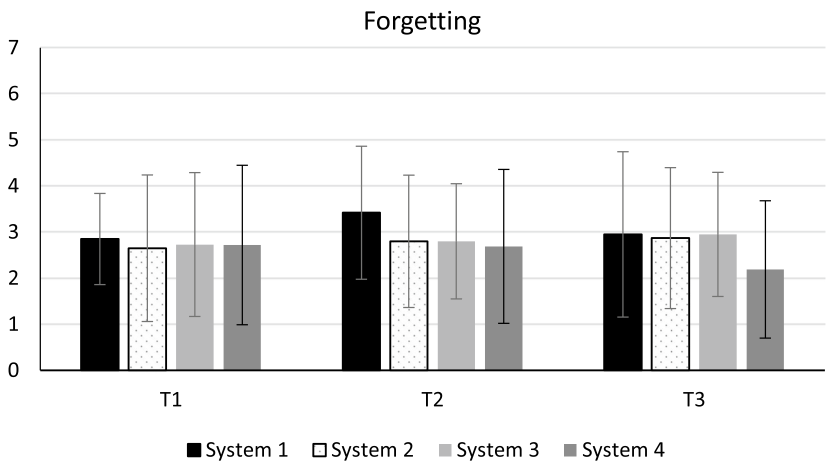


Fig. 3 Mean development of forgetting in case studies

real world, we found higher trust in the system to result in lower strain as reported by users. Strain continues to decrease over consecutive measurement points if trust continues to increase [49]. With trust scores of systems 1, 2, and 4 increasing from T1 to T3, we see a mirroring decrease in strain scores. Figure 3 presents the development of intentional forgetting in the case studies measured on a 7-point Likert scale.

While system properties were found to have a stronger correlation with trust compared to context and service structures, the question remains what actions organizations can take to actively foster employees' trust development. To explore these organizational strategies, we conducted interviews. Our analysis of the interviews identified three key strategies that potentially impacted the success of companies in introducing the information system: (1) identifying the most suitable system, (2) customizing the system to align with the company's needs, and (3) continuously evolving and improving the system.

1. Identifying the most suitable system. Before introducing the system, interview partners emphasized the importance of identifying the most suitable system for their company's specific needs. Each company conducted a detailed analysis of existing processes, allowing them to assess their usefulness and define requirements for the new system. This phase helped establish core functionalities and potential deal-breakers, aiding in the efficient filtering of available systems. Companies had to prioritize functions, recognizing that no system would meet every need perfectly. Engaging stakeholders in this process was crucial for both system selection and employee motivation during the transition.
2. Customizing the system to align with the company's needs. Rather than simply adapting the system to match current processes, they stressed the importance of critically assessing and, if necessary, modifying these processes to align with the system. Each interviewee also highlighted the necessity of extensive testing to

assess the system's practicality. Engaging experts from various departments in testing was seen as vital to ensuring a comprehensive evaluation and facilitating knowledge transfer across teams.

3. Continuously evolving and improving the system. Quick evaluations helped companies identify and resolve issues early, boosting user confidence and demonstrating a commitment to ongoing enhancement. Gradually adding new functions and maintaining close communication with users to build trust and ensure ongoing satisfaction. Actively involving users in system improvements fostered a sense of trust, as employees felt their needs were being heard and addressed. This approach helped ensure that the system could adapt as organizational needs evolved, reinforcing trust in both the system and the organization's commitment to continuous improvement. Ultimately, the interviews also requested approaches to actively govern trust. However, the character of trust poses a big challenge for organizations trying to govern user trust in an IS.

4.2 Managing Trust in Organizations

Historic research has shown that management approaches are instrumental in shaping solutions for the IS domain [8, 66, 67, 69]. A new approach, namely trust management, can focus on the two constitutional challenges related to the dynamic nature of trust: (1) building trust [59] and (2) repairing trust [2, 35]. In the subsequent part, we consolidate and elucidate the essential components required for trust management in organizational settings based upon the case study results for trust antecedents, moderators, and outcomes. Trust management is particularly relevant and beneficial if organizations are willing to continuously improve and change the deployed IS. This is twofold. First, trust management activities have the capability to comprehend the needs of users. Thus, the organization can adapt to the issues or concerns that were issued by the users (directly or indirectly via trust scores). Second, trust management provides a structured approach to managing and navigating transitions of trust dynamics as the IS evolves [41].

During trust building, changes in the antecedents of trust are particularly meaningful. Without a history of interaction, trust is mainly influenced by cues such as stereotypes and surrounding structures, which shape an individual's predisposition toward technology [35]. As users gain more exposure to the IS over time, they accumulate the necessary knowledge to make informed trust decisions based on their experiences [44]. Trust violations during the trust-building phase have a more apparent negative impact compared to violations occurring in later stages when trust is more stable. Thus, violations are especially harmful during trust confirmation or when attention to a project is at its peak [59]. In high-attention scenarios (e.g., the IS is launched; an update is released; a new feature is published), an incident in the form of reneging and incongruence has the strongest impact on trust decline [51]. Users have a specific expectation of the properties or functionality of a system. Not meeting these expectations can elicit negative emotions toward the IS [31, 38].

Furthermore, different mechanisms are at play for trust repair [1, 10]. For instance, verbal responses like apologies and explanations could be adapted as short-term solutions, with organizations offering apologies and explanations on behalf of the IS [10]. On the other hand, long-term violations might require mechanisms like improving the ISs trustworthiness through configurations that enhance functionality and reliability [41]. Thus, to restore trust to its pre-existing level, a combination of short-term and long-term trust repair mechanisms is required [10]. However, the nature of the trust violation, such as its frequency and severity, will impact the effectiveness of the trust repair strategies [35]. Lastly, while many trust repair strategies have been established in literature [29, 31], they have not been extensively tested in employee-IS relationships.

The phases of the trust management framework presented in this work combine the process theory of enterprise system success [39], the trust lifecycle [59], and business process management [67, 69]. While these concepts are established (management) approaches in their respective IS domains, focusing on trust as a dynamic construct within IS relationships necessitates a novel modus operandi. The framework presented in the following distinguishes between two main developments: (1) *adoption* and (2) *maintenance*.
For adoption, we integrate the first three phases of the process theory [39], i.e., chartering, project, and shakedown. Moreover, in line with previous longitudinal research on trust in IT artifacts [59], we consider that trust can exist in two states: (1) *building* and (2) *repair*. As a result, the terminology from [39] is adopted to a trust context. 'Chartering' is replaced with *pre-interaction* [59]. 'Project' becomes *confirmation*. Lastly, 'shakedown' is changed to *building*.
Maintenance (phase four of the process theory) is linked to the phases of the business process management life cycle [69]: Evaluation, analysis, configuration, and enactment. Yet, the boundary between these states is not rigid, as both can coexist and concurrently influence trust. Trust building primarily occurs during the adoption phase and gradually decreases over time. Trust repair, however, involves strategies and mechanisms to restore trust after a transgression that has diminished or negatively affected trust levels [10]. Notably, our proposed trust management model incorporates research on trust repair by [22, 36].
Lastly, we incorporate the trust lifecycle by [59] in our model. This model encompasses five phases: initial trust building, trust building, trust stability, trust dissolution, and trust repair.

The trust management framework, depicted in Fig. 4, encompasses a linear phase adoption, followed by a cyclic structure hereafter referred to as the *trust management cycle*. The latter comprises four phases. In the subsequent sections, each phase is described by providing a dedicated paragraph for each, along with a table that outlines the typical activities involved in trust management, potential errors or issues, metrics for evaluation, and possible outcomes associated with each phase.
Although building on each other, it is emphasized that smaller feedback loops exist, which are not depicted explicitly in the model. For instance, findings of an analysis step might be used to improve data gathering before actioning points of improvement.

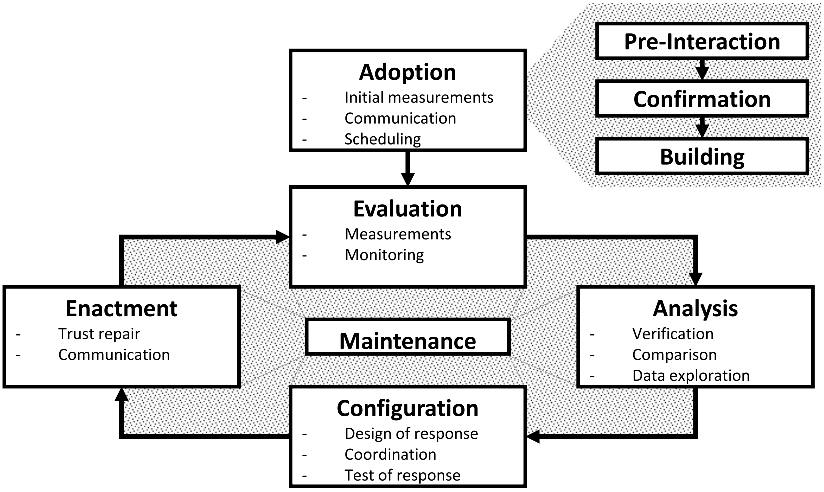


Fig. 4 Trust management framework

5 Discussion

This work investigates antecedents, outcomes, and moderators of trust in ISs in organizational contexts. It contributes to the existing body of research by affirming earlier propositions for trust building in ISs [24, 41, 44] and by advancing the existing findings with evidence from longitudinal research. The outcomes of this study's trust research propose a user's accumulated experience with an IS serves as a moderator for trust antecedents, such as trustworthiness [48, 49]. Furthermore, the study underscores the favorable outcomes associated with trusting IS usage within organizational settings. The level of trust in an organizational IS is significantly influenced by individual expectations. In unison with earlier research [41, 45], this work uncovers positive correlations between various system properties and trust. The results align with previous research that postulated a user's perceived trustworthiness of an IS is the trustor's evaluation of the inherent system characteristics [40]. Thus, a system's reliability, credibility, and usability significantly predict a user's perceived trustworthiness of an IS [48, 49]. Furthermore, it is crucial to recognize significant variations in the importance of individual characteristics of an IS when it comes to trust building. Specifically, trust disposition holds greater significance prior to the initial usage of a system.

This, again, is influenced by other factors (e.g., age, gender, and culture [25]), which can explain the varying levels to which individuals trust an IS before it is used. Ultimately, we confirm the idea of trust-building mechanisms involving technology to be dynamic. Regarding the effects of trustful ISs usage in organizations, the studies performed in this work indicate that trust in IS has various positive effects. Supporting earlier research [24, 45, 63], the results found trust to be positively related to well-being. However, when involving work tasks, trust in the system does not necessarily increase well-being; instead, the decline of well-being for all individuals performing work tasks is softened when the system is trusted. Trustful usage of the IS, therefore, contributes to a 'buffering' phenomenon. Additionally, trust in an IS is positively related to freeing cognitive resources of a use. Confirming prior research on the topic of intentional forgetting with IS [24, 45], the results of this work prove trust in an organizational IS is associated with releasing cognitive resources. Further, the results of this work show that high trust in an IS is associated with higher performance.

Naturally, freeing cognitive resources via intentional forgetting effects is one of the reasons for this outcome. At the same time, improved well-being can also contribute to increased performance. However, the underlying condition is always trust. Lastly, trust in a IS can increase an individual's willingness to rely on the information a user receives from that system [18]. When present, trust can lead to users being more likely to believe that the information provided is accurate, complete, and unbiased.

Building on the results of the topics above, this work contributes to a novel understanding of trust management for IS in organizations. The existing theoretical knowledge on trust, its influencing factors, and its outcomes are contextualized with management literature to construct the trust management approach. First and foremost, trust management in ISs is situated in the stages of trust building and repair. The understanding presented in this work aligns with the limited yet present literature focusing on the longitudinal dimensions of trust in IS contexts [28, 59]. Ultimately, this work shows that the management of trust in IS is a complex endeavor. Especially the management of expectations toward the IS, i.e., what inherent characteristics can be and cannot be expected, is crucial. Not meeting set expectations can result in trust violations.

6 Conclusion

This work offers a scientifically grounded definition of trust relationships, providing a foundation for classifying current and future research on trust. The proposed model distinguishes three actors and five trust directions between them. Building on this, the study synthesizes empirical and theoretical insights on the antecedents, outcomes, and moderators of trust.

It extends longitudinal IS trust research [59] and contributes to the emerging field of trust management [8], focusing on the employee–IS relationship in organizations. Drawing from IS [39, 69], management [6], and trust research [10, 22], the model serves as a basis for identifying gaps, integrating findings, or guiding future investigations.

For practitioners, trust in technology must be seen as dynamic. As influencing factors shift over time, organizations must adapt practices—such as communicating IS roadmaps—to maintain trust. Currently, trust strategies tend to focus outward, primarily targeting consumers [4], while internal trust toward IS often emerges incidentally. To unlock benefits like motivation and satisfaction [19], trust management should be institutionalized through evaluations, communication, and targeted interventions.

This work, while theory-driven and empirically informed, has limitations. Its conceptual model has yet to be practically validated, reflecting broader challenges in trust research [13]. The effectiveness of trust repair strategies in employee–IS contexts also remains unexplored [10]. Additionally, the interpretation of data is inherently subjective, leaving room for alternative readings.

Future research should consider developing benchmarks to evaluate trust strategies systematically. Quantifiable metrics would support both academic rigor and practical assessment of trust management efforts.

Finally, in light of the third rationality of ISs [32], IS implementation should be complemented by strategies that account for cultural and interpersonal dynamics. Viewing trust merely as a 'hygiene factor' [8, 14] risks undermining the full potential of technology. Instead, trust management should be understood as a dynamic, integrative construct that draws on insights from psychology and IS research.

Acknowledgements We thank all employees that gave us valuable insight into their work with the new information system. Furthermore, we thank Dr. Christoph Nohe and Dr. Miriam Höddinghaus for her support during data collection and preparation and Dr. Celeste Brennecka for her support with proofreading the manuscript.
Competing Interests The authors have no conflicts of interest to declare that are relevant to the content of this chapter.

References

1. Bachmann, R., Gillespie, N., & Priem, R. (2015). Repairing trust in organizations and institutions: Toward a conceptual framework. *Organization Studies, 36*(9), 1123–1142. https://doi.org/10.1177/0170840615599334
2. Bapna, R., Qiu, L., & Rice, S. (2017). Repeated interactions versus social ties: Quantifying the economic value of trust, forgiveness, and reputation using a field experiment. *MIS Quarterly, 41*(3), 841–866. https://doi.org/10.25300/MISQ/2017/41.3.08
3. Bjork, R. A. (1970). Positive forgetting: The noninterference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior, 9*(3), 255–268. https://doi.org/10.1016/S0022-5371(70)80059-7
4. Bozic, B. (2017). Consumer trust repair: A critical literature review. *European Management Journal, 35*(4), 538–547. https://doi.org/10.1016/j.emj.2017.02.007
5. Bravo, E. R., Santana, M., & Rodon, J. (2015). Information systems and performance: The role of technology, the task and the individual. *Behaviour & Information Technology, 34*(3), 247–260. https://doi.org/10.1080/0144929X.2014.934287
6. Cadle, J., & Yeates, D. (2004). *Project management for information systems* (4th ed.). Pearson Education.
7. Carnevale, D. G., & Wechsler, B. (1992). Trust in the public sector: Individual and organizational determinants. *Administration & Society, 23*(4), 471–494. https://doi.org/10.1177/009539979202300404
8. Chasin, F., Riehle, D. M., & Rosemann, M. (2019). Trust management—An information systems perspective. In *ECIS 2019—Proceedings of the 27th European Conference on Information*

Systems. Association for Information Systems, Stockholm & Uppsala, Sweden. https://aisel.aisnet.org/ecis2019_rip/57

9. DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: A ten-year update. *Journal of Management Information Systems, 19*(4), 9–30. https://doi.org/10.1080/07421222.2003.11045748
10. Dirks, K. T., Lewicki, R. J., & Zaheer, A. (2009). Reparing relationships within and between organizations: Building a conceptual foundation. *Academy of Management Review, 34*(1), 68–84. https://doi.org/10.5465/amr.2009.35713285
11. Easterby-Smith, M., & Lyles, M. A. (2011). In praise of organizational forgetting. *Journal of Management Inquiry, 20*(3), 311–316. https://doi.org/10.1177/1056492611408508
12. Elster, J. (1989). *The cement of society: A survey of social order*. Cambridge University Press. https://doi.org/10.1017/CBO9780511624995
13. Emamjome, F., & Rosemann, M. (2021). Managing trust—A design theory and design principles. In *Proceedings of the 42nd International Conference on Information Systems (ICIS 2021)* (pp. 1–17). Association for Information Systems, Austin, Texas, USA. https://aisel.aisnet.org/icis2021/dig_innov/dig_innov/22
14. Emerson, R. M. (1993). Power-dependence relations. In M. E. Olsen, M. N. Marger, & Fonseca, V. (Eds.), *Power in modern societies* (pp. 48–58). Routledge.
15. Eppler, M. J., & Mengis, J. (2004). The concept of information overload - A review of literature from organization science, accounting, marketing, MIS, and related disciplines. *The Information Society, 20*(5), 325–344. https://doi.org/10.1080/01972240490507974
16. Fogg, B.J., & Tseng, H. (1999). The elements of computer credibility. In *CHI '99: Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 80–87). https://doi.org/10.1145/302979.303001
17. Gefen, D. (2000). E-commerce: The role of familiarity and trust. *Omega, 28*(6), 725–737. https://doi.org/10.1016/S0305-0483(00)00021-9
18. Gefen, D., Benbasat, I., & Pavlou, P. (2008). A research agenda for trust in online environments. *Journal of Management Information Systems, 24*(4), 275–286. https://doi.org/10.2753/MIS0742-1222240411
19. Geister, S., Konradt, U., & Hertel, G. (2006). Effects of process feedback on motivation, satisfaction, and performance in virtual teams. *Small Group Research, 37*(5), 459–489. https://doi.org/10.1177/1046496406292337
20. Giffin, K. (1967). The contribution of studies of source credibility to a theory of interpersonal trust in the communication process. *Psychological Bulletin, 68*(2), 104–120. https://doi.org/10.1037/h0024833
21. Gill, H., Boies, K., Finegan, J. E., & McNally, J. (2005). Antecedents of trust: Establishing a boundary condition for the relation between propensity to trust and intention to trust. *Journal of Business and Psychology, 19*, 287–302. https://doi.org/10.1007/s10869-004-2229-8
22. Gillespie, N., & Dietz, G. (2009). Trust repair after an organization-level failure. *Academy of Management Review, 34*(1), 127–145. https://doi.org/10.5465/amr.2009.35713319
23. Hall, J. V., & Krueger, A. B. (2018). An analysis of the labor market for Uber's driver-partners in the United States. *Industrial and Labor Relations Review, 71*(3), 705–732. https://doi.org/10.1177/0019793917717222
24. Hertel, G., Meeßen, S. M., Riehle, D. M., Thielsch, M. T., Nohe, C., & Becker, J. (2019). Directed forgetting in organisations: The positive effects of decision support systems on mental resources and well-being. *Ergonomics, 62*(5), 597–611. https://doi.org/10.1080/00140139.2019.1574361
25. Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors, 57*(3), 407–434. https://doi.org/10.1177/0018720814547570
26. Hou, C. K. (2012). Examining the effect of user satisfaction on system usage and individual performance with business intelligence systems: An empirical study of Taiwan's electronics industry. *International Journal of Information Management, 32*(6), 560–573. https://doi.org/10.1016/j.ijinfomgt.2012.03.001

27. Irani, Z., Ezingeard, J. N., Grieve, R. J., & Race, P. (1999). A case study strategy as part of an information systems research methodology: a critique. *International Journal of Computer Applications in Technology, 12*(2–5), 190–198. https://doi.org/10.1504/IJCAT.1999.000203
28. Kanawattanachai, P., & Yoo, Y. (2002). Dynamic nature of trust in virtual teams. *The Journal of Strategic Information Systems, 11*(3–4), 187–213. https://doi.org/10.1016/S0963-8687(02)00019-7
29. Kim, P. H., Dirks, K. T., & Cooper, C. D. (2009). The repair of trust: A dynamic bilateral perspective and multilevel conceptualization. *Academy of Management Review, 34*(3), 401–422. https://doi.org/10.5465/amr.2009.40631887
30. Kluge, A., & Gronau, N. (2018). Intentional forgetting in organizations: The importance of eliminating retrieval cues for implementing new routines. *Frontiers in Psychology, 9*(51), 1–17. https://doi.org/10.3389/fpsyg.2018.00051
31. Kramer, R. M., & Lewicki, R. J. (2010). Repairing and enhancing trust: Approaches to reducing organizational trust deficits. *The Academy of Management Annals, 4*(1), 245–277. https://doi.org/10.5465/19416520.2010.487403
32. Kumar, K., van Dissel, H. G., & Bielli, P. (1998). The merchant of Prato-revisited: Toward a third rationality of information systems. *MIS Quarterly, 22*(2), 199–226. https://doi.org/10.2307/249395
33. Lankton, N. K., McKnight, D. H., & Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. *Journal of the Association for Information Systems, 16*(10), 880–918. https://doi.org/10.17705/1jais.00411
34. Lee, A. S. (1999). Rigor and relevance in MIS research: Beyond the approach of positivism alone. *MIS Quarterly, 23*(1), 29–33. https://doi.org/10.2307/249407
35. Lewicki, R. J., & Brinsfield, C. (2017). Trust repair. *Annual Review of Organizational Psychology and Organizational Behavior, 4*, 287–313. https://doi.org/10.1146/annurev-orgpsych-032516-113147
36. Lewicki, R. J., & Bunker, B. B. (1996). Developing and maintaining trust in work relationships. In Kramer, M. Roderick, & T. R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research* (pp. 114–139). Sage. https://doi.org/10.4135/9781452243610
37. Lewis, J. D., & Weigert, A. (1985). Trust as a social reality. *Social Forces, 63*(4), 967–985. https://doi.org/10.1093/sf/63.4.967
38. Lount, R. B., Jr., Zhong, C. B., Sivanathan, N., & Murnighan, J. K. (2008). Getting off on the wrong foot: The timing of a breach and the restoration of trust. *Personality and Social Psychology Bulletin, 34*(12), 1601–1612. https://doi.org/10.1177/0146167208324512
39. Markus, M. L., & Tanis, C. (2000). The enterprise systems experience—From adoption to success. In B. Zmud & M. F. Price (Eds.), *Framing the domains of IT research: Glimpsing the future through the past* (pp. 173–207). Pinnaflex Educational Resources.
40. Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review, 20*(3), 709–734. https://doi.org/10.2307/258792
41. McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems, 2*(2), 1–25. https://doi.org/10.1145/1985347.1985353
42. McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). The impact of initial consumer trust on intentions to transact with a web site: A trust building model. *The Journal of Strategic Information Systems, 11*(3–4), 297–323. https://doi.org/10.1016/S0963-8687(02)00020-3
43. McKnight, D. H., Cummings, L. L., & Chervany, N. L. (1998). Initial trust formation in new organizational relationships. *The Academy of Management Review, 23*(3), 473–490. https://doi.org/10.2307/259290
44. Meeßen, S. M., Thielsch, M. T., & Hertel, G. (2019). Trust in management information systems (MIS). *Zeitschrift für Arbeits-und Organisationspsychologie A&O, 64*(1), 6–16. https://doi.org/10.1026/0932-4089/a000306
45. Meeßen, S. M., Thielsch, M. T., Riehle, D. M., & Hertel, G. (2020). Trust is essential: Positive effects of information systems on users' memory require trust in the system. *Ergonomics, 63*(7), 909–926. https://doi.org/10.1080/00140139.2020.1758797

46. Müller, L. S., & Hertel, G. (2023, November). Trusting information systems in everyday work events—Effects on cognitive resources, performance, and well-being. *Ergonomics*, 1–18. https://doi.org/10.1080/00140139.2023.2286910
47. Müller, L. S., Meeßen, S. M., Thielsch, M. T., Nohe, C., Riehle, D. M., & Hertel, G. (2020). Do not disturb!: Trust in decision support systems improves work outcomes under certain conditions. In *MuC '20: Proceedings of Mensch und Computer 2020* (pp. 229–237). Association for Computing Machinery, Magdeburg, Germany. https://doi.org/10.1145/3404983.3405515
48. Müller, L. S., Nohe, C., Reiners, S., Becker, J., & Hertel, G. (2023). Adopting information systems at work: A longitudinal examination of trust dynamics, antecedents, and outcomes. *Behaviour & Information Technology*, 1–33.
49. Müller, L. S., Nohe, C., Reiners, S., Becker, J., & Hertel, G. (2024). *Developing trust in information systems at the workplace—A longitudinal field study*.
50. Petter, S., DeLone, W., & McLean, E. R. (2013). Information systems success: The quest for the independent variables. *Journal of Management Information Systems, 29*(4), 7–62. https://doi.org/10.2753/MIS0742-1222290401
51. Piccoli, G., & Ives, B. (2003). Trust and the unintended effects of behavior control in virtual teams. *MIS Quarterly, 27*(3), 365–395. https://doi.org/10.2307/30036538
52. Prat, N., & Madnick, S. (2008). Measuring data believability: A provenance approach. In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)*. IEEE. https://doi.org/10.1109/HICSS.2008.243
53. Rainer, R. K., &Prince, B. (2021). *Introduction to information systems: Supporting and transforming business* (9th ed.). John Wiley & Sons.
54. Rainer, R. K., & Prince, B. (2023). *Introduction to information systems* (10th ed.). John Wiley & Sons.
55. Rosemann, M. (2019). Trust-aware process design. In T. Hildebrandt, B. F. van Dongen, M. Röglinger, & J. Mendling (Eds.), *Business Process Management. BPM 2019. Lecture Notes in Computer Science* (Vol. 11675, pp. 305–321). Springer. https://doi.org/10.1007/978-3-030-26619-6_20
56. Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *The Academy of Management Review, 23*(3), 393–404. https://doi.org/10.5465/amr.1998.926617
57. Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An integrative model of organizational trust: Past, present, and future. *The Academy of Management Review, 32*(2), 344–354. https://doi.org/10.5465/amr.2007.24348410
58. Sidorova, A., Evangelopoulos, N., Valacich, J. S., & Ramakrishnan, T. (2008). Uncovering the intellectual core of the information systems discipline. *MIS Quarterly, 32*(3), 467–482. https://doi.org/10.2307/25148852
59. Söllner, M., & Pavlou, P. A. (2016). A longitudinal perspective on trust in IT artefacts. In *ECIS 2016—Proceedings of the 24th European Conference on Information Systems (ECIS)* (pp. 1–17). İstanbul, Turkey. https://aisel.aisnet.org/ecis2016_rp/52
60. Tashakkori, A., & Teddlie, C. (2003). Issues and dilemmas in teaching research methods courses in social and behavioural sciences: US perspective. *International Journal of Social Research Methodology, 6*(1), 61–77. https://doi.org/10.1080/13645570305055
61. Taylor, C. (1992). *Sources of the self: The making of the modern identity*. Harvard University Press.
62. Thielsch, M. T., & Hirschfeld, G. (2019). Facets of website content. *Human-Computer Interaction, 34*(4), 279–327. https://doi.org/10.1080/07370024.2017.1421954
63. Thielsch, M. T., Meeßen, S. M., & Hertel, G. (2018). Trust and distrust in information systems at the workplace. *PeerJ, 6*(1), 1–26. https://doi.org/10.7717/peerj.5483
64. Timotius, E., & Octavius, G. S. (2022). Stress at the workplace and its impacts on productivity: A systematic review from industrial engineering, management, and medical perspective. *Industrial Engineering and Management Systems, 21*(2), 192–205. https://doi.org/10.7232/iems.2022.21.2.192

65. Vanneste, B. S. (2016). From interpersonal to interorganisational trust: The role of indirect reciprocity. *Journal of Trust Research, 6*(1), 7–36. https://doi.org/10.1080/21515581.2015.1108849
66. van Aken, J. E. (2004). Management research based on the paradigm of the design sciences: The quest for field-tested and grounded technological rules. *Journal of Management Studies, 41*(2), 219–246. https://doi.org/10.1111/j.1467-6486.2004.00430.x
67. van der Aalst, W. M. P. (2013). Business process management: A comprehensive survey. *International Scholarly Research Notices, 2013*, 1–37. https://doi.org/10.1155/2013/507984
68. Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS Quarterly, 37*(1), 21–54. https://doi.org/10.25300/MISQ/2013/37.1.02
69. Weske, M. (2007). *Business process management: Concepts, languages, architectures*. Springer. https://doi.org/10.1007/978-3-540-73522-9

A User-Centered Development Pipeline for Persuasive Systems

Benjamin Weyers and Nico Feld

1 Introduction

Forgetting is an integral element of human memory [33, 40, 41] enabling adaptivity to the ever-changing environment and requirements of everyday life [6]. In this regard, *intentional forgetting* represents the process of the active regulation of knowledge and memory applied by a human being [5, 21, 30]. Besides episodic and semantic, procedural knowledge is an integral piece of knowledge kept in human memory and is the basis for the execution of tasks [26]. A specific type of procedures are habits, which can be defined as automatically enacted (activated by an internal or externally perceived trigger) behavioral pattern [18] entangled with a low demand on cognitive resources [4, 8]. Despite this positive effect of a low demand on cognitive resources habits may have on work-related tasks (by increasing productivity and lower the risks of errors due to more free cognitive resources), habits may be perceived as dysfunctional. A well-known example of such a dysfunctional habit is constantly checking for emails [25] or accepting irrelevant phone calls, as highlighted later in Sect. 5. In the first case, a suitable alternative behavior might be closing the email application or only checking them at specific intervals. In the second case, an additional step could be embedded into the routine to check the call's relevance according to the caller's number. Thus, it seems to be a suitable approach to apply the mechanism of intentional forgetting to procedural knowledge, where we focus on habits as a specific type of procedural knowledge in this work. Sonnentag et al. [35] have identified certain potential for implementation intentions [12] and vigilant

B. Weyers (✉) · N. Feld
Computer Science, Human-Computer Interaction, Trier University
Universitätsring 15, Trier, Germany
e-mail: weyers@uni-trier.de

N. Feld
e-mail: feldn@uni-trier.de

J. O. Berndt et al. (eds.), *Intentional Forgetting with Intelligent Systems*, Progress in IS,
https://doi.org/10.1007/978-3-032-17621-9_5

monitoring as mechanisms for adapting work-related habits. This might lay out the path towards mapping these cognitive constructs to actual system behavior.

In the case of digital systems, persuasive systems are useful to support adaptation processes. Persuasive systems are "computerized software or information systems designed to reinforce, change or shape attitudes or behaviors or both without using coercion or deception" [28, p. 164]. Such systems can be found in various domains, such as in supporting sustainable behavior [23] or supporting fitness through devices such as fitness trackers [16, 37]. Previously presented research results show that persuasive systems are more effective in case their function is individualized to the user and to the specific behavior [11, 29]. The research area of HCI, as well as research in the area of interactive system engineering and design, investigates various types of development and engineering processes for a requirement-driven and human-centered creation [34]. A well-known design paradigm in this regard is the "user-centered design process" [1], which is part of ISO 9241-210[1] standard. Applying user-centered design methods to engineering and designing persuasive systems seems promising in supporting the intentional forgetting of work-related habits.

As previously outlined, habits are a type of procedural knowledge with specific characteristics. Empirical research [35] suggests that work-related habits (besides others) show a high level of individualization. This emphasizes the need to design processes that both address the target group's requirements and include mechanisms as well as implementation approaches for accounting for individual behavior. For example, one may consider sports to be an integral element in maintaining a high level of health. Consequently, it is obvious that supporting bodily activity is an integral part of supporting health. This line of argument is employed to explain why fitness trackers are successfully used by the general public. Individualization of persuasion for these type of persuasive systems (as outlined as one of the major design guidelines for persuasive systems [29]) can be oriented along quite simple parameters, such as age or gender in the case of fitness trackers. Using findings from health research, certain rules of thumb can be applied and are likely to produce significant effects when applied by the user (e.g., [16]). However, in the case of work-related habits, such general rules seem to be less obvious. Therefore, this work presents a software engineering approach that supports the integration of individualized models of habits into persuasive systems to enable a user-centered and habit-related persuasion action.

This chapter presents an approach to persuasive systems, obtained by the composition of a pipeline from previous work [7, 19, 20, 35] conducted in the areas of user-centered design, formal modeling, and software development. In detail, this chapter presents the composition of various design steps and methodologies, which focus on the integration of process-related human behavior or habits to be adapted through persuasive systems from a subjectively perceived malfunctioning habit to a new and better one. In specific, the following contributions have been integrated:

---

[1] www.iso.org/standard/77520.html.

- Interview-based gathering informal data for work-related behavior [7, 35],
- producing bias-reduced or -free semi-formal models from informal (interview) input data [20],
- generating formal (reference net-based) models from semi-formal (BPMN) models [19], and
- demonstrating the use of such models in the implementation of a persuasive system in an immersive (VR-based) scenario.

We specifically focus on gathering a model of human habit as an essential point for the persuasive system. In addition, we will give more insight into potential implementations for persuasive systems at the office workplace as one potential context and use case.

2 A User-Centered Development Pipeline for the Implementation of Individualized Persuasive System in the Domain of Work-related Habits

Building interactive systems needs to employ suitable software engineering techniques and principles [2]. The suitability of software engineering approaches depends on the domain of the software being developed, which, in this work, is persuasive systems. To outline a better framing for the discussion of such an engineering approach, this section first presents a simple architecture for persuasive systems, followed by the presentation of a data processing pipeline that outputs a model enabling the persuasive system to identify deviations from a wanted towards an unwanted habit. The latter builds the foundation for the well framed persuasive intervention by the system.

2.1 A System Architecture for Persuasive Systems

In the following, we present our system architecture, which is based on the approach by Fogg [10] on developing persuasive systems as *tools*. Figure 1 illustrates the outline of our architecture. In this architecture, a *sensor component* perceives the user's behavior as well as its surroundings or context. This information on the user's status is communicated to the *persuasion model*. The integral functionality of the persuasion model is to enable the system to differentiate between unwanted and wanted behavior. Thus, the persuasion model includes a fusion of a model of wanted and unwanted behavior, which allows the persuasive system to asses which behavior the user is currently showing. In addition, the persuasion model includes additional logic that triggers the *persuasion strategy component* in case the user is currently showing unwanted behavior and persuasion towards the wanted behavior should be attempted. The model is embedded into an implementation to communicate those persuasion attempts to the persuasive strategy component [10]. The latter defines the

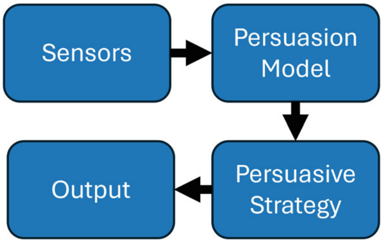


Fig. 1 A simple software architecture for persuasive systems, where the communication between the system and the outer world is modeled and implemented by the sensor and output components. Decision-making on when to persuade is implemented in the persuasion model considering information on wanted and unwanted behaviors, where the persuasion strategy defines the actual output generated being triggered by the persuasion model

actual type of feedback communicated to the user through the Output component, e.g., showing a notification to the user.

In the following section, we will explain the architecture using a well-known example: fitness trackers. Fitness trackers are a category of persuasive systems widely used. A fitness tracker can be worn by a user as a smartwatch on his/her wrist. Equipped with sensors, these watches can measure the wearer's movement activity. In this case, e.g., steps are counted, including information about heart rate. This information is available through the sensory input component. Combined with information about the user, such as age, weight, or gender, the persuasion model of such a fitness tracker would identify the difference between the current achieved training goal (e.g., an optimal number of steps) and an optimal training outcome based on the demographics of the user. By means of the persuasive strategy and based on the identified difference, the system gives feedback to the user, e.g., in textual form. This text is displayed to the user through the system's output channel, e.g., using the smartwatch's display accompanied with a trigger, e.g., a vibration, to catch the user's attention.

There are various works focusing on the design of persuasive strategies. Halko and Kientz [13] present eight main types of persuasive strategies. Furthermore, they differentiate between instructional style, social feedback, and motivation type. Oinas-Kokunnen also highlights various design guidelines for persuasive systems [29], including the need for a high level of individualization.

Previous works focus mainly on the design of effective persuasive strategies in the sense of output. However, an important but neglected aspect is developing a persuasion model from a user-centered perspective. They present various high-level aspects without giving clear advice on how to individualize a persuasion model in the sense of this work. One possible solution might be to gather formal (executable) descriptions of the users' behavior by applying user-centered engineering approaches, such as focus groups or quantitative evaluations. These descriptions can then be integrated in an exchangeable fashion into a persuasive system, thus, making the persuasion model interchangeable. However, this raises the challenge regarding gathering input data from the user to be transformed into a persuasion model. It can be assumed that such data is mainly informal input from, e.g., interviews. In the next section, we will tackle this challenge with an integrated pipeline outlined below. It specifically focuses on work-related habits as a matter of adaption and forgetting. Still, the pipeline might also be a pattern for other domains in which persuasive systems can be applied and a high level of individualization of the persuasion model is necessary due to the behavior to be adapted.

2.2 A Pipeline for Engineering a Persuasion Model

As outlined in the introduction, we focus on the implementation of persuasive systems to support the adaptation of dysfunctional habits. Because of that, we emphasize the formalization of the original and unwanted behavior as well as the alternative behavior, which should replace the unwanted one. In the following subsections, we will outline the developed pipeline shown in Fig. 2.

For the pipeline, we assume informal data resulting from interviews with users. Based on transcripts from these interviews, we have developed a manual transformation method [20] with the goal of creating mostly unbiased "Business Process Model and Notation" (BPMN) models [9]. Subsequently, an algorithmic transformation approach is applied using these BPMN models representing the current and wanted work-related habits (presented in [19]). In a two-step process, the individual habit descriptions are first translated into reference nets [17], a special type of Petri net, whereas in a second step, these nets are matched onto each other to fuse similar parts. Embedded into a persuasive system architecture as outlined above, this model is able to identify whether the current behavior is following the current or wanted behavior. With a final addition of controlling elements to the fused net, relevant information for the persuasion can be triggered and sent to the persuasion model. The latter will be demonstrated with an example at the end of this chapter.

The following sections outline each step in more detail and present previously published results [7, 19, 20].

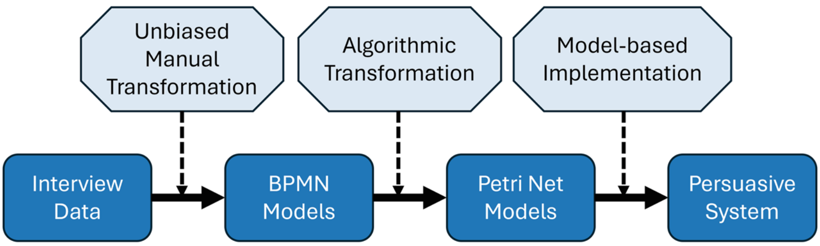


Fig. 2 Pipeline for the generation of formal persuasion models from informal input data, here interviews with users

3 From Informal Data to Semi-Formal Models

Besides other potential approaches for gathering data about work-related habits, we used an interview-based approach as has been outlined in our previous work [7]. In the interviews, we specifically asked potential users to think aloud about their habits. In this regard, we aimed to enable the interviewees to put themselves into the situation of showing the unwanted habit as well as to envision the wanted alternative. We relate to this type of interview as *think-aloud interviews*, as it borrows integral mechanisms from the well-established approach of think-aloud protocols [14]. The major difference is that for think-aloud protocols used in HCI and interactive systems engineering, the user is confronted with the study object (e.g., a user interface or interactive system) while he/she is thinking-aloud during use/interaction with the system. In our case, we asked persons to put themselves into a situation where they showed the to-be-described behavior.

In a preliminary online study, office workers were first asked to identify unwanted habits in their work life and then describe an alternative, desired habit. We wanted to identify work-related habits, understand their structure, and generate data for developing, testing, and analyzing the pipeline outlined previously. In addition, we used the data to guide potential psychological constructs and classify and identify certain characteristics. For instance, it turned out that work-related habits are highly individual [35].

Participants were supported with a magnetic whiteboard to model a BPMN (Business Process Model and Notation) representation of their habits. In addition, they got support from an interviewer, who gave them an introduction to BPMN models and various examples. As material for modeling, they were provided with magnetic markers representing a subset of BPMN nodes and could also add written notes with whiteboard markers for additional context. Subsequently, they were asked to recall a specific situation in their work life where they exhibited an unwanted or dysfunctional habit. They then modeled this behavior and imagined a desired alternative, which they also represented using the same materials.

In the second part of the study, participants verbally described both the unwanted and desired behaviors, following the previously outlined think-aloud approach. The order of BPMN modeling and verbal interviews was randomized across participants. Interviews lasted between 20 and 60 minutes, depending on the complexity of the responses. Combining visual modeling and verbal interviews, this mixed-method approach provided a deeper understanding of participants' habits and desired behavior changes.

We were able to collect BPMN models from the study outlined in the last section. Still, we observed an important limitation: without the active assistance of the interviewers, no syntactically correct models would have been generated, even though we had reduced the number of BPMN nodes to make the task more manageable for the participants. This highlights the challenges involved in guiding users who may not have prior experience with modeling or specifically with BPMN, as well as the importance of expert support in ensuring that the models are syntactically correct and meaningful. The gathered interview, in contrast, is not structured enough to serve as input data for a subsequent generation of formal models to be embedded into a persuasive system as a persuasion model.

Therefore, we presented a transformation approach to generate BPMN models from interview data with the control of bias introduced by the modeler [20]. The transformation process consists of three steps, where all three are conducted by a human modeler. In the first step, the transcripted (think aloud) interview is segmented into smaller text entities. This *segmentation* is done by means of verbs used in the sentences. This accounts for the relevant identification of tasks as central element in BPMN models. In the second step, the segments created are classified into various types of segments. The classes applied in this *classification* step focus on the later transformation into BPMN nodes or sub nets. Here, rules are applied explaining the semantical interpretation of these classes. Material is provided guiding the modeler in the classification.[2] In the final *modeling* step, the classified segments are transformed into BPMN (sub)nets by applying pre-defined templates also provided in the transformation material. Following the sequential order of the segments, the various subnets are fused into one coherent BPMN model.

This guided transformation approach has been investigated in terms of a potential reduction of bias induced by the modeler in an empirical study with forty-one participants. All details of the study are available in [20]. Study material, raw data, and translation model can be found on Zenodo.[3] In a between-subject design, the participants either used the manual provided or their modeling skills without further support. However, both groups got the same material explaining what BPMN is and what the goal of the modeling process aims at. Specifically, all participants went through a web-based learning application that explained BPMN and tested the knowledge gained by means of example models created by the participants. All participants had previous knowledge in modeling and had a computer science background. In total, two interviews of different complexity were translated into corresponding BPMN process models. Both groups received a previously segmented version of the interviews, such that the experimental group only conducted classification and transformation.

The results showed that in the experimental group, more segments were translated into fewer different models compared to the control group when focusing on the same segment. Thus, it seems that the translation manual supports the reduction of individual bias. Still, the study showed low power, which may argue for the potential of missing effects. However, we still observe much variation in the models produced by the groups, which argues that bias still affects the created models. In future work, the support by artificial intelligence seems to be promising if we consider this initial work on transformation as a starting point. Based on the proposed transformation manual, an algorithmic approach supported by knowledge-based systems being able to "understand" the semantics of segments might be a solution.

[2] https://zenodo.org/records/4704593.

[3] https://zenodo.org/records/4704593.

4 From Semi-formal Models to Formal Descriptions

To generate the differential model to identify the deviation of a user from the wanted behavior or sticking to the current, potentially unwanted behavior, we assume the corresponding BPMN models of both processes/habits are given. This might be the case using the previously presented transformation approach in Sect. 3. Previous work has considered Petri nets for the formalization of BPMN models [24, 31]. Considering the potential use of Petri nets in executable context through simulators, this type of formalization seemed to be suitable as it has been shown that BPMN models can be translated into Petri nets as well as being executed in actual implementations of interactive systems [38, 39]. In this work, we use reference nets [17], which are a specific type of Petri nets.

Reference nets are a specific type of colored Petri net used for modeling and analyzing complex, dynamic systems. They extend traditional Petri nets by incorporating object-oriented concepts, such as tokens that can carry structured data and references to other nets. This enables hierarchical modeling, where individual nets (also related to net instances) can interact with each other. For the coordination of various net instances, the concept of so-called *synchronous channels* can be facilitated. This concept also enables the coordination of firing transitions with calling methods in Java code and vice versa. For this, the simulator *Renew*[4] has been implemented, which not only executes reference net models but also enables synchronization with Java code. This makes it possible to use reference nets as an integral element of a Java-based implementation and, with this, makes it possible to use a reference net-based model as an integral part of the persuasion model as outlined above.

We adapted the transformation approach for BPMN by Chinosi & Trombetta [9] to represent more context [19]. The obtained approach is more adequate for the purpose of habit descriptions and execution as a persuasion model. Examples of the transformation templates are presented in Fig. 3. This figure visualizes the various types of subnets created from BPMN nodes. Each subnet starts with a transition and ends with a place. This enables a simple connection of subnets to each other. Furthermore, this makes it possible to fully automatically execute the transformation focusing on elementary parts in the BPMN model in the first step, where the full process is then created in a second step by simply combining the various subnets created.

Some subnets are also showing the needed synchronous channel inscriptions. These outline the connection points into the system's implementation, where sensor information is converted into trigger events forwarded to the reference nets. This is also the reason why, e.g., tasks have a transition for the start and end of a task. It is assumed that the system is able to identify (through context and task modeling as well as a sensory system) the start and end of tasks as well as occurring events.

For the fusion and splitting of the process, BPMN defines so-called gateway nodes. Two examples of such nodes are presented on the right of Fig. 3. Here, the upper example is a so-called XOR gateway, where either one of all outgoing processes

[4] http://www.renew.de/.

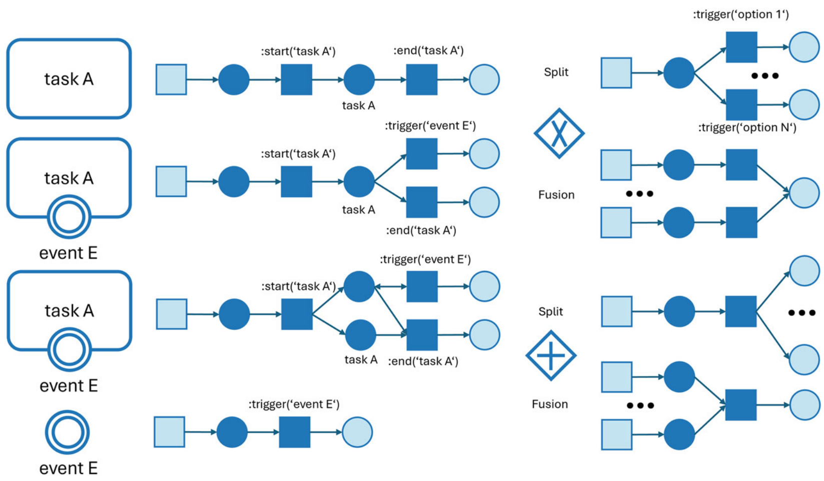


Fig. 3 BPMN to reference net transformation templates—a selection

is triggered (in case of splitting processes), and single processes are fused into one in case of fusion of processing. The second example (on the right lower part of Fig. 3) shows an AND gateway, where in case of splitting, all outgoing processes are triggered simultaneously, and in case of fusion, all incoming processes need to be finished before the single outgoing processes get started.

Developing a persuasion model to implement a persuasive system requires a structure that distinguishes between their current (unwanted) habit and the target (wanted) habit. Based on this observation and considering the previously outlined transformation algorithm, we outlined an approach to fuse reference nets presenting the current and target into one coherent reference net, which is able to identify the divergence of a user from the target to the current, thus unwanted habit.

The net fusion is based on a multi-step matching process, which includes the node inscription as well as the nets' structure [19]. The first step is a *syntactic matching*, where node labels are compared using the string edit distance (SED), which measures the number of changes needed to transform one string into another, ignoring spaces. Since the same person creates both diagrams, similar labels are likely. A low SED, therefore, indicates the same activity or task.

The second step is a *semantic matching*. While the SED is useful for string comparison, it is not always reliable to find the matching labels. A synonym dictionary can improve label matching by replacing words with similar meanings, identifying *check e-mails* and *read e-mails* as the same activity. Word stemming [15] further enhances accuracy by reducing words like reading and reads to read, leading to more precise SED calculations.

In the third step, a *structural matching* of the nets is performed. Therefore, the graph structure is analyzed after matching the nodes and their labels. Matching nodes are renamed with the same letter, while non-matching nodes are assigned unique characters. The reference net is then converted into a string representation, and the string edit distance is calculated. This conversion helps identify the nodes where the graphs differ.

Structural matching helps to identify direct relationships between nodes but struggles with indirect connections. Using a *behavioral matching* compensates for this. In behavioral matching, we compare the reachability graphs of the reference nets, which highlights the advantages of using reference nets as formalism.

The resulting net is a fusion of target and current behavior, which may be slightly extended by control structures, which then triggers the persuasion strategy in the system architecture. The next section will present an implementation of such a system, which demonstrates the method's applicability, including a potential use case.

5 Implementation of Persuasive Systems Using Formal Models of Human Behavior

When both the current and target behavior are represented within a single reference net, additional functionality can be embedded to define the logic of the persuasive system. This includes integrating logic for monitoring user behavior, triggering interventions, and processing system states in response to observed deviations. Within the reference net, transitions can be associated with specific events that signify key moments in the user's behavioral process. These transitions can trigger interventions, such as displaying visual cues, sending reminders, or modifying the system environment to encourage desired behaviors. By defining these elements within the reference net structure, the logic of the persuasive system remains entirely within the model, ensuring that all decisions and interactions follow the structured workflow established by the net.

To execute and interact with this logic in a real-world system, simulators like *Renew* can be utilized. *Renew* not only enables the simulation of the defined logic but also provides a direct interface for Java code integration. This integration is crucial, as it enables reference nets to both call Java methods and be triggered by external Java events, significantly simplifying the process of embedding persuasive functionality into a larger software framework. The system can seamlessly interact with sensors, external devices, and other application components by leveraging synchronous channel communication, connecting the persuasive strategy to real-world outputs as described in the software architecture in Sect. 2.1.

To evaluate this approach, we implemented a prototype of a persuasive system using the proposed architecture in a virtual environment. The prototype consists of two key components: the virtual environment along with the simulated sensors and the persuasive system, a server application that runs the persuasive model via *Renew*, manages communication with the sensors, and triggers persuasive actions through the persuasive strategy, like turning lights on (as a predefined output). Using a virtual environment enables us to test and develop persuasive systems without the restrictions of a physical environment, like simulating incoming calls or tracking the user's gaze directions, which would require additional hardware to implement in a physical environment.

To implement the communication of the persuasive system and the persuasive strategy, the reference net transitions are mapped to specific outputs and sensor events using synchronous channels, representing the "Sensor" and "Output" components of the software architecture:

- **Sensor transitions** are linked to external events, such as a user's inactivity or specific interactions with objects in the virtual environment. These transitions are eligible to fire only when the corresponding external event is detected.
- **Output transitions** initiate predefined actions within the virtual environment when they fire. For instance, a transition might trigger a reminder or change the color of an indicator light in the virtual office.

For example, in the first evaluation, we designed a virtual office of a real-estate manager tasked with common office work, like organizing their files & folders and writing and printing new documents while frequently getting interrupted by phone calls. Simply answering or declining every phone call has been implemented as an unwanted behavior. The wanted behavior would be to first check the phone number on a number board to see if the caller is an important client before answering or declining the call. Figure 4 shows an example reference net that includes the wanted (blue) and unwanted behavior (orange) that is generated by the presented pipeline described in Sects. 3 and 4. In the wanted behavior, after the phone starts ringing, the user looks at the number board before accepting or declining the call. When the user directly tries to accept or decline the call without looking at the number board, they show the unwanted behavior, which is detected by the reference net. The green area now shows the integrated logic to persuade the user, who is currently showing

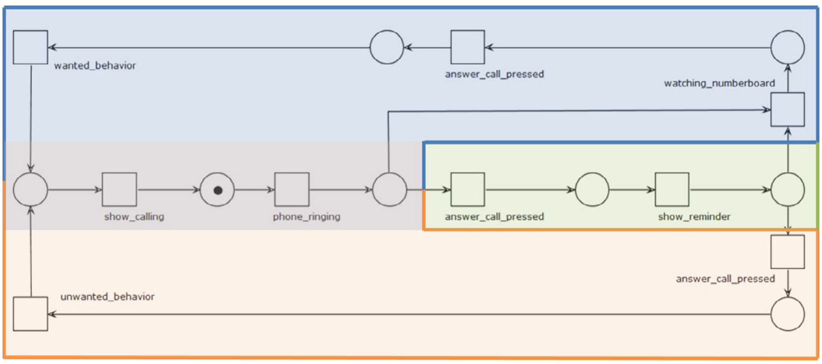


Fig. 4 This diagram shows the reference net of a persuasive model for handling interrupting phone calls. Blue nodes represent the desired behavior, checking the number board before answering, while orange nodes indicate the undesired behavior of answering immediately. The green area highlights the integrated logic of a potential persuasive system that intervenes with a reminder notification

unwanted behavior, by displaying a reminder to look at the number board. If the user looks at the number board, they again show the wanted behavior. However, if they again press accept or decline, they show unwanted behavior. Now, once the user reaches the "show_reminder" state, *Renew* calls a Java method, specified by the persuasive strategy, that triggers a specific output. In this example, the persuasive strategy defined the persuasive action as a text notification on the phone to remind the user to look at the number board.

Using a virtual reality-based test environment made it possible to track the state of the user, e.g., by tracking the user's gazing direction, as well as easily triggering interruptions in the form of phone calls. The *Renew* simulation tool linked the reference net with Java-based components that monitored user activity, like the gaze direction or phone presses, and triggered persuasive actions when unwanted behaviors were detected.

The integration with Java offers several significant advantages. It allows the system to reuse existing Java libraries and components, reducing development effort and enabling seamless communication with external devices and services. This makes it possible to build a highly modular and extensible system that can evolve over time. Additionally, the ability to call Java code directly from a reference net simplifies the implementation of complex logic, such as querying databases, analyzing sensor data, or integrating with cloud-based services.

A key strength of this prototype is its adaptability and ease of customization. The control routines can be tailored for individual users, allowing for the creation of personalized persuasive systems. Since the environment and control routines are loosely coupled, different sensors and outputs/actions can be added or removed with minimal effort. This ensures that the system can be easily extended to support new use cases or physical environments.

By leveraging the power of reference nets and the bidirectional interaction with Java code through *Renew*, the persuasive system achieves a high level of logical consistency, precision, and adaptability. This makes it possible to create interventions that are not only reproducible but also highly responsive to real-time user behavior, providing a robust foundation for persuasive systems in both virtual and physical environments.

6 Evaluation of Effectiveness of Persuasive Systems at the Work Place

The previously presented use case of interrupting phone calls includes an implementation of a persuasive system for supporting the forgetting of unwanted work-related habits. In the next step, we are currently conducting a controlled user study to evaluate the effectiveness of persuasive systems in the workplace. This study aims to investigate how different technologies can get users' attention to remind them of the target behavior while negatively affecting the main task as little as possible. We

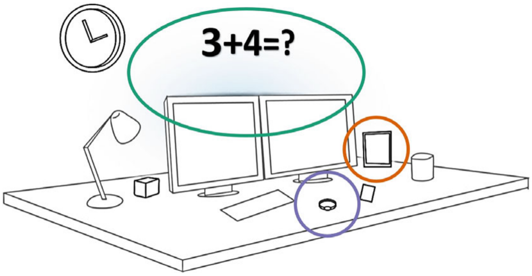


Fig. 5 Overview of conditions used in the user study. Green: In-situ projection with color-changing light. Orange: Digital picture frame. Purple: Smartwatch. The control condition did not use any technology

are comparing three technologies: a smartwatch, an in-situ projection system, and a digital picture frame. Each technology delivers a persuasive cue consisting of two components: an attention cue and an information cue, as shown in Fig. 5.

The attention cue is designed to capture the user's focus and differs based on the technology used. The smartwatch uses vibration as its attention cue, the in-situ projection system employs a color-changing light, and the digital picture frame shows an image as a visual cue. The information cue is consistent across all conditions and takes the form of a short text. In real-world applications, this information cue could be a context-specific reminder of the desired behavior the persuasive system aims to promote.

Due to privacy constraints in workplace environments, the use of personal sensors or external data, such as physical activity or incoming calls or emails, is restricted. As a result, the only unrestricted feasible sensor available for triggering persuasive cues is time. Therefore, the timing of interruptions is predefined and cannot adapt to real-time user behavior or context.

Before starting the current evaluation, we conducted a preliminary study to explore a suitable task design to measure the impact of these technologies on performance. The study compared the same three system variants. Participants performed an anagram-based cognitive task while the system interrupted them at predefined intervals. The information cue displayed a preselected text of the three possibilities: *"Please stay tuned!"*, *"Please don't lose focus!"*, or *"Please stay focused!"*. The participants were asked to press a button on the keyboard when noticing the interruption.

The results of the preliminary study yielded significant differences between a control condition and the three technologies but revealed three main aspects that have to be addressed for a more robust and insightful investigation:

- **Increased Sample Size Required:** The initial sample size was insufficient to draw definitive conclusions about the relative effectiveness of the three technologies. Therefore, the current study involves 180 participants to ensure sufficient statistical power.
- **Redesign of Main Task:** The anagram task used in the preliminary study did not induce a constant level of cognitive load, which is essential for evaluating interruptions. In response, we redesigned the main task into a resource allocation

task. To maintain a consistent cognitive load, we also introduced a puzzle task that participants engage with during potential downtimes, such as when all resources are currently allocated.

- **Interruption as a dual task:** The participants' comments and the experimenter's observations indicated that the participants acknowledged the interruption already when noticing the attention cue without looking at the information cue. Therefore, we added a simple math problem (e.g., $3 + 5 = ?$) as an information cue the users had to solve by entering the correct solution on the keyboard. The task performance of this task can then be used as an indicator of the technology's ability to get the user's attention.

The final study involves 180 participants who are randomly assigned to one of four conditions: three technology-based intervention conditions and a control group without interruptions. Each condition has $N = 45$ participants. During the study, participants perform a simulated office task for 40 minutes. The main task is a resource allocation task in which participants take the role of a ride-rental manager responsible for distributing a limited fleet of vehicles to various events. Successful resource allocation is rewarded with points, which is the main measure of task performance. Resource allocation tasks are a widely used type of task to assess the impact of interruption, as they induce a constant level of mental load throughout [3, 36].

To minimize user downtime and maintain a consistent cognitive load, a puzzle task is introduced during idle moments, such as when all vehicles are currently rented out. This puzzle task keeps participants engaged and allows them to earn additional points and temporary resources to be used in the main task.

As explained earlier, to assess the effectiveness of each technology in gaining attention, we have integrated a secondary task at fixed intervals. Participants are prompted to solve the math problem presented in the information cue as a response to the interruption. These interruptions are triggered at regular intervals, starting 10 minutes into the task and recurring every 7 minutes for a total of five interruptions.

The smartwatch delivers interruptions through vibration, providing discrete tactile feedback that is expected to minimize disruption while remaining noticeable. The in-situ projection system uses visual cues like color changes or ambient light to attract attention. Although this approach may be more intrusive, it offers an engaging way of delivering feedback. Finally, the digital picture frame displays an image as its attention cue, blending into the office environment with subtle and less disruptive cues.

We are collecting both objective and subjective data to assess the effectiveness of these technologies. Objective measures include task performance, interruption recognition rates, and resumption lag, which is the time taken to return to the primary task after an interruption [3]. A shorter resumption lag is considered an indicator of lower disruption caused by the persuasive technology. Subjective measures are being captured through standardized questionnaires to assess concentration [22], fatigue [27], and flow [32], both before and after the task.

We expect the smartwatch to have the shortest resumption lag due to its immediate and direct feedback. The in-situ projection system is expected to be highly effective in gaining attention, but it may be perceived as more disruptive in shared environments. The digital picture frame, while less intrusive, might have lower recognition rates, due to its subtle attention cue.

The findings from this ongoing study will provide valuable insights into the design of persuasive systems for office environments. The results will be employed to balance the need for attention capture with minimal disruption. Future work might build on these findings through longitudinal studies, exploring the long-term impact of such technologies on behavior adaptation and habit formation.

7 Summary

This chapter introduces a structured approach to designing and implementing persuasive systems that help users modify their behaviors by intentionally forgetting unwanted work-related habits. The development pipeline ensures that user insights are systematically transformed into structured models, guiding the creation of personalized and adaptive persuasive interventions.

The process begins with collecting user data through structured interviews, capturing insights about work-related habits and behavioral transitions. These insights are then represented as semi-formal BPMN models, which allow for clear visualization of current and desired behaviors. We presented a manual and guided approach for creating BPMN models from interviews with a reduced influence of modeler biases. BPMN models are converted into reference nets to facilitate deeper analysis and automation, which formalize the logic of intervention triggering and system responses.

To implement these persuasive systems, we integrated reference nets with Java using the *Renew* simulator. This integration enables seamless communication between the system and external devices, such as sensors and actuators, ensuring that interventions are delivered in real-time and tailored to user needs. The system's modular architecture allows for easy modifications and adaptations, making it highly flexible and scalable.

A major advantage of this architecture is its adaptability. Because the system separates control logic from environmental components, new sensors, outputs, and feedback mechanisms can be incorporated without significant reconfiguration. This makes it applicable to a wide range of domains, including workplace productivity, digital well-being, and personalized learning environments.

By embedding intervention logic within reference nets, the system ensures reproducibility, consistency, and precision in delivering behavior change interventions. The structured approach enables interventions to be systematically tested and refined, improving their effectiveness over time. Future research will focus on evaluating the long-term impact of these interventions and extending the framework to additional use cases beyond work-related behaviors, such as health monitoring and habit formation in everyday life.

Acknowledgements We thank Sabine Sonnentag, Wilken Wehrt, and Yuen Law for their contributions to developing the pipeline. Sabine Sonnentag, Wilken Wehrt, and Leon Többen contributed significantly to designing, evaluating, and discussing the presented empirical evaluations and studies. Julian Sawatzki conducted the prestudy for the evaluation, and Florian Langel implemented the VR prototype. This work has been conducted in the context of the DFG-funded project "iVAA" (318151256) as part of the priority program SPP 1921.

References

1. Abras, C., Maloney-Krichmar, D., Preece, J., et al. (2004). User-centered design. In W. Bainbridge (Ed.), *Encyclopedia of human-computer interaction* (pp. 445–456). Sage Publications.
2. Aït-Ameur, Y., Bowen, J., Campos, J., Palanque, P., & Weyers, B. (2021). Heterogeneous models and modelling approaches for engineering of interactive systems. *Interacting with Computers, 33*(1), 1–2.
3. Altmann, E., and Trafton, J. (2004). *Task interruption: Resumption lag and the role of cues*. In Proceedings of the 26th Annual Conference of the Cognitive Science Society.
4. Bargh, J. A. (2014). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In *Handbook of social cognition* pp. 1–40. Psychology Press.
5. Bjork, E. L., Bjork, R. A., & Anderson, M. C. (2013). Varieties of goal-directed forgetting. In *Intentional forgetting* (pp. 103–137). Psychology Press.
6. Bjork, R. A. (2013). Intentional forgetting in perspective: Comments, conjectures, and some directed remembering. In *Intentional forgetting* (pp. 453–481). Psychology Press.
7. Bowen, J., Weyers, B., & Liu, B. (2023). Creating formal models from informal design artefacts. *International Journal of Human-Computer Interaction, 39*(15), 3141–3158.
8. Chae, H., & Choi, J. N. (2019). Routinization, free cognitive resources and creativity: The role of individual and contextual contingencies. *Human Relations, 72*(2), 420–443.
9. Chinosi, M., & Trombetta, A. (2012). Bpmn: An introduction to the standard. *Computer Standards & Interfaces, 34*(1), 124–134.
10. Fogg, B. J. (1998). Persuasive computers: Perspectives and research directions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 225–232).
11. Fogg, B. J. (2009). A behavior model for persuasive design. In *Proceedings of the 4th International Conference on Persuasive Technology* (pp. 1–7).
12. Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. *Advances in Experimental Social Psychology, 38*, 69–119.
13. Halko, S. (2010). Personality and persuasive technology: An exploratory study on health-promoting mobile applications.
14. Jääskeläinen, R. (2012). Think-aloud protocol. In *Handbook of translation studies* (Vol. 1, pp. 371–373). John Benjamins Publishing Company.
15. Jivani, A. G., et al. (2011). A comparative study of stemming algorithms. *International Journal of Computer Applications in Technology, 2*(6), 1930–1938.
16. Konstanti, C., & Karapanos, E. (2018). An inquiry into goal-setting practices with physical activity trackers. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–6).
17. Kummer, O., Wienberg, F., Duvigneau, M., Köhler, M., Moldt, D., & Rölke, H. (2000). Renew–The reference net workshop. In *Tool Demonstrations, 21st International Conference on Application and Theory of Petri Nets, Computer Science Department* (pp. 87–89), Aarhus University, Aarhus, Denmark.
18. Lally, P., & Gardner, B. (2013). Promoting habit formation. *Health Psychology Review, 7*(sup1), S137–S158.

19. Law, Y. C., Wehrt, W., Sonnentag, S., & Weyers, B. (2017). Generation of information systems from process models to support intentional forgetting of work habits. In Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems (pp. 27–32).
20. Law, Y. C., Wehrt, W., Sonnentag, S., & Weyers, B. (2023). Obtaining semi-formal models from qualitative data: From interviews into BPMN models in user-centered design processes. *International Journal of Human-Computer Interaction, 39*(3), 476–493.
21. Lehman, M., & Malmberg, K. J. (2011). Overcoming the effects of intentional forgetting. *Memory & Cognition, 39*, 335–347.
22. Lindner, C., Lindner, M., & Retelsdorf, J. (2019). Measuring self-control depletion in achievement situations: A validation of the 5-item brief state self-control capacity scale. *Diagnostica*.
23. Lockton, D., Nicholson, L., Cain, R., & Harrison, D. (2014). Persuasive technology for sustainable workplaces. *Interactions, 21*(1), 58–61.
24. Lohmann, N., Verbeek, E., & Dijkman, R. (2009). Petri net transformations for business processes–a survey. In *Transactions on Petri Nets and Other Models of Concurrency II: Special Issue on Concurrency in Process-Aware Information Systems* (pp. 46–63). Springer.
25. Mazmanian, M., Orlikowski, W. J., & Yates, J. (2013). The autonomy paradox: The implications of mobile email devices for knowledge professionals. *Organization Science, 24*(5), 1337–1357.
26. McCormick, R. (1997). Conceptual and procedural knowledge. *International Journal of Technology and Design Education, 7*, 141–159.
27. McNair, D., Lorr, M., & Droppleman, L. (1992). Profile of mood states manual (Rev.). Educational and Industrial Testing Service, San Diego.
28. Oinas-Kukkonen, H., & Harjumaa, M. (2008). Towards deeper understanding of persuasion in software and information systems. In *First International Conference on Advances in Computer-Human Interaction* (pp. 200–205). IEEE.
29. Oinas-Kukkonen, H., & Harjumaa, M. (2009). Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems, 24*(1), 28.
30. Payne, B. K., & Corrigan, E. (2007). Emotional constraints on intentional forgetting. *Journal of Experimental Social Psychology, 43*(5), 780–786.
31. Raedts, I., Petković, M., Usenko, Y. S., Van Der Werf, J. M., Groote, J. F., & Somers, L. (2007). Transformation of BPMN models for behaviour analysis. In *The 5th International Workshop on Modelling, Simulation, Verification and Validation of Enterprise Information Systems* (Vol. 2, pp. 126–137). SCITEPRESS.
32. Rheinberg, F., Vollmeyer, R., Engeser, S. (2007). Die erfassung des flow-erlebens. In J. Stiensmeier-Pelster, & F. Rheinberg (Eds.), *Diagnostik von Motivation und Selbstkonzept, Tests und Trends; N.F.* (Vol. 2, pp. 261–279). Hogrefe.
33. Roediger, H. L., Weinstein, Y., & Agarwal, P. K. (2010). Forgetting: Preliminary considerations. In *Forgetting* (pp. 15–36). Psychology Press.
34. Sharp, H., Rogers, Y., & Preece, J. (2023). Interaction design: beyond human-computer interaction.
35. Sonnentag, S., Wehrt, W., Weyers, B., & Law, Y. C. (2022). Conquering unwanted habits at the workplace: Day-level processes and longer term change in habit strength. *Journal of Applied Psychology, 107*(5), 831.
36. Trafton, J., Altmann, E., Brock, D. P., & Mintz, F. (2003). Preparing to resume an interrupted task: effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human Computer Studies, 58*, 583–603. https://doi.org/10.1016/S1071-5819(03)00023-5
37. Wang, Y., Wu, L., Lange, J. P., Fadhil, A., & Reiterer, H. (2018). Persuasive technology in reducing prolonged sedentary behavior at work: A systematic review. *Smart Health, 7*, 19–30.
38. Weyers, B., Burkolter, D., Luther, W., & Kluge, A. (2012). Formal modeling and reconfiguration of user interfaces for reduction of errors in failure handling of complex systems. *International Journal of Human-Computer Interaction, 28*(10), 646–665.
39. Weyers, B., Luther, W., & Baloian, N. (2011). Interface creation and redesign techniques in collaborative learning scenarios. *Future Generation Computer Systems, 27*(1), 127–138.

40. Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology, 55*(1), 235–269.
41. Wixted, J. T. (2005). A theory about why we forget what we once knew. *Current Directions in Psychological Science, 14*(1), 6–9.

Self-(re)organizing and Especially Forgetful Personal Knowledge Assistants to Support Information Management and Knowledge Work

Christian Jilek, Heiko Maus, Tobias Tempel, Christian Frings, Claudia Niederée, and Andreas Dengel

1 Introduction

For more than 20 years now, the *Smart Data & Knowledge Services (SDS)* department (formerly *Knowledge Management* department) at the *German Research Center for Artificial Intelligence (DFKI)* has investigated methods for supporting users' (personal) information management (PIM) [63] and knowledge work [30] activities.

C. Jilek (✉) · H. Maus · A. Dengel
Smart Data and Knowledge Services Department, German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany
e-mail: christian.jilek@dfki.de

H. Maus
e-mail: heiko.maus@dfki.de

A. Dengel
e-mail: andreas.dengel@dfki.de

T. Tempel
Ludwigsburg University of Education, Ludwigsburg, Germany
e-mail: tobias.tempel@ph-ludwigsburg.de

C. Frings
Department of Cognitive Psychology, University of Trier, Trier, Germany
e-mail: chfrings@uni-trier.de

C. Niederée
L3S Research Center, Hannover, Germany
e-mail: niederee@l3s.de

A. Dengel
Department of Computer Science, RPTU Kaiserslautern-Landau, Kaiserslautern, Germany

J. O. Berndt et al. (eds.), *Intentional Forgetting with Intelligent Systems*, Progress in IS,
https://doi.org/10.1007/978-3-032-17621-9_6

Assuming that knowledge[1] emerges with the individual and eventually spreads into groups like the team, the department or the company, our solutions are typically centered around an *Organizational* or *Corporate Memory*, i.e. a computer system that continuously collects, updates and structures knowledge and information in the organization and context-sensitively, specifically and actively provides them for various tasks (see Abecker et al. [1] and Dengel et al. [32]). Especially in systems developed in our department, there is a *Knowledge Graph (KG)* [82] based on ontologies [93] at its core to represent (parts of) users' mental models (like persons, organizations, locations, topics, tasks, events, etc.) and interconnect them with personal and corporate information (like files, mails, bookmarks, databases, etc.). Based on such a Corporate Memory, *Personal Knowledge Assistants (PKA)* operating on users' individual desktops as well as the corporate intranet (cloud service) can be implemented. In interdisciplinary research, we and our partners (e.g. see co-authors) have worked on several variants and aspects of such assistants over the years. In this paper, we would like to provide a bit of an overview and some highlights of our research of the last decade whose focus was in particular on further increasing the degree of automation in PIM and Knowledge Work (KW) support. One major aspect in this regard was *digital forgetting* and in particular *Managed Forgetting* [59] serving as a "final puzzle piece" to realize true *self-organization* in this area: Without the ability to also select information items to be archived, condensed, possibly deleted, a system—as autonomous as it may be—can more or less only "move items around" without ever "shedding ballast", even if indicators would advise to do so. In order to decide which support measures to apply, the system observes user behavior and uses specifically tailored knowledge graphs for sense making and representing the user's information sphere.

The rest of this paper is structured as follows: Sect. 2 introduces necessary background information and foundations. A short overview of related work is given in Sect. 3. The paper's main part is Sect. 4, which details our research on self-organizing personal knowledge assistants. Last, the paper is concluded in Sect. 5.[2]

2 Background and Foundations

Self-organizing Context Spaces (or *cSpaces*, for short) combines three key concepts: the *Semantic Desktop* as an ecosystem, *Managed Forgetting* (a variant of intentional forgetting) and *Context Spaces*, which are a way of explicating user contexts. This

[1] Regarding the definition of "knowledge", we adopt the *knowledge ladder* by North et al. [81]: In short, it starts with *symbols* at the bottom. Adding *syntax* leads to *data*, adding *meaning* leads to *information* and *connecting information* including personal context, experience and expectations leads to *knowledge*.

[2] Note: This paper is a condensed version tailored to the aspects of Managed Forgetting and cSpaces. Readers additionally interested in knowledge graph construction and industry use cases of DFKI's corporate memory solution, CoMem, are kindly referred to a 73-page preprint available online [62]. Both versions contain excerpts of two of the authors' PhD theses [55, 90], which can also be used for further reading, since they cover the topics in even greater detail (over 500 pages).

section introduces these concepts in more detail. In addition, screenshots of some of cSpaces' prototypical implementations are shown in Figs. 1 and 2. Section 4.3 (second paragraph) also describes a typical interaction with such a prototype.

2.1 The Semantic Desktop as an Ecosystem

When creating Corporate Memory systems and PKAs based on them, we first need to solve the problem of *data acquisition*, i.e. tapping into and connecting existing data sources on users' personal computers, the intra- and internet, and building a *Knowledge Graph (KG)* from them (c.f. Schröder [90] for a detailed overview). Especially solutions developed in our department involve the *Semantic Desktop* [31, 86] as an ecosystem bridging users' local devices with the Corporate Memory residing on the intranet. Technically, the Semantic Desktop aims at bringing *Semantic Web* [16] technology to users' desktops. In short, information items (files, emails, bookmarks, etc.) are treated as a Semantic Web resources, each uniquely identified (by a *URI* [15, 17]) and accessible and queryable using the *Resource Description Framework (RDF)* [23, 27]. *Ontologies* [93] allow users to express personal mental models that interconnect these information items with their mental concepts like persons, organizations, locations, projects, topics, tasks, events, etc. An easy way offered by the Semantic Desktop is users simply tagging items with these concepts. Such a resulting semantic network at the core of the Semantic Desktop is therefore also called the *Personal Information Model* [87], or *PIMO* for short. As mentioned in the introduction, knowledge eventually spreads into groups, which is also reflected by the system: shared parts of users' individual PIMOs form a *Group Information Model (GIMO)* [70] as part of the Corporate Memory. In more recent terms, one may speak of PIMO and GIMO as a *Personal KG* [11] or *Corporate/Enterprise KG* [38], respectively. Several features of our recent Semantic Desktop prototype that is part of our productively used Corporate Memory system, *CoMem*, are illustrated in Maus et al. [71] and Jilek et al. [58] and the CoMem website.[3] Our department was among the Semantic Desktop pioneers bringing up the idea and technology in the early 2000s [85]. Although the hype around the topic finally subsided about a decade later, our department kept working in this area continuously. One of the more recent features using this technology as an ecosystem is the aforementioned *Managed Forgetting*, which is discussed in the next section.

[3] https://comem.ai/.

2.2 Managed Forgetting—A Variant of Intentional Forgetting

According to Dragan and Decker [34] (2012), one of the reasons why the Semantic Desktop was still not widespread at the time, although being superior to conventional desktops as pointed out by Franz et al. [37], was the absence of a "killer app", i.e. a highly beneficial feature (possibly not to be found elsewhere) making users adopt the technology. Managed Forgetting tries to fill this gap with its goal to increase the degree of automation of the Semantic Desktop in PIM and KW support scenarios aiming for a more tidied-up information sphere of the user, reduced cognitive load (especially avoiding overload), faster task switching, etc.

Managed Forgetting was envisioned in the EU project *ForgetIT*[4] (2013–2016) that brought together an interdisciplinary team of eleven partners from different universities, research institutes and companies across Europe. The idea was initially published by members of the team in two position papers [64, 80] and later extended by us in the *Managed Forgetting*[5] project (2016–2023) emphasizing more on self-(re)organization to finally read as follows:
"Managed Forgetting is an approach performed by a computer system to replace the binary keep-or-delete paradigm with an escalating set of measures instead. These measures range from 1) temporal hiding and inhibition, to 2) condensation, aggregation and summarization, to 3) adaptive reorganization, synchronization, archiving and deletion. The selection and application of measures, i.e. determining what to forget and what to focus on, is performed in a self-organizing and decentralized way based on observed evidences (typically gathered by user activity tracking). Managed Forgetting relies on two variants of information value assessment inspired by human memory and cognition: *Memory Buoyancy* for short- and medium-term information value and the *Preservation Value* as a long-term counterpart. The goal of Managed Forgetting is to complement not copy or replace human memory" [55].

The definition especially mentions Memory Buoyancy and Preservation Value as two variants of *Information Value Assessment*, i.e. measures trying to assign a score to information items reflecting their present or future value for the user. For example, a train ticket may have a high Memory Buoyancy score while one is on the business trip, but its value rapidly decreases once the trip is over and the ticket cannot be used for another ride. Although aiming for different perspectives (short/medium- vs. long-term), both concepts incorporate similar dimensions in their calculations like user activity and user investment, gravity, social aspects, etc. [58, 71]. In particular, Memory Buoyancy follows the metaphor that items losing relevance gradually "sink away", while those that are important (again) are pushed closer to the user by their higher buoyancy [64, 80]. Both concepts will be addressed in more detail in Sect. 4.

Another aspect worth pointing out is the implied paradigm shift allowing the system to reorganize items on a person's computer and especially select some of

[4] https://www.forgetit-project.eu/.
[5] http://www.spp1921.de/projekte/p4.html.en.

them for deletion (as well as executing the deletion at some point in time). Since the time computers became *personal* computers (i.e. a great number of people having one or more devices for themselves), users could rely on information being stored on their computer would remain there (hardware malfunction or manual deletion aside). This changes when introducing such self-organizing and especially forgetting-enabled ("forgetful") information systems (FIS [59], for short). They impose new challenges to usability, users' feeling of control, trust, etc. Imagine a user entering keywords into the search field of an FIS and no (or seemingly incomplete) results are shown. Users would wonder whether they entered the "right" keywords or whether they really saved the information item they are now looking for. Searching FIS as well as the aforementioned challenges are addressed more thoroughly in Sect. 4.

In the Managed Forgetting approach, deciding, for example, which items to forget and which to focus on heavily relies on *context*: the same information item could be very important in one context while being totally irrelevant in another. Thus, another rather recent feature based on Semantic Desktop technology we envisioned is *Context Spaces* [57] as introduced in the next section.

2.3 Context Spaces—User Context as an Explicit Interaction Element

When speaking of "context" we actually mean "a 'sense-giving environment' for a (given) nucleus, i.e. a context tries to represent relevant information items and their relations describing the given situation. Such a nucleus can be an activity (e.g. writing a scientific paper), an event (a meeting) or an information item itself (a PDF document, email, etc.). Because of the dynamics of situations, a context evolves over time. The context of a large research task (later containing many documents, links, notes, etc.), for example, could spawn from a small context having only an email calling for participation as its nucleus" [39]. Since every information item could become the nucleus of such a context, this definition leaves great freedom to the user. What belongs together in a person's mind should be representable as a context (or *Context Space* [57]) on their computer. In one of our industry use cases, there was a context space with a ticket as its nucleus for every failure incident reported by a customer or system health monitor. Like in folders, items in the same context should then also be displayed together, so their relatedness is clearly visible to the user countering the well-known *project fragmentation problem in PIM* [13], i.e. items of the same context being spread across various applications (files in the file system, mails in the email client, bookmarks in the web browser, etc.). This latter aspect is addressed by injecting context spaces into as many applications as possible. Details will be presented in Sect. 4.

In summary, with Context Spaces we see context as an active element users can work *in* and interact *with*. There are similarities to traditional folders but the idea of context spaces extends this concept as explained in the next section.

Having introduced these concepts, the next section presents *Self-organizing Context Spaces* [55] as a variant of a PKA that combines these three ideas.

2.4 Self-organizing Context Spaces as a PKA Variant

The idea of Self-organizing Context Spaces is to combine Semantic Desktop technology with Managed Forgetting and Context Spaces in the following way [55]:

Context Spaces. Have context as an explicit interaction element: *Context Spaces* allow knowledge workers to work *with* (i.e. a "tangible" object similar to a folder) and *in* contexts (i.e. immersion). These Context Spaces should be available in as many applications as possible fulfilling the vision of a single unified hierarchy structure (e.g. a tree or a directed acyclic graph) replacing separate file, mail, web browser and similar folder structures. Additionally, they provide necessary contextual information to select and perform appropriate support measures.

Managed Forgetting. Apply measures of *Managed Forgetting* like temporal hiding, condensation, reorganization, etc., to support knowledge workers with new kinds of services and higher degrees of automation not available in traditional systems. Exploiting the high amount of contextual (meta-)information makes this possible. Capabilities like condensation, summarization and ultimately deletion may significantly help with regard to long-term scalability of such a system (another reason for failure of preceding Semantic Desktop prototypes).

Semantic Desktop. Use the *Semantic Desktop* as an ecosystem to capture and represent a user's personal information model (PIMO) as well as contextual meta-information. The latter is either gathered automatically by means of user activity tracking or manually by users working with Semantic Desktop features like tagging or dragging items to a certain Context Space. An enhanced PIMO explicitly contains all of a user's Context Spaces.

As mentioned in the introduction, Managed Forgetting is not seen in isolation but as a component to realize self-organization. Thus, cSpaces aims at supporting the whole lifecycle of a context, from its initial spawning, to potential actions like splitting or temporal hiding in its main phase, to merging, condensation and potentially forgetting, archiving or deletion in its late phase.[6] More details on these operations are provided in Jilek [55] and a selection is also presented in Sect. 4.

The whole concept of Context Spaces and especially Self-organizing Context Spaces is explained in great detail in Jilek [55].

[6] Note: Technically speaking, cSpaces is not a self-organizing system according to the definition by Heylighen and Gershenson [53], i.e. it does not organize *itself* (briefly speaking). Using the term, we refer to the way cSpaces maintains its content (i.e. context spaces and their items) in a self-organizing and decentralized way by using operations like the ones described before.

3 Related Work

To our best knowledge, there is no approach strongly related to Self-organizing Context Spaces (cSpaces). This section thus provides a brief overview of related work for each of its key concepts and ideas in separation. For further literature pointers see Jilek et al. [62] or our individual papers on a particular topic.

Semantic Desktop and Similar Approaches. In general, there is the inspiring vision of *Memex* by Vannevar Bush [21] for many Semantic Desktop and similar approaches. As mentioned before, our department has quite a long track record in the area of organizational/corporate memories in general and the Semantic Desktop in particular. Apart from the well-known *NEPOMUK*[7] (2006–2008), the more recent *ForgetIT*[8] (2013–2016), *Managed Forgetting*[9] (2016–2023) and *SensAI*[10] (2020–2023) need to be mentioned. Apart from these projects, there is also work by Microsoft Research like *MyLifeBits* [42], the *Lumière Project* [54] and the assistant *Cortana*.[11] There was the *CALO* project (Cognitive Assistant that Learns and Organizes) with Semantic Desktops like *SEMEX* [33] or *IRIS* [25]. The well-known *Enron Email Dataset* [65] was also collected and prepared in this project. However, its most prominent outcome is presumably the assistant *Siri*[12] that started as a spin-off later bought by Apple.[13] Besides Cortana and Siri, there are also *Amazon Alexa*[14] and *Google Assistant*[15] (formerly *Google Now*) as further assistants by major corporations. Google also provided the 2011 discontinued *Google Desktop*.[16] Other works are *TaskTracer* [35] by Dietterich, Shen, Stumpf *et al.* or papers on user activity tracking and task detection [50, 51] by Gyllstrom *et al.*, who also published the information value assessment approach *LostRank* [49] (see next section). The group of Abela, Staff, Handschuh, Scerri *et al.* (some of whom were also members of the NEPOMUK project) worked on several related topics and presented solutions like *DCON* [88], *PiMx(T)* [3] or *(X-)iDeTaCt* [2, 4]. Last, there were three large projects, *APOSDLE*[17] (2006–2010) [68], *ACTIVE*[18] (2008–2011) [98] and *SWELL*[19] (2011–2017) [66], each publishing 60 or more papers, several of them relevant. For a slightly more detailed overview see Jilek [55].

---

7 https://nepomuk.semanticdesktop.org/.

8 https://www.forgetit-project.eu/.

9 http://www.spp1921.de/projekte/p4.html.en.

10 https://comem.ai/sensai/.

11 https://www.microsoft.com/en-us/cortana/.

12 https://www.apple.com/siri/.

13 https://web.archive.org/web/20220308041153/http://www.ai.sri.com/project/CALO.

14 https://developer.amazon.com/alexa.

15 https://assistant.google.com/.

16 https://web.archive.org/web/20110824073151/http://desktop.google.com:80/.

17 https://web.archive.org/web/20190503192712/http://www.aposdle.tugraz.at/.

18 https://web.archive.org/web/20110305113841/http://active-project.eu/.

19 https://web.archive.org/web/20210612234945/http://swell-project.net/.

Digital Forgetting and especially Intentional Forgetting. With regard to digital and especially intentional forgetting see other chapters of this book. Apart from the work by our colleagues, other works are, for example, in the area of *Robotic Memories* [47] or *Life Logging* [48]. With regard to PIM, Schmidt [89] applied an "aging mechanism as a form of controlled forgetting" in user context management and Bergman et al. [14] proposed a *GrayArea* as "an additional folder feature that allows the users to drag information items of low importance to a designated location at the bottom of a folder". There is also the area of *Machine Unlearning*, i.e. the removal of individual data points from a trained model (e.g. [19, 22, 46]).

Regarding Information Value Assessment (see the introduced Memory Buoyancy and Preservation Value), related approaches are, for example, the ideas of *waste data* [52] or *information waste* in the file system [99] or on the internet [6]. Turczyk et al. [96] investigated *file valuation* in the area of *Information Lifecycle Management (ILM)*, which seeks to store files on different storage systems according to their (business) value. Gyllstrom and Pedersen [49] proposed *LostRank*, an approach to estimate which documents are most likely to be lost for the user (e.g. important but not recently used) and are thus harder to re-find. Sappelli et al. [84] presented an approach for email importance estimation. Last, there is work by Attard, Brennan et al. (e.g. [9, 10, 20]) using the mostly synonymous term of *data value assessment*. Some of these approaches are more business centric, and others aim at more personal use cases. However, to the authors' best knowledge, Memory Buoyancy and Preservation Value are the only approaches that were designed and implemented according to findings of cognitive psychology about human memory and cognition.

Context Spaces and Self-organization Aspects in this Regard. Apart from our own prior research, especially work of the aforementioned ACTIVE project and the area of *Semantic File Systems* are to some extent related to the idea of working with and in Context Spaces and Self-organizing Context Spaces.

ACTIVE assumed that users are aware of the concept of context [45] and allowed them to select their current one. As a support measure, recently used document lists were updated according to the selected context [45, 98].

Allowing users to actually work in Context Spaces ultimately requires making them available in many (ideally all) of the users' applications. Especially the file system is interesting in this regard since it is available in all applications. Thus, it is an effective approach to utilize virtual or semantic file systems. To the authors' best knowledge, the paper that first mentioned the idea of a semantic file system is the one by Gifford et al. [43] (1991). Since then, more than 700 publications cited the paper, whereas only a fraction of them presents a semantic file system as well. However, seeing a user's knowledge work context as an entity to traverse has not been realized by any of these systems. The same is true for self-organization: none of these systems is self-(re)organizing except the one by Mesnier et al. [72], which addresses and applies self-reorganization in the scenario of low-level file system organization but not with respect to users' knowledge work contexts.

4 Self-organizing Personal Knowledge Assistants

Self-organizing Context Spaces, or *cSpaces* for short, were mainly developed as part of a PhD thesis of one of the authors (Jilek [55]) from 2018 to 2022—its development is ongoing for further research. It is a variant of a self-organizing PKA and actually not (yet) a single application containing all features mentioned in the following. Instead, features were developed step by step over several years and also re-combined or tailored toward certain research questions. Especially before work on cSpaces had started, experiments were conducted with the already productively used *CoMem* system. Over the years, CoMem, whose development started in 2011, further matured coming closer to an industry-ready product. This also made it necessary to "outsource" certain experiments and research questions to more lightweight, experimentation-friendly environments that are typically also more "fragile", i.e. less stable and maintained. cSpaces can thus be seen as a reduced, experimentation-tailored version of CoMem paving the way for potential future CoMem features.

In the remainder of the section, we first present the basic interaction cycle with systems like cSpaces or CoMem, followed by an overview of different user interfaces. Next, we share our insights on how working with and in *Context Spaces*, i.e. context an explicit interaction element, was perceived by test users so far and which benefits we could already identify or even quantify, respectively. This is followed by an overview of support measures offered by CoMem and cSpaces as well as further studies conducted, for example with regard to searching and trust in such highly autonomous systems.

4.1 Interaction Cycle

The interaction with cSpaces is basically a cycle of six steps [55, 59]:

1. **Evidence collection**: First, there is evidence collection, i.e. the system tracking user behavior like reading/writing files, clicking websites, etc. We, for example, created *tiny headless plug-ins* that simply *send out* in-app events to the Semantic Desktop—we therefore called them *plug-outs* [57]. More details on such plug-ins as well as Semantic Desktop architecture and re-engineering using plug-outs can be found in Jilek [55].
2. **Information extraction**: Next, the collected user activity evidence snippets are analyzed using, for example, information extraction methods. We mainly perform a variant of ontology-based Named Entity Recognition (NER), in particular identifying mentioned entities of the KG in the snippets. Items like files, email, websites, etc., are then automatically annotated with a *hasSuggestedTopic* relation, e.g. <DFKI website> *hasSuggestedTopic* <machine learning>. Since our system is a real-time assistance system, this analysis needs to be performed as fast as possible (preferably in only a few milliseconds) and desirably with a bit

of tolerance toward lexical variations like inflections, so the system does not "lose track" of users' activities by not detecting mentioned entities. In Jilek et al. [60], we presented a specialized NER approach combining (near-)real-time requirements with inflection tolerance.

3. **Context elicitation**: The third step is context elicitation. Given the evidence snippet(s) and the analysis result, the system tries to identify reasonable context candidates. Since users are directly interacting with the Semantic Desktop like explicitly selecting a context to work in or dragging items into a context space, the task can also be stated as predicting whether the user is still in the same context or whether they switched to another one (and if that is the case, to which one). Several context elicitation methods were presented in Jilek [55].
4. **Information value assessment**: In the fourth step, Memory Buoyancy and Preservation Value scores are calculated (see later parts of this section).
5. **Self-(re)organization measures**: Based on these scores, support measures may be triggered like temporal hiding, reorganization, condensation, etc. Details will follow in the remainder of the section.
6. **User interface updates**: Steps 5 and 6 are the actual support measures, whereas step 5 can be seen as the back-end part and step 6 addresses the front-end, i.e. updating the GUI appropriately.

After step 6, the interaction goes into its next iteration back at evidence collection. There are two additional aspects to be mentioned:

0. **Data storage and knowledge graph (PIMO)**: First, there is the data foundation (PIMO as a personal KG, databases, etc.) mentioned in the last section. This not only comprises data storage and connection to other systems but also context modeling. cSpaces extends the context model by Schwarz [91, 92], which itself is an extension of the one by Maus [69]. Both researchers already introduced various dimensions like organizational or causal aspects. These were extended in Jilek et al. [59] and Jilek [55] by hierarchy, forgetting and focal aspects. Storing and querying all this data imposes the challenge of combining structured and indexed search in near-real-time scenarios of highly evolving data (every click of the user may change the KG, the contexts, etc.), which is a large topic of its own. In Jilek [55], requirements for such a data store were summarized and an early prototype was developed and presented.
7. **Socio-technical issues and cognitive psychology aspects**: Several of our experiments, for example focusing on cognitive load or user's trust (e.g. see Sects. 4.3 and 4.6), were supported by research partners of cognitive psychology and ergonomics.

Having introduced the basic interaction cycle with the system, the next section presents some of its user interfaces.

4.2 User Interfaces

So far, three different user interfaces have been realized for cSpaces: a dashboard, a sidebar and transparent injections into existing systems like file or web browsers. Each of these interfaces offers a bit more familiarity to the user.

Dashboard. With the potentially screen-filling size of a *dashboard*, there is lot of space to visualize information and offer interaction possibilities for search, tagging, etc. There is, however, the drawback of having to leave the current application and switch to another app (the dashboard) typically taking up the screen's whole space obscuring the user's view on all or most other applications.

Sidebar. A *sidebar* mitigates this problem by occupying far less space but, on the downside, all widgets need to be more compact to fit into the sidebar, which typically means dropping visualization or input elements that would have been shown in the dashboard. Screenshots of different cSpaces sidebars are shown in Figs. 1 and 2.

Transparent injections. The highest familiarity with the environment is achieved using *transparent injections* of cSpaces into existing systems like the file, mail or web browser, calendar, etc. As presented in Jilek et al. [57], cSpaces uses standard protocols like *SMB* [73] (files), *IMAP* [26] (mails) or *WebDAV* [44] and its extensions *CalDAV* [29] and *CardDAV* [28] (calendar, address book) or similar quasi-standards like *WebExtensions* [74] (web and mail browser), supported by all major operating systems to inject contexts on a very low level without touching any operating system code. Especially the file system is interesting in this regard since it is available in all applications and thus makes cSpaces available in all of them. Higher-level actions like tagging items or writing comments could become cumbersome using only these injections. Therefore, a sidebar or dashboard is used as a complement. An example of a file system injection is given in Fig. 3: the figure shows a user browsing a specific context whereas the lower variant has been reorganized in the system.

4.3 Working with and in Context Spaces

One major idea of cSpaces is to enable users to actually work *with* (i.e. a "tangible object" similar to a folder) and *in* (immersion) context spaces. This especially allows for unified browsing of contexts that are close(r) to a person's mental model and easy ways of tagging/associating items with these contexts. "Modeling" (in the sense of users making more of their mental model explicit for the machine) should be easy and subtly possible for the user, e.g. dragging a file to a context makes the system automatically create an *isContainedIn* relation. If the context represents a calendar event, e.g. a business meeting, then the meeting can be associated with entities found in that file, etc. cSpaces addresses all three *pressing requirements for future PIM systems* stated by Warren [97]: combat information overload, ease context switching and support information integration across a variety of applications.

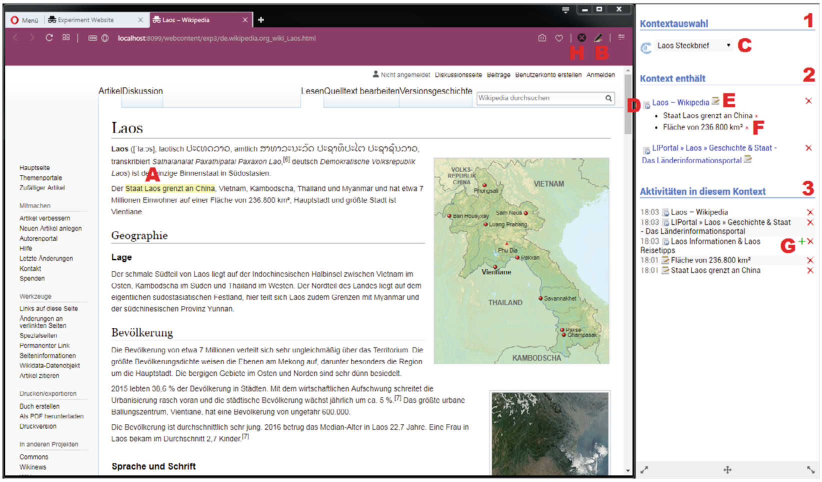


Fig. 1 cSpaces prototype for several user studies

We have already conducted multiple studies to gather user feedback on the system and to quantify potential benefits. Figure 1 shows a typical experimentation scenario with cSpaces: users interact with the sidebar (right) next to their usual applications, in this case a web browser (left). In this example, the sidebar consisted of three parts: the context selection (1), a "context contains" section filled with the items of the selected context (2) and an activity history showing the (last) actions performed in the context in reverse-chronological order (3). The screenshot particularly shows a web research scenario, in which participants investigated a certain topic, e.g. information about a country or famous person. They could add items (webpages) to the currently selected context (C) by clicking on the plus symbol next to the webpage's entry in the activity history (G). The webpage then appeared in the "context contains" area (D). They could also write notes associated with the webpage (E) or use a highlighter in the web browser (A and B), whose selected text passages were automatically converted into notes (F). To ensure that all browser tabs (except for the experiment instructions) were closed when participants were, for example, asked to switch contexts, a special tab closing app was developed (H). This prototype or slightly varying ones were used in several studies presented in the following.

Cognitive Offloading Effects

In a first study, participants were asked to do web research on a certain topic. The cSpaces assistant in a variant as described above captured their respective context, they could take notes, etc. In the middle of their task (after a few minutes), they were interrupted to switch to another completely unrelated context, the solving of modular arithmetic problems by Bogomolny [18] (referring to Gauß [40]). These equations are known to "place heavy demands on working memory" as stated by Beilock and Carr [12] referring to Ashcraft [7] and Ashcraft and Kirk [8]. Before switching to the math context, participants were informed whether their previous context (web research) would be kept or deleted. In case of the former, they could rely on the system keeping their context and bringing it back up when needed later. Participants whose context was kept were actually able to solve more equations than the deletion group (10.3 vs. 9.4, $n = 48$, $p < 0.01$). Although it was only a weak effect, the difference in means was significant. Participants whose context was kept were able to "cognitively offload" [83] their so far collected facts in the research task to cSpaces freeing working memory capacity, which then was additionally available for the math task. Our colleagues of cognitive psychology had already shown the effect in a scenario of encoding and recall of word material [94]. With this experiment, we were able to transfer these findings to a more complex scenario like knowledge work. For details please see Gauselmann et al. [39].

In a follow-up experiment [41], we could show the same effect more directly using *functional Near-Infrared Spectroscopy (fNIRS)*, "a neuroimaging technology for mapping the functioning human cortex" [36].

Effects on Task Resumption/Switching

In a second study, we investigated effects on context/task switching with the goal of quantifying participants' so-called *Task Resumption Lag (TRL)* [5], which we understand as the time a person needs to re-adjust their mind to the former context in order to continue where they left off. Participants were asked to plan a barbecue evening that made them do research like checking emails by guests, in which they stated what they would like to eat or bring for the party, vendor websites with prices, etc. Step by step, they had to fill an order list with amounts of food and beverages to be ordered. From time to time (in each case after a few minutes), they were interrupted to work on an unrelated task (working on a wiki entry about autonomous driving). Each participant was interrupted and a few minutes later resumed the main task three times. As shown in Fig. 2, the experiment had three groups: Group G1, a control group, had a non-interactive sidebar: it only showed a concise content analysis of the currently browsed email or website (1), basically the most prominent entities found in the text. Group G2 additionally had the possibility to add items to a single context-free folder as well as adding notes to them (3). Their activities were captured in an activity history list (4). Group G3 also had these possibilities but with the additional feature of being able to switch contexts (2). As a consequence, items, notes and activities stayed separated by contexts and were not mixing up. Figure 2 shows how content of the two contexts—from the experimenter's view: main task and distraction task—colored in yellow and red, respectively, mix up in the sidebar of G2, whereas they stay separated for G3. After an initial tutorial to get to know the system's different features, each participant worked in total for 20 minutes on the main task and 12 minutes on the distraction task. The experiment contained 40 different webpages—emails were shown as part of an online mail client, so the whole experiment could

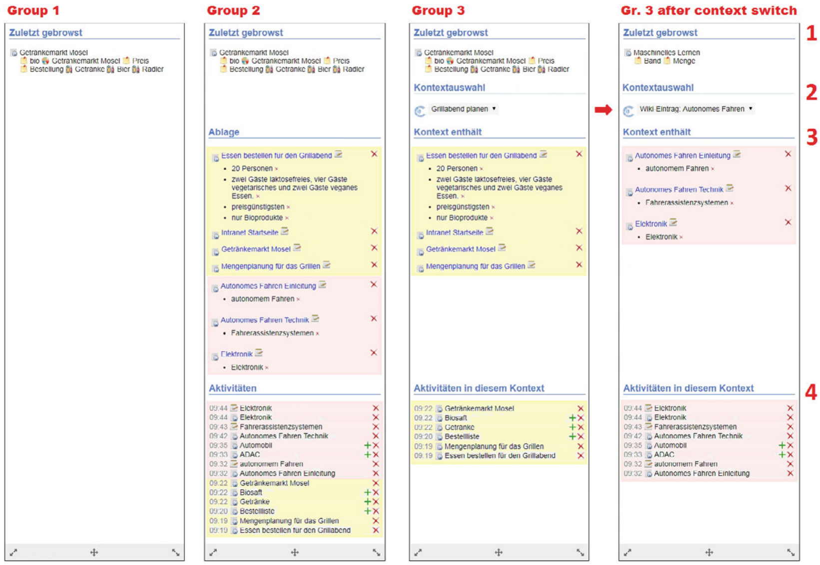


Fig. 2 Task switching with three different variants of the cSpaces sidebar

be conducted using just a web browser. With the experiment having 51 participants, we captured 153 observed context switches as well as one hour of video material and 1000 logged activities on average per participant. Participants were not asked to explicitly state when they felt their switch back was complete (mainly due to the experiment initially having a slightly different focus; however, the rich amount of obtained data could also be used for the TRL analysis in a secondary study). The TRL was measured by analyzing the captured evidences. It was assumed that a person was back in their former context (1) when getting back to pages actually containing content needed to fill the remaining lines of the order list or (2) adding new items or notes to the current context. To detect (1), all 40 web pages used in the study were classified as being either offerings and prices, amounts of food or beverages brought or requested by guests, navigation pages, spam without any relevant information or the order list to enter results. Next, this classification was applied to the activity logs in order to calculate the TRL for each of the 153 context switches. Last, these results were compared to the captured video material and corrected where necessary. For instance, this was the case when users navigated through several pages very quickly, staying on each page for less than a second. Even if a page classified as content was among those, it was assumed that the detention time was too short to actually conceive anything written on that page. Thus, the next content page visit or sidebar addition in the activity log was used to calculate the TRL.

The average TRLs for each group were as follows: 18.3 seconds for G1, 11.0 seconds for G2 and 10.5 seconds for G3. cSpaces-supported groups were thus about 40% faster in resuming their former context than the control group. The small difference between G2 and G3 (not significant) can be explained by only having two different contexts, of which one even was a distraction task and discovered as such by some participants: we observed rare cases of G2 participants taking the time to tidy up their sidebar by removing items and notes after returning from the distraction to the main task. It is worth noting that in post-experiment interviews, 78% of G2 participants stated (or at least insinuated) the need for a context-sensitive sidebar not mixing things up. More contexts would have presumably been necessary to see a larger TRL difference between G2 and G3. However, the differences of G2 and G3, respectively, to G1 were significant ($p < 0.001$) and can be classified as a strong effect. Further details on the experiment can be found in Jilek et al. [61] and Jilek [55].

Further Insights of a Multi-month User Study

As a complement to the previously mentioned short-term studies, we set up a multi-month user study, in which a smaller group of participants could use a more advanced prototype for up to five months (some users joined later in the study). The group of participants consisted of seven fellow researchers and students. They were asked to integrate the application into their daily work, which let to some of them using the app nearly every work day and others only a few days per month. In total, 249 days of use could be captured, consisting of 46,552 individual user events, 173 context spaces, 165 browsed files and 8393 browsed webpages. The study ended with a structured interview of 45 questions on 15 aspects of cSpaces. For example, all participants clearly agreed that

- the ideas behind cSpaces were easy to understand,
- learning how to use the app was easy and its usage could easily be remembered,
- taking notes on items in a context was exactly where they would like to have such notes,
- switching between different user interfaces like dashboards and sidebars would help in being more efficient and
- that they had not seen something like cSpaces in existing software applications.

Besides rating and describing their experience with the current prototype, they were in some cases also asked to extrapolate their gained experience to envisioned or early features they knew about (e.g. by seeing them in demos) but which were not part of the current prototype, yet. Such items were for example:

- whether items were right where participants would like to have them using context spaces,
- whether such context spaces were more aligned with their mental model than with traditional systems,
- whether cSpaces would allow for faster context switches,
- whether cSpaces helped in keeping their information sphere more tidied up or
- cSpaces showing only the current context while temporarily hiding others reduced cognitive load while working.

Being a research prototype, the cSpaces application still had usability issues that were less problematic in short-term experiments. However, in a more productive medium/long-term use, some "quality of life" improvements would have been helpful, e.g. clustering of the activity history, introducing better scrolling for certain widgets, etc. Note that these were not conceptual problems: users could give separate ratings for the current and the envisioned prototype and some problems were clearly attributed to the current but not the envisioned one. In general, participants were asked to give their honest opinions and ratings, and they actually pointed out the problems above. Thus, a courtesy bias [67], i.e. known or friendly persons giving more positive answers than they would have with an unknown experimenter or the prototype of an unknown person, can probably be ruled out for this study. In summary, one may conclude that the current prototype was already perceived as helpful (in total, ratings were slightly on the positive side of the spectrum), whereas even more potential was seen in the envisioned or early features (ratings very much on the spectrum's positive side). For a detailed break-down of all results see Jilek [55].

4.4 User Support Measures

This section gives more insights into user support measures offered by cSpaces and/or CoMem. They are an implementation of Managed Forgetting comprising ideas like condensation, temporal hiding or reorganization (see Sect. 2 for the complete definition).

Condensation and Summarization

An example of this category is **PIMO Diary**, a diary generator proposed in Jilek et al. [56]. It takes a person's personal KG (PIMO) and a given timespan as inputs and tries to cluster items and events in a way that makes the diary interesting to read/browse. There should be, for example, a certain amount of diversity within the diary, i.e. events belonging to the same project or topic should rather become one cluster instead of several ones about more or less the same. The app generates headlines for each entry and shows associated media like photos, the most important associated concepts of the KG as well as a word cloud of the most important terms. Users can zoom-in and zoom-out or shift the focus toward certain topics. PIMO Diary is basically a clustering approach using text and concept embeddings. Our colleagues also investigated summarization methods on **microblogs** [75–77, 79].

Temporal Reorganization, Fading Out and Resurfacing

A feature directly driven by Memory Buoyancy (MB) is **Fading out and Resurfacing**. MB tries to assign a score to each information item that represents the current value for the user. It was created according to findings of human memory and cognition and its design principles in CoMem were presented in Maus et al. [71]. A first evaluation was conducted in Tran et al. [95]. Later, Jilek et al. [58] presented an advanced version to better cope with different contexts and sudden switches.

Another example has already been presented in Fig. 2, when users switched from one context to another, the context sidebar (as well as all injections although not shown in the screenshot) was dynamically reorganized to fit the newly selected context's content. While this can already be seen as a simple form of reorganization, the idea can be further extended. Using so-called **Intelligent Folder Injections (IFI)**, the current context's content can be further divided with regard to certain criteria. For example, an IFI could contain less important items of this context that are, however, associated with hot topics of other contexts. A context's last focus could be another IFI. Another example we often use is the binary division of a context's content in an active and a forgotten part, whereas the forgotten part is hidden by default but accessible using another IFI. Figure 3 shows the example of a real-world context, the one of a project not visited for several years. After returning to that context, users would typically find a situation like in the upper screenshot: an overwhelming number of files and folders and it is at first unclear what were the important items. In the lower screenshot, items with low MB scores are hidden by default (accessible via a "forgotten" IFI) and only the most relevant remain. Since screenshots can hardly get across the dynamics of interacting with such a system, we encourage readers to watch a short technical demonstration video available online.[20]

In the aforementioned multi-month user study, participants unanimously agreed to the statements that IFIs are easy to understand and that they are an easy way to work with an highly autonomous assistant like cSpaces and the results of its hiding, condensation or reorganization measures. The previous examples apply to temporal as well permanent reorganization. Further examples of the latter are presented in the next section.

Permanent Reorganization

The last section showed examples that can be adjusted to work temporally as well as permanently. Another example of permanent reorganization is related to the *Preservation Value (PV)*. The PV was designed as a long-term counterpart to Memory Buoyancy. A first prototypical implementation in CoMem focusing on **photo preservation** assessed six dimensions: *user investment* (e.g. usage or number of annotations or comments added to an item), *gravity* (e.g. type-based heuristics or connectivity in the KG, in particular closeness to important things), *social graph* (e.g. relations to important persons, certain persons on a photo), *popularity* (e.g. ratings, number of views), *coverage* (e.g. reaching a certain coverage of a user's data, e.g. at least one photo per photo collection) and *quality* (e.g. photo quality). In a three-week

---

[20] https://comem.ai/cspaces/ or https://arxiv.org/src/1805.02181v1/anc/demo_video.mp4.

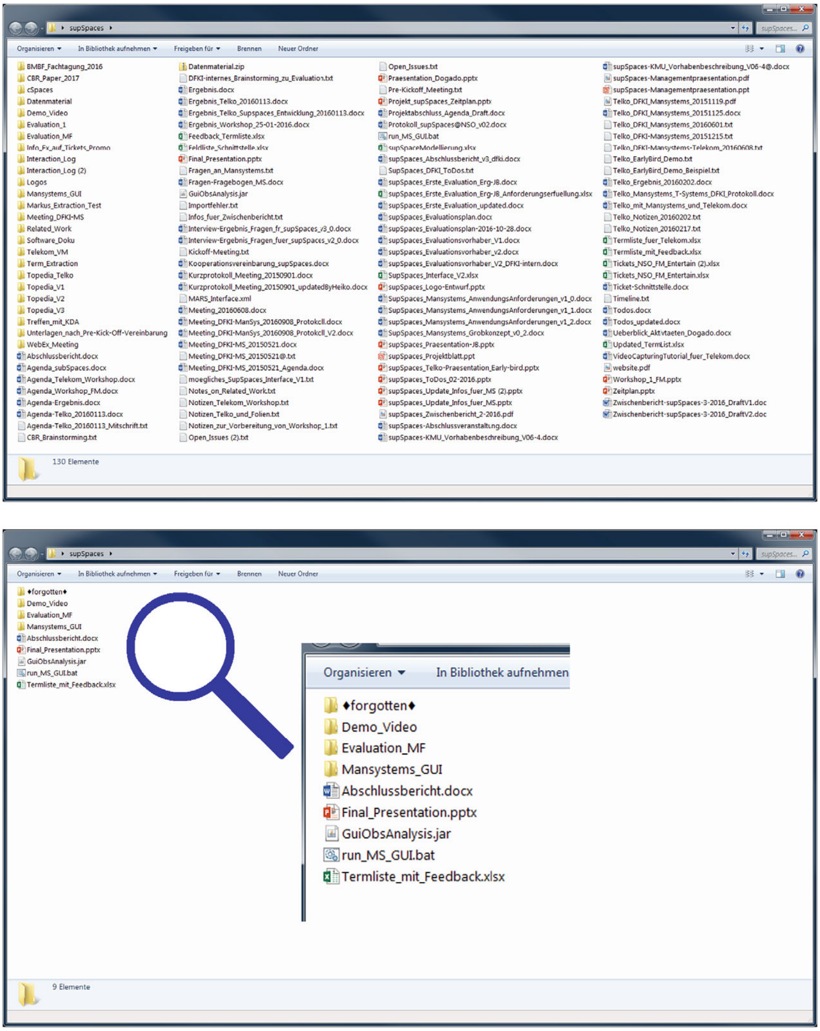


Fig. 3 Example of reorganization

user study involving ten participants, that could use CoMem to manage their photo collections, we could show that the system was capable of selecting meaningful photo collections to be preserved. The system selected between 55% and 83% of users' photos to be preserved (mean: 72%, median: 73%) and eight of ten participants agreed with 87% to 96% of its decisions (mean: 91%, median: 93%). For one participant the agreement was only 76% and a last one disagreed to such a large extent that using the feedback interface was omitted. For details of this study or more **general preservation methods in CoMem** as well as **cSpaces' automated context management**, an ongoing topic under heavy development, see Jilek [55].

Additionally, a reorganization approach applied to **Stack Overflow**[21] can be found in Nguyen et al. [78].

Due to their autonomous character and the aforementioned paradigm shift, self-organizing and especially forgetting-enabled PKA raise questions with regard to search and trust that are discussed in the following.

4.5 Searching Forgetting-Enabled Information Systems like cSpaces

In the beginning of this paper, we already mentioned that searching *forgetting-enabled information systems (FIS)* is more challenging than traditional search. For example, if no or seemingly incomplete search results are shown, users may start to wonder whether they entered the "right" keywords or whether they really saved the item they are now looking for. Being aware that the system is forgetting-enabled, they may also ask themselves whether the searched items have been forgotten by the system and are therefore not showing up. This may actually be correct because a user is maybe looking for something not used, mentioned or otherwise "stimulated" (e.g. by mentioning related topics) for a very long time and now wants to remember or go back to the item. We addressed these issues in several FIS search prototypes, of which two are shortly presented in the following. Our first prototype was developed as part of CoMem. With a slider users could decrease the Memory Buoyancy threshold leading to more and more "forgotten" items to show up. This led, however, to pathological behavior by users dragging the slider to "low MB" instead of rephrasing their query if search results were not satisfactory. Issues and doubts as described above were still present. We thus developed a second prototype as shown in Fig. 4. Besides the typical input field (A) and the list of results (E), there is a coverage indicator (B) showing how many found items are in the "active" part of the system and how many in the "forgotten part". Users thus get an idea of what actually is or was there without being additionally flooded by possibly rightfully forgotten items. In addition, contextual clustering is performed, showing topic clusters for active (C) and forgotten parts (D) of the user's information sphere allowing for further drill-down of the search. More details and evaluation results can be found in Jilek [55].

4.6 Trust in Highly Autonomous Assistants like cSpaces

In scenarios involving a highly autonomous PKA like cSpaces that reorganizes a person's information sphere, trust is particularly important. This comprises trust in search results (see last section) and the system's actions and not losing the feeling

[21] https://stackoverflow.com/.

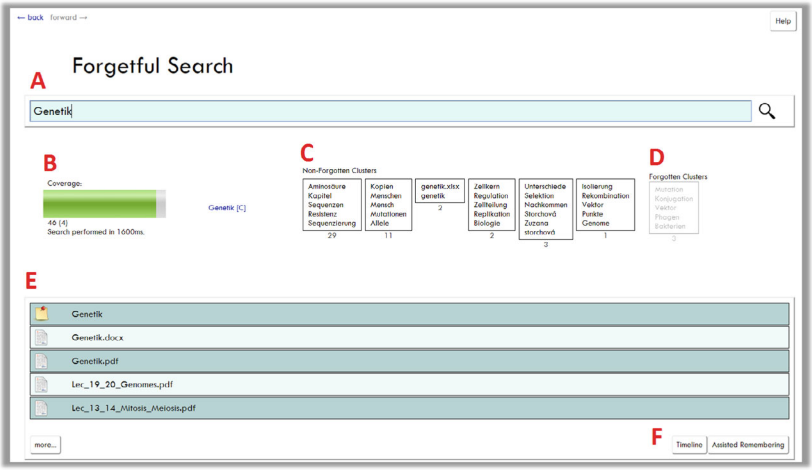


Fig. 4 FIS search prototype II: Coverage indication and contextual clustering

of control. Together with colleagues and students,[22] we conducted an online study, in which 140 participants assumed the role of a knowledge worker being supported by cSpaces in the form of a sidebar on the one side of the screen. In three different situations, participants were shown three different variants of status messages of cSpaces. These messages were phrased to correspond to levels of the *Situation Awareness-based Agent Transparency (SAT) model* by Chen et al. [24]. A first variant only stated *goals and actions* (e.g. a new context has been created by cSpaces). A second variant additionally explained *reasoning* behind the action (e.g. the last four actions strongly addressed a certain project and therefore the context has been created). A third variant also gave *projections* on the assistant's future actions (e.g. if additional evidences confirmed the context switch, the user's last three actions would also be assigned to the newly created context). Findings of the study were as follows:

- Trust and intention to use increased as a function of higher transparency.
- There was no moderation effect of propensity to trust.
- Transparency perceptions differed across situations.
- Participants formed a preference.

In conclusion, one can say that more explanations were not necessarily perceived as more transparency, and for different situations different levels were preferred. An assistant's explanations should thus be adaptable and their amount adjustable. For example, in the beginning, when working with such a system is still a new experience for users, more explanations are presumably helpful. After some time, when users

[22] A.-S. Ulfert-Blank, J. Knabe, and C. Jilek: *Forgetting-enabled AI systems: Exploring the role of transparency perceptions and trust* (paper in preparation).

are more accustomed to the system, they may prefer to regulate down their amount. For further details please see Jilek [55].

5 Conclusions

In this paper, we gave a retrospective overview of a decade of research in our department toward self-organizing personal knowledge assistants in evolving corporate memories. It was typically inspired by real-world problems and often conducted in interdisciplinary collaborations with research and industry partners. We summarized a selection of past experiments and results on *(Self-organizing) Context Spaces* combining the Semantic Desktop, Managed Forgetting and Context Spaces as a novel approach to PIM and KW support. Past results were complemented by an overview on related work and our latest findings not published so far, like a multi-month user study and another one on user's trust in such a system. The approach already allows for further increasing the automation in PIM and KW, whereas some contributions are only first steps in new directions with untapped potential left to explore. One of the very recent topics we are currently working on is the efficient combination of *Large Language Models* [100] and Knowledge Graphs in our PIM and KW support scenarios.

Acknowledgements The research reported in this article was funded by grants DE 420/19-2, FR 2133/11-1, NI 1760/1-2, TE 891/4-2 of the German Research Foundation (DFG), and by grant 01IW20007 by the German Federal Ministry of Education and Research.

References

1. Abecker. A., Bernardi, A., Hinkelmann, K., Kühn, O., & Sintek, M. (1998, June). Toward a technology for organizational memories. *IEEE Intelligent Systems*, *13*, 40–48.
2. Abela, C., & Staff, C. (2016). Behaviour mining for automatic task-keeping and visualisations for task-refinding. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval (CHIIR'16), Carrboro, North Carolina, USA* (pp. 23–32). ACM. https://doi.org/10.1145/2854946.2854966.
3. Abela, C., Staff, C., & Handschuh, S. (2015a). Collecting and analysing personal information management data. In J. Debattista, M. d'Aquin, & C. Lange (Eds.), *Proceedings of the First DIACHRON Workshop on Managing the Evolution and Preservation of the Data Web co-located with 12th European Semantic Web Conference (ESWC 2015)*, volume 1377 of *CEUR Workshop Proceedings*. CEUR-WS.org. http://ceur-ws.org/Vol-1377/paper3.pdf.
4. Abela, C., Staff, C., & Handschuh, S. (2015b). Automatic task-cluster generation based on document switching and revisitation. In A. Cristea, J. Masthoff, A. Said, & N. Tintarev (Eds.), *Posters, demos, late-breaking results and workshop proceedings of the 23rd conference on User Modeling, Adaptation, and Personalization (UMAP 2015), Dublin, Ireland, June 29–July 3, 2015*, volume 1388 of *CEUR Workshop Proceedings*. CEUR-WS.org. http://ceur-ws.org/Vol-1388/DeCat2015-paper1.pdf

5. Altmann, E. M., & Trafton, J. G. (2004). Task interruption: Resumption lag and the role of cues. In *Annual meeting of the cognitive science society* (Vol. 26).
6. Amrit, C., Wijnhoven, A. B. J. M., & Beckers, D. (2015). Information waste on the world wide web and combating the clutter. In *European Conference on Information Systems 2015 completed research papers*. https://doi.org/10.18151/7217266. https://research.utwente.nl/files/232243236/Amrit2015information.pdf
7. Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, *44*(1), 75–106. https://doi.org/10.1016/0010-0277(92)90051-I. https://www.researchgate.net/profile/Mark-Ashcraft/publication/222340465_Ashcraft_M_H_Cognitive_arithmetic_a_review_of_data_and_theory_Cognition_44_75-106/links/5a2aee1ca6fdccfbbf8520fd/Ashcraft-M-H-Cognitive-arithmetic-a-review-of-data-and-theory-Cognition-44-75-106.pdf
8. Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General, 130*(2), 224–237. https://doi.org/10.1037/0096-3445.130.2.224. https://www.apa.org/news/press/releases/xge1302224.pdf
9. Attard, J., & Brennan, R. (2018). A semantic data value vocabulary supporting data value assessment and measurement integration. In *20th International Conference on Enterprise Information Systems—Volume 2: ICEIS* (pp. 133–144). INSTICC, SciTePress. https://doi.org/10.5220/0006777701330144. https://doras.dcu.ie/22982/1/iceis_attardandBrennan.pdf
10. Attard, J., & Brennan, R. (2019). Dave: A semantic data value vocabulary to enable data value characterisation. In S. Hammoudi, M. Śmiałek, O. Camp, & J. Filipe (Eds.), *Enterprise information systems, 20th International Conference, ICEIS 2018, Funchal, Madeira, Portugal, March 21–24, 2018, Revised Selected Papers* (pp. 239–261). Springer. https://doi.org/10.1007/978-3-030-26169-6_12. https://www.researchgate.net/profile/Rob-Brennan/publication/334722723_DaVe_A_Semantic_Data_Value_Vocabulary_to_Enable_Data_Value_Characterisation/links/5d529f424585153040709a4a/DaVe-A-Semantic-Data-Value-Vocabulary-to-Enable-Data-Value-Characterisation.pdf
11. Balog,K., & Kenter, T. (2019). Personal knowledge graphs: A research agenda. In Y. Fang, Y. Zhang, J. Allan, K. Balog, B. Carterette, & J. Guo (Eds.), *2019 ACM SIGIR International Conference on Theory of Information Retrieval, ICTIR 2019, Santa Clara, CA, USA, October 2–5, 2019* (pp. 217–220). ACM. https://doi.org/10.1145/3341981.3344241
12. Beilock, S. L., & Carr, T. H. (2005). When high-powered people fail: Working memory and "choking under pressure" in math. *Psychological Science, 16*(2), 101–105. https://doi.org/10.1111/j.0956-7976.2005.00789.x. https://hpl.uchicago.edu/files/2018/07/Psychological-Science2005-1b1qwwq.pdf
13. Bergman, O., Beyth-Marom, R., & Nachmias, R. (2006). The project fragmentation problem in personal information management. In *SIGCHI Conference on Human Factors in Computing Systems (CHI '06)* (pp. 271–274). ACM. https://doi.org/10.1145/1124772.1124813. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.100.6405&rep=rep1&type=pdf
14. Bergman, O., Beyth-Marom, R., & Nachmias, R. (2008). The user-subjective approach to personal information management systems design: Evidence and implementations. *Journal of the American Society for Information Science and Technology, 59*(2), 235–246. https://doi.org/10.1002/asi.20738. https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.20738
15. Berners-Lee, T., Fielding, R., & Masinter, L. (1998, August). Uniform resource identifiers (URI): Generic syntax. RFC 2396. https://www.rfc-editor.org/rfc/rfc2396.html
16. Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web: A new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American, 284*(5), 34–43. http://web.archive.org/web/20040309192648/http://www.sciam.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21
17. Berners-Lee, T., Fielding, R., & Masinter, L. (2005, January). Uniform resource identifier (URI): Generic syntax. RFC 3986. https://www.rfc-editor.org/rfc/rfc3986.html
18. Bogomolny, A. (1996). Modular arithmetic. Cut the Knot—Interactive Mathematics Miscellany and Puzzles (Website). https://www.cut-the-knot.org/blue/Modulo.shtml

19. Bourtoule, L., Chandrasekaran, V., Choquette-Choo, C. A., Jia, H., Travers, A., Zhang, B., Lie, D., & Papernot, N. (2021). Machine unlearning. In *2021 IEEE Symposium on Security and Privacy (SP)* (pp. 141–159). IEEE. https://doi.org/10.1109/SP40001.2021.00019. https://www.researchgate.net/profile/David-Lie-4/publication/337855762_Machine_Unlearning/links/617993aea767a03c14be4702/Machine-Unlearning.pdf
20. Brennan, R., Attard, J., Petkov, P., Nagle, T., & Helfert, M. (2019). Exploring data value assessment: A survey method and investigation of the perceived relative importance of data value dimensions. In *21st International Conference on Enterprise Information Systems—Volume 1: ICEIS* (pp. 200–207). INSTICC, SciTePress. https://doi.org/10.5220/0007723402000207. https://pdfs.semanticscholar.org/e566/26c76b06d66758fb79c206c5ed733ecafbd8.pdf
21. Bush, V. (1945, March). As we may think. *Atlantic Monthly, 176*(1), 641–649. https://doi.org/10.1145/227181.227186. https://www.ias.ac.in/article/fulltext/reso/005/11/0094-0103
22. Cao, Y., & Yang, J. (2015). Towards making systems forget with machine unlearning. In *2015 IEEE symposium on security and privacy* (pp. 463–480). IEEE. https://doi.org/10.1109/SP.2015.35. https://www.ieee-security.org/TC/SP2015/papers-archived/6949a463.pdf
23. Carroll, J., & Klyne, G. (2004, February). Resource description framework (RDF): Concepts and abstract syntax. W3c recommendation, W3C. https://www.w3.org/TR/2004/REC-rdf-concepts-20040210/
24. Chen, J. Y. C., Procci, K., Boyce, M., Wright, J., Garcia, A., & Barnes, M. (2014, April). Situation awareness-based agent transparency. Technical Report ARL-TR-6905, U.S. Army Research Laboratory, Aberdeen Proving Ground, MD 21005-5425. https://apps.dtic.mil/sti/pdfs/ADA600351.pdf
25. Cheyer, A., Park, J., & Giuli, R. (2005). Iris: Integrate. relate. infer. share. In S. Decker, J. Park, D. Quan, & L. Sauermann (Eds.), *ISWC 2005 Workshop on The Semantic Desktop—Next Generation Information Management & Collaboration Infrastructure. Galway, Ireland, November 6, 2005*, volume 175 of *CEUR Workshop Proceedings* (pp. 59–73). CEUR-WS.org. http://ceur-ws.org/Vol-175/17_park_iris_final.pdf
26. Crispin, M. (1988, July). Interactive mail access protocol: Version 2. RFC 1064. https://www.rfc-editor.org/rfc/rfc1064.html
27. Cyganiak, R., Wood, D., & Lanthaler, M. (2014, February). RDF 1.1 concepts and abstract syntax. W3c recommendation, W3C. https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/
28. Daboo, C. (2011, August) . Carddav: vcard extensions to web distributed authoring and versioning (webdav). RFC 6352. https://www.rfc-editor.org/rfc/rfc6352.html
29. Daboo, C., Desruisseaux, B., & Dusseault, L. (2007, March). Calendaring extensions to webdav (caldav). RFC 4791. https://www.rfc-editor.org/rfc/rfc4791.html
30. Davenport, T. H. (2005). *Thinking for a living: How to get better performances and results from knowledge workers*. Harvard Business School Press. ISBN 1-59139-423-6.
31. Decker, S., & Frank, M. (2004, May). The social semantic desktop. DERI Technical Report 2004-05-02, Digital Enterprise Research Institute (DERI).
32. Dengel, A., Abecker, A., Bernardi, A., van Elst, L., Maus, H., Schwarz, S., & Sintek, M. (2002). Konzepte zur gestaltung von unternehmensgedächtnissen [design concepts for organizational memories]. *KI—Künstliche Intelligenz, German Journal of Artificial Intelligence, 16*(1), 5–11. https://www.dfki.uni-kl.de/~elst/papers/KI-Heft_OM-Prinzipien-draft.pdf
33. Dong, X., Halevy, A., Nemes, E., Sigurdsson, S., & Domingos, P. (2004). Semex: Toward on-the-fly personal information integration. In *Workshop on Information Integration on the Web (IIWEB)*. https://pal.sri.com/wp-content/uploads/publications/calo/2004/semex_iiweb.pdf
34. Dragan, L., & Decker, S. (2012). Knowledge management on the desktop. In *Knowledge engineering and knowledge management* (pp. 373–382). Springer. https://doi.org/10.1007/978-3-642-33876-2_33
35. Dragunov, A. N., Dietterich, T. G., Johnsrude, K., McLaughlin, M., Li, L., & Herlocker, J. L. (2005). Tasktracer—A desktop environment to support multi-tasking knowledge workers. In *10th International Conference on Intelligent User Interfaces (IUI '05), San Diego, California, USA* (pp. 75–82). ACM. https://doi.org/10.1145/1040830.1040855. https://web.engr.oregonstate.edu/~tgd/publications/iui2005-tasktracer.pdf

36. Ferrari, M., & Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fnirs) development and fields of application. *NeuroImage, 63*(2), 921–935, . https://doi.org/10.1016/j.neuroimage.2012.03.049. https://iss.com/resources/pdf/publications/fnirs-history.pdf
37. Franz, T., Scherp, A., & Staab, S. (2009). Are semantic desktops better?: Summative evaluation comparing a semantic against a conventional desktop. In *5th International Conference on Knowledge Capture*, K-CAP '09 (pp. 1–8). ACM. ISBN 978-1-60558-658-8.
38. Galkin, M., Auer, S., Vidal, M.-E., & Scerri, S. (2017). Enterprise knowledge graphs: A semantic approach for knowledge management in the next generation of enterprise information systems. In S. Hammoudi, M. Smialek, O. Camp, & J. Filipe (Eds.), *ICEIS 2017—Proceedings of the 19th International Conference on Enterprise Information Systems, Volume 2, Porto, Portugal, April 26–29, 2017* (pp. 88–98). SciTePress. https://doi.org/10.5220/0006325200880098
39. Gauselmann, P., Runge, Y., Jilek, C., Frings, C., Maus, H., & Tempel, T. (2022). A relief from mental overload in a digitalized world: How context-sensitive user interfaces can enhance cognitive performance. *International Journal of Human-Computer Interaction*. https://doi.org/10.1080/10447318.2022.2041882. In press.
40. Gauß, C. F. (1966). *Disquisitiones arithmeticae*. 1801. https://doi.org/10.1007/978-1-4939-7560-0. Translated into English by A. A. Clark, Yale University Press.
41. Geißler, C., Gauselmann, P., Jilek, C., Maus, H., Frings, C., & Tempel, T. (2023). A functional near-infrared spectroscopy study on the prefrontal correlates of cognitive offloading via a personal knowledge assistant. *Scientific Reports, 13* (13938).
42. Gemmell, J., Bell, G., Lueder, R., Drucker, S., & Wong, C. (2002). Mylifebits: Fulfilling the memex vision. In *10th ACM International Conference on Multimedia, MULTIMEDIA '02, Juan-les-Pins, France* (pp. 235–238). ACM. https://doi.org/10.1145/641007.641053. https://www.microsoft.com/en-us/research/wp-content/uploads/2002/12/mylifebitsmm02.pdf
43. Gifford, D. K., Jouvelot, P., Sheldon, M. A., & O'Toole Jr, J. W. (1991). Semantic file systems. *ACM SIGOPS Operating Systems Review, 25*: 16–25. https://doi.org/10.1145/121133.121138. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.147.2488&rep=rep1&type=pdf
44. Goland, Y., Whitehead, E., Faizi, A., Carter, S., & Jensen, D. (1999, February). Http extensions for distributed authoring—webdav. RFC 2518. https://www.rfc-editor.org/rfc/rfc2518.html
45. Gomez-Perez, J. M., Grobelnik, M., Ruiz, C., Tilly, M., & Warren, P. (2009). Using task context to achieve effective information delivery. In *1st Workshop on Context, Information and Ontologies (CIAO '09)* (pp. 3:1–3:6). ACM, . https://doi.org/10.1145/1552262.1552265. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.494.600&rep=rep1&type=pdf
46. Gupta, V., Jung, C., Neel, S., Roth, A., Sharifi-Malvajerdi, S., & Waites, C. (2021). Adaptive machine unlearning. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, & J. Wortman Vaughan (Eds.), *Advances in neural information processing systems* (Vol. 34, pp. 16319–16330). Curran Associates. https://proceedings.neurips.cc/paper/2021/file/87f7ee4fdb57bdfd52179947211b7ebb-Paper.pdf
47. Gurrin, C., Lee, H., & Hayes, J. (2010). iforgot: A model of forgetting in robotic memories. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 93–94). IEEE. https://doi.org/10.1109/HRI.2010.5453255
48. Gurrin, C., Smeaton, A. F., & Doherty, A. R. (2014, June). Lifelogging: Personal big data. *Foundations and Trends in Information Retrieval, 8*(1), 1–125. https://doi.org/10.1561/1500000033
49. Gyllstrom,K., & Pedersen, E. R. (2010). Not gone, but forgotten: Helping users re-find web pages by identifying those which are most likely to be lost. *Proceedings of the Workshop on Desktop Search at SIGIR*.
50. Gyllstrom, K., & Stotts, D. (2006). *Unsupervised task extraction from a stream of window focus events* (Technical Report TR06-023). University of North Carolina. http://www.cs.unc.edu/techreports/06-023.pdf
51. Gyllstrom, K., Soules, C., & Veitch, A. (2008). Activity put in context: Identifying implicit task context within the user's document interaction. In *2nd International Symposium on Information*

Interaction in Context (IIiX '08), London, United Kingdom (pp. 51–56). ACM. https://doi.org/10.1145/1414694.1414707. http://www.cs.unc.edu/techreports/08-007.pdf

52. Hasan, R., & Burns, R. (2013). The life and death of unwanted bits: Towards proactive waste data management in digital ecosystems. In *Third International Conference on Innovative Computing Technology (INTECH 2013)* (pp. 144–148). https://doi.org/10.1109/INTECH.2013.6653665. https://arxiv.org/pdf/1106.6062v2.pdf
53. Heylighen, F., & Gershenson, C. (2003). The meaning of self-organization in computing. *IEEE Intelligent Systems, 18*(4).
54. Horvitz,E., Breese, J., Heckerman, D., Hovel, D., & Rommelse, K. (1998). The lumière project: Bayesian user modeling for inferring the goals and needs of software users. *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, Madison WI, July, 1998* (pp. 256–265). https://www.microsoft.com/en-us/research/wp-content/uploads/2016/11/lum.pdf
55. Jilek, C. (2024). *Self-organizing context spaces to support information management and knowledge work* (PhD thesis). RPTU Kaiserslautern-Landau.
56. Jilek, C., Maus, H., Schwarz, S., & Dengel, A. (2015). Diary generation from personal information models to support contextual remembering and reminiscence. In *Workshop on Human Memory-Inspired Multimedia Organization and Preservation (HMMP 2015), Proceedings of the 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW 2015), Turin, Italy, June 29–July 3, 2015* (pp. 1–6). IEEE. https://doi.org/10.1109/ICMEW.2015.7169753
57. Jilek, C., Schröder, M., Schwarz, S., Maus, H., & Dengel, A. (2018). Context spaces as the cornerstone of a near-transparent and self-reorganizing semantic desktop. In A. Gangemi, A. L. Gentile, A. G. Nuzzolese, S. Rudolph, M. Maleshkova, H. Paulheim, J. Z. Pan, & M. Alam (Eds.), *The Semantic Web: ESWC 2018 Satellite Events—ESWC 2018 Satellite Events, Heraklion, Crete, Greece, June 3–7, 2018, Revised Selected Papers* (pp. 89–94). Springer. https://doi.org/10.1007/978-3-319-98192-5_17. https://arxiv.org/pdf/1805.02181.pdf
58. Jilek, C., Chwalek, J., Schwarz, S., Schröder, M., Maus, H., & Dengel, A. (2019, May). Advanced memory buoyancy for forgetful information systems. *AIS Transactions on Enterprise Systems* 4.1. https://doi.org/10.30844/aistes.v4i1.11. https://www.aes-journal.com/index.php/ais-tes/article/view/11/11
59. Jilek, C., Runge, Y., Niederée, C., Maus, H., Tempel, T., Dengel, A., & Frings, C. (2019b, March). Managed forgetting to support information management and knowledge work. *KI—Künstliche Intelligenz, German Journal of Artificial Intelligence, 33*(1), 45–55. https://doi.org/10.1007/s13218-018-00568-9. https://arxiv.org/pdf/1811.12155.pdf
60. Jilek, C., Schröder, M., Novik, R., Schwarz, S., Maus, H., & Dengel, A. (2019c). Inflection-tolerant ontology-based named entity recognition for real-time applications. In M. Eskevich, G. de Melo, C. Fäth, J. P. McCrae, P. Buitelaar, C. Chiarcos, B. Klimek, & M. Dojchinovski (Eds.), *2nd Conference on Language, Data and Knowledge (LDK 2019), Leipzig, Germany, May 20–23, 2019*, volume 70 of *OpenAccess Series in Informatics (OASIcs)* (pp. 11:1–11:14). Schloss Dagstuhl. https://doi.org/10.4230/OASIcs.LDK.2019.11. https://drops.dagstuhl.de/opus/volltexte/2019/10375/pdf/OASIcs-LDK-2019-11.pdf
61. Jilek, C., Gauselmann, P., Chwalek, J., Tempel, T., & Dengel, A. (2021). Quantifying benefits of a personal knowledge assistant on task resumption. In *Arbeit HumAIne Gestalten—67. Frühjahrskongress der Gesellschaft für Arbeitswissenschaft e.V. (GfA 2021), Ruhr-Universität Bochum, Online, March 3–5, 2021*. GfA-Press. ISBN 978-3-936804-29-4.
62. Jilek, C., Schröder, M., Maus, H., Schwarz, S., & Dengel, A. (2023). Towards self-organizing personal knowledge assistants in evolving corporate memories. *Computing Research Repository (CoRR)*. https://arxiv.org/pdf/2308.01732.pdf
63. Jones, W. (2008). *Keeping found things found: The study and practice of personal information management*. Morgan Kaufmann. ISBN 978-0-12-370866-3. https://doi.org/10.1016/B978-0-12-370866-3.X5001-2
64. Kanhabua, N., Niederée, C., & Siberski, W. (2013). Towards concise preservation by managed forgetting: Research issues and case study. In *10th International Conference on Preservation of Digital Objects (iPres)*.

65. Klimt, B., & Yang, Y. (2004). Introducing the enron corpus. In *CEAS 2004—First Conference on Email and Anti-Spam, July 30–31, 2004, Mountain View, California, USA*. https://ceas.cc/papers-2004/168.pdf
66. Kraaij, W., Verberne, S., Koldijk, S., de Korte, E., van Dantzig, S., Sappelli, M., Shoaib, M., Bosems, S., Achterkamp, R., Bonomi, A., Schavemaker, J., Hulsebosch, B., Wabeke, T., Vollenbroek-Hutten, M., Neerincx, M., & van Sinderen, M. (2020). Personalized support for well-being at work: An overview of the swell project. *User Modeling and User-Adapted Interaction, 30*(3), 413–446, . https://doi.org/10.1007/s11257-019-09238-3. https://link.springer.com/content/pdf/10.1007/s11257-019-09238-3.pdf
67. León, F. R., Lundgren, R., Huapaya, A., Sinai, I., & Jennings, V. (2007). Challenging the courtesy bias interpretation of favorable clients' perceptions of family planning delivery. *Evaluation Review, 31*(1), 24–42. https://doi.org/10.1177/0193841X06289044. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.845.3731&rep=rep1&type=pdf
68. Lindstaedt, S., & Mayer, H. (2006). A storyboard of the aposdle vision. In W. Nejdl, & K. Tochtermann (Eds.), *Innovative approaches for learning and knowledge sharing* (pp. 628–633). Springer Berlin Heidelberg. https://doi.org/10.1007/11876663_64. https://www.know-center.at/filemaker/pdf_publikationen/A
69. Maus, H. (2001). Workflow context as a means for intelligent information support. In *Modeling and Using Context: 3rd International and Interdisciplinary Conference (CONTEXT 2001)*, volume 2116 of *LNAI* (pp. 261–264). Springer. https://doi.org/10.1007/3-540-44607-9_20. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.23.9363&rep=rep1&type=pdf
70. Maus, H., Schwarz, S., & Dengel, A. (2013). Weaving personal knowledge spaces into office applications. In M. Fathi (Ed.), *Integration of practice-oriented knowledge technology: Trends and prospectives* (pp. 71–82). Springer. https://doi.org/10.1007/978-3-642-34471-8_6. https://pimo.opendfki.de/wp9-pilot/intro/MausSchwarzDengel12_submitted.pdf
71. Maus, H., Jilek, C., & Schwarz, S. (2018). Remembering and forgetting for personal preservation. In V. Mezaris, C. Niederée, & R. H. Logie, editors, *Personal multimedia preservation: Remembering or forgetting images and video* (pp. 233–277). Springer. https://doi.org/10.1007/978-3-319-73465-1_7
72. Mesnier, M., Thereska, E., Ganger, G. R., Ellard, D., & Seltzer, M. (2004). File classification in self-* storage systems. In *Proceedings of the International Conference on Autonomic Computing, New York, NY, USA, May 17–18, 2004* (pp. 44–51). IEEE. https://doi.org/10.1109/ICAC.2004.1301346. https://www.pdl.cmu.edu/PDL-FTP/ABLE/CMU-PDL-04-101.pdf
73. Microsoft Corporation. (2021, June). *Server message block (SMB) protocol versions 2 and 3. Microsoft Docs. Windows Protocols. Open Specifications* (Technical report) MS-SMB2 v20210625, Microsoft Corporation. https://docs.microsoft.com/openspecs/windows_protocols/ms-smb2
74. Mozilla Project. (2015). *Webextensions*. Mozilla Wiki. https://wiki.mozilla.org/WebExtensions
75. Nguyen, T. H., & Rudra, K. (2022a). Towards an interpretable approach to classify and summarize crisis events from microblogs. In *Proceedings of the ACM Web Conference 2022*, WWW '22 (pp. 3641–3650). Association for Computing Machinery. ISBN 9781450390965.
76. Nguyen, T. H., & Rudra, K. (2022b). Rationale aware contrastive learning based approach to classify and summarize crisis-related microblogs. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, CIKM '22 (pp. 1552–1562). Association for Computing Machinery. https://doi.org/10.1145/3511808.3557426
77. Nguyen, T. H., Hoang, T.-A., & Nejdl, W. (2019). *Efficient summarizing of evolving events from Twitter streams* (pp. 226–234). https://doi.org/10.1137/1.9781611975673.26
78. Nguyen, T. H., Nguyen, T., Hoang, T.-A., & Niederée, C. (2021). *On the feasibility of predicting questions being forgotten in stack overflow*. https://arxiv.org/abs/2110.15789
79. Nguyen, T. H., Shaltev, M., & Rudra, K. (2022). Crisicsum: Interpretable classification and summarization platform for crisis events from microblogs. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, CIKM '22 (pp. 4941–4945). Association for Computing Machinery. https://doi.org/10.1145/3511808.3557191

80. Niederée, C., Kanhabua, N., Gallo, F., & Logie, R. H. (2015). Forgetful digital memory: Towards brain-inspired long-term data and information management. *SIGMOD Record, 44*(2), 41–46. https://doi.org/10.1145/2814710.2814718. https://sigmodrecord.org/publications/sigmodRecord/1506/09_vision_Niederee.pdf
81. North, K., Brandner, A., & Steininger, T. (2016). *Die Wissenstreppe: Information—Wissen—Kompetenz [The knowledge ladder: information—Knowledge—Expertise]* (pp. 5–8). Springer Fachmedien Wiesbaden. ISBN 978-3-658-11250-9. https://doi.org/10.1007/978-3-658-11250-9_2
82. Paulheim, H. (2017). Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web, 8*(3), 489–508. https://doi.org/10.3233/SW-160218
83. Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002. http://samgilbert.net/pubs/Risko2016TiCS.pdf
84. Sappelli, M., Verberne, S., & Kraaij, W. (2013, November). Combining textual and non-textual features for e-mail importance estimation. In K. Hindriks (Ed.), *Proceedings of the 25th Benelux Conference on Artificial Intelligence (BNAIC 2013), Delft, November 7–8, 2013* (pp. 168–174). https://repository.ubn.ru.nl/bitstream/handle/2066/122950/122950.pdf
85. Sauermann, L. (2009). *The Gnowsis semantic desktop approach to personal information management* (PhD thesis). TU Kaiserslautern. https://www.dfki.de/fileadmin/user_upload/import/4261_Sauermann2009phd.pdf
86. Sauermann, L., Bernardi, A., & Dengel, A. (2005). Overview and outlook on the semantic desktop. In S. Decker, J. Park, D. Quan, & L. Sauermann (Eds.), *ISWC 2005 Workshop on The Semantic Desktop—Next Generation Information Management & Collaboration Infrastructure. Galway, Ireland, November 6, 2005*, volume 175 of *CEUR Workshop Proceedings* (pp. 74–91). CEUR-WS.org. http://ceur-ws.org/Vol-175/18_sauermann_overviewsemdesk_final.pdf
87. Sauermann, L., van Elst, L., & Dengel, A. (2007). PIMO—A framework for representing personal information models. In *Proceedings of I-Media '07 and I-Semantics '07* (pp. 270–277). Know-Center, Austria.
88. Scerri, S., Attard, J., Rivera, I., Valla, M., & Handschuh, S. (2012). DCON: Interoperable context representation for pervasive environments. In *AAAI Workshops, Technical Report WS-12-05* (pp. 90–97).
89. Schmidt, A. (2006). A layered model for user context management with controlled aging and imperfection handling. In T. R. Roth-Berghofer, S. Schulz, & D. B. Leake (Eds.), *Modeling and Retrieval of Context, Second International Workshop, MRC 2005, Edinburgh, UK, July 31-August 1, 2005, Revised Selected Papers* (pp. 86–100). Springer Berlin Heidelberg. https://doi.org/10.1007/11740674_6. https://www.researchgate.net/profile/Andreas-Schmidt-11/publication/220743719_A_Layered_Model_for_User_Context_Management_with_Controlled_Aging_and_Imperfection_Handling/links/0912f50c7aa1d84d5e000000/A-Layered-Model-for-User-Context-Management-with-Controlled-Aging-and-Imperfection-Handling.pdf
90. Schröder, M. (2022). *Building knowledge graphs from messy enterprise data* (PhD thesis). TU Kaiserslautern.
91. Schwarz, S. (2005). A context model for personal knowledge management. In T. R. Roth-Berghofer, S. Schulz, & D. B. Leake (Eds.), *2nd International Workshop on Modeling and Retrieval of Context (MRC '05) in conjuction with IJCAI-05*, volume 146 of *CEUR Workshop Proceedings*. CEUR-WS.org. http://ceur-ws.org/Vol-146/paper4.pdf
92. Schwarz, S. (2006). A context model for personal knowledge management applications. In T. R. Roth-Berghofer, S. Schulz, & D. B. Leake (Eds.), *Modeling and Retrieval of Context—Second International Workshop, MRC 2005, Edinburgh, UK, July 31–August 1, 2005, Revised Selected Papers*, Lecture Notes in Artificial Intelligence (pp. 18–33). Springer Berlin Heidelberg. https://doi.org/10.1007/11740674_2
93. Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data & Knowledge Engineering, 25*(1), 161–197. https://doi.org/

10.1016/S0169-023X(97)00056-6. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.97.3940&rep=rep1&type=pdf
94. Tempel, T., & Frings, C. (2016). Directed forgetting benefits motor sequence encoding. *Memory & Cognition, 44*(3), 413–419. https://doi.org/10.3758/s13421-015-0565-8. https://link.springer.com/content/pdf/10.3758/s13421-015-0565-8.pdf
95. Tran, T. A., Schwarz, S., Niederée, C., Maus, H., & Kanhabua, N. (2016). The forgotten needle in my collections: Task-aware ranking of documents in semantic information space. In *2016 ACM on Conference on Human Information Interaction and Retrieval (CHIIR '16), Carrboro, North Carolina, USA* (pp. 13–22). ACM. https://doi.org/10.1145/2854946.2854971. https://homes.cs.aau.dk/~nattiya/papers/CHIIR2016-task-aware-ranking.pdf
96. Turczyk, L. A. (2009). *Information lifecycle management—Eine Methode zur Wertzuweisung von Dateien [A method for valuation of files]* (PhD thesis). TU Darmstadt.
97. Warren, P. (2013, July). Personal information management: The case for an evolutionary approach. *Interacting with Computers, 26*(3), 208–237. https://doi.org/10.1093/iwc/iwt034. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.975.4643&rep=rep1&type=pdf
98. Warren, P., Thurlow, I., Davies, J., Dolinsek, I., Ruiz, C., & Gomez-Perez, J. M. (2010). Context as a tool for organizing and sharing knowledge. In V. Ermolayev, J. M. Gomez-Perez, P. Haase, & P. Warren (Eds.), *Second Workshop on Context, Information and Ontologies (CIAO-2010), Lisbon, Portugal, October 11, 2010*, volume 626 of *CEUR Workshop Proceedings*. CEUR-WS.org. http://ceur-ws.org/Vol-626/regular1.pdf
99. Wijnhoven, F., Amrit, C., & Dietz, P. (2014). Value-based file retention: File attributes as file value and information waste indicators. *Journal of Data and Information Quality, 4*(4), 15:1–15:17. https://doi.org/10.1145/2567656. https://www.researchgate.net/profile/Chintan-Amrit/publication/274483776_Value-Based_File_Retention/links/5851183c08aecb6bd8d21d2e/Value-Based-File-Retention.pdf
100. Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Liu, P., Nie, J.-Y., & Wen, J.-R. (2024). *A survey of large language models*. https://arxiv.org/abs/2303.18223

Intentional Forgetting Operators—Formal Foundations and Empirical Support

Christoph Beierle, Daniel Brand, Diana Howey, Gabriele Kern-Isberner, Kai Sauerwald, Sara Todorovikj, and Marco Ragni

1 Introduction

In modern times, the amount of stored data and knowledge structures grows at exponential rates. Particularly in the time of Big Data and complex knowledge structures, the ability to focus on relevant information, while disregarding irrelevant details, is essential. The human mind excels in this regard and allows for flexible adaptation and application of new information. Even though the mind is not without its limitations—sometimes leading to forgetting relevant information or retaining irrelevant ones—it offers forgetting mechanisms allowing individuals to efficiently process and retrieve the most relevant information, which can be further used for reasoning, decision making or problem solving. One approach describing these processes is activation-based functions [3], which propose that each memory unit, often referred to as chunks, has an activation level. If the activation is over a given threshold, the information can be retrieved, otherwise, it remains inaccessible. The cognitive architecture ACT-R [3]

C. Beierle (✉) · K. Sauerwald
Faculty of Mathematics and Computer Science, FernUniversität in Hagen, Hagen, Germany
e-mail: christoph.beierle@fernuni-hagen.de

K. Sauerwald
e-mail: kai.sauerwald@fernuni-hagen.de

D. Brand · S. Todorovikj · M. Ragni
Predictive Analytics, TU Chemnitz, Chemnitz, Germany
e-mail: daniel.brand@metech.tu-chemnitz.de

S. Todorovikj
e-mail: sara.todorovikj@metech.tu-chemnitz.de

M. Ragni
e-mail: marco.ragni@hsw.tu-chemnitz.de

D. Howey · G. Kern-Isberner
Department of Computer Science, TU Dortmund University, Dortmund, Germany
e-mail: gabriele.kern-isberner@tu-dortmund.de

J. O. Berndt et al. (eds.), *Intentional Forgetting with Intelligent Systems*, Progress in IS,
https://doi.org/10.1007/978-3-032-17621-9_7

models this basic memory and forgetting mechanism as part of a broader theory of human cognition. However, the application to information systems that account for relevance and identify factors has not yet been comprehensively investigated. The aim of our research is to develop a psychologically grounded framework for understanding intentional forgetting processes in humans, providing a formal modeling structure and deriving best implementation practices.

The field of intentional forgetting is characterized by its multidisciplinary nature. On the theoretical level, the focus is on the formal characterization and categorization of forgetting [19]. The methods originate from knowledge representation, a subfield of artificial intelligence, as well as cognitive modeling. Hence, we will review intentional forgetting from the perspectives of cognition and formal logics. By refining and extending previous work [5, 8], we explore the logical foundations and theoretical aspects that underpin intentional forgetting from a commonsense perspective of knowledge representation and reasoning. We identify and examine various forgetting operators and their interrelationships, and characterize them axiomatically in a general framework. By illustrating how these forgetting operators can be modeled using, e.g., ranking functions [43], this chapter demonstrates a structured approach for understanding and implementing forgetting processes.

On a cognitive level, we consider further the distinction of declarative and procedural memory, well established in psychology and neuroscience (see, e.g., [44]). The unique characteristics of these memory types entail differences in forgetting processes. Building on the theoretical foundations of the operators, we investigate the cognitive mechanisms underlying intentional forgetting through new experimental paradigms: First, as a paradigm testing the transition between declarative and procedural memory, we introduced the *alien task*, where individuals needed to explicitly revise rules for a classification task. Thereby, the rules themselves could be considered declarative information, but also had a procedural training component due to their repeated application. Second, the *counting game* designed specifically to investigate how individuals apply and revise structured rules under conditions that require forgetting. This experiment is more demanding on a procedural level and aligns closely to the operations introduced within the formal framework. Both formal and cognitive investigations contribute to a deeper understanding of how to describe and reverse-engineer intentional forgetting (IF) in humans—to make it applicable in systems and organizations.

2 A Computer Science Perspective on Forgetting

In the following, we will present and discuss several types of change which involve forgetting. These types of forgetting have been inspired by (intentional) forgetting in commonsense reasoning where, e.g., we have to forget about a certain product that we want to buy but that is no longer available (contraction), or we want to replace our intention to buy this product by the wish to buy another product that seems to be more suitable for our needs (revision), or we forget about irrelevant details when solving a problem (abstraction). Obviously, such different kinds of forgetting are based on different cognitive processes, and our aim here is to characterize these types of forgetting by their outcomes, and to present adequate formalizations to reach these outcomes. Such formalizations also yield a solid basis for realizing and understanding intentional forgetting in AI systems [9, 46].

To provide a uniform logic-based framework for formalization, we use an abstract model in which an agent is equipped with an epistemic state Ψ (also called *belief state* in this paper) and an inference relation $\mid\!\approx$ between epistemic states and formulas. We make no further assumptions about how this belief state is represented, except that Ψ makes use of a logical language $\mathcal{L}_\Sigma$ over a signature Σ. Thus, the notion of belief state should be understood in a very broad sense. For instance, Ψ might be a set of logical formulas, a Bayesian network or a total preorder on possible worlds. The relation $\Psi \mid\!\approx \varphi$ holds if an agent with belief state Ψ infers φ, where φ is a formula which can be evaluated by the semantics of Ψ, e.g., a propositional formula, or a conditional. Depending on Ψ, the relation $\mid\!\approx$ can be a deductive inference relation, a non-monotonic inference relation based on conditionals, a probabilistic inference relation, etc. In the following, Ψ will denote the prior state of the agent and Ψ° the posterior state after the forgetting resp. change operation. In addition to the abstract framework, we will also develop a refinement of it employing ordinal conditional functions (OCF) [43], more intuitively called ranking functions. OCFs are capable of handling also conditional beliefs in a qualitative way. Besides having a concrete language for expressing beliefs and their changes at hand, choosing OCFs also illustrates that our abstract framework can properly catch logics that are much more powerful than, e.g., propositional logic and that it also allows for change operations with respect to a conditional belief. Most importantly in the context of modeling human reasoning, OCFs provide a cognitively adequate logical framework that avoids fallacies and paradoxes having been observed when classical logic is taken as the norm and that allows for defining rationality of reasoning on a clear conditional-logical base [18, 40].

As a logical base for our general epistemic states or OCFs, respectively, we use a finitely generated propositional language $\mathcal{L}_\Sigma$ with atoms $\Sigma = \{a, b, c, \ldots\}$, and with formulas $A, B, C, \ldots$ For conciseness of notation, we will omit the logical *and*-connector, writing AB instead of $A \wedge B$, and overlining formulas indicates negation, i.e., $\overline{A}$ means $\neg A$. For an atom a, let $\dot{a}$ denote any of its literals $a, \overline{a}$. Let Ω_Σ denote the set of possible worlds over Σ; Ω_Σ will be taken here simply as the set of all propositional interpretations over Σ. As usual, $\omega \models A$ means that the propositional formula $A \in \mathcal{L}$ holds in the possible world $\omega \in \Omega$, and $Mod(A) = \{\omega \mid \omega \models A\}$ denotes the set of all such possible worlds. The symbol $\top$ represents any tautology. We will use ω for both the model and the corresponding complete conjunction containing all atoms suitably either in positive or in negative form. By introducing a new binary operator $|$, we obtain the set $(\mathcal{L} \mid \mathcal{L}) = \{(B|A) \mid A, B \in \mathcal{L}\}$ of conditionals over $\mathcal{L}$. $(B|A)$ formalizes "*if A* "*then usually B*" and establishes a plausible connection between the *antecedent A* and the *consequent B*. Conditionals with tautological antecedents are taken as plausible statements about the world. Following De Finetti [20], a conditional $(B|A)$ is *verified* (*falsified*) by a world ω iff $\omega \models AB$ ($\omega \models A\overline{B}$).

If $\omega \not\models A$, the conditional is *not applicable* to ω. The negation of $(B|A)$ is defined as its counter conditional $(\overline{B}|A)$. A finite set of conditionals is called a *belief base*.

Because conditionals go well beyond classical logic, they require a richer setting for their semantics than classical logic. *Ordinal conditional functions (OCFs)*, (also called *ranking functions*) $\kappa : \Omega \to \mathbb{N} \cup \{\infty\}$ with $\kappa^{-1}(0) \neq \emptyset$, were introduced (in a more general form) first by Spohn [43]. They express degrees of plausibility of propositional formulas A by specifying degrees of disbelief of their negations $\overline{A}$. Thus, the more plausible a world ω is, the smaller is its rank $\kappa(\omega)$. An epistemic state can be respresented by an OCF κ. More formally, $\kappa(A) := \min\{\kappa(\omega) \mid \omega \models A\}$, so that $\kappa(A \vee B) = \min\{\kappa(A), \kappa(B)\}$. Hence, due to $\kappa^{-1}(0) \neq \emptyset$, at least one of $\kappa(A), \kappa(\overline{A})$ must be 0. To avoid technical peculiarities, we only consider OCFs $\kappa : \Omega \to \mathbb{N}$ within a finite range here; extending our approach to the general case is straightforward. A conditional $(B|A)$ is accepted in the epistemic state represented by κ, written as

$$\kappa \models (B|A), \text{ iff } \kappa(AB) < \kappa(A\overline{B}), \tag{1}$$

i.e., iff the verification AB of $(B|A)$ is more plausible than its falsification $A\overline{B}$. This is well in line with commonsense reasoning—people often believe (or accept) a conditional "if A then B" if the most plausible examples of $A \wedge B$ are more plausible than the most plausible examples of $A \wedge \overline{B}$. For a propositional formula A, we have $\kappa \models A$ iff $\kappa \models (A|\top)$ iff $\kappa(A) < \kappa(\overline{A})$ iff $\kappa(\overline{A}) > 0$, since at least one of $\kappa(A), \kappa(\overline{A})$ must be 0. Hence, κ instantiates Ψ in the abstract framework. Instead of abstract beliefs A we consider conditionals $(B|A)$, also including propositional beliefs via $A \equiv (A|\top)$, and $\approx\!\!|$ is instantiated to $\models$ as given in Eq. (1).

Example 1 Let $\Sigma = \{p, b, f\}$ with p meaning "penguin", b "bird" and f "able to fly". "Birds normally fly" is modeled with the conditional $r_1 = (f|b)$, "penguins normally do not fly" with $r_2 = (\overline{f}|p)$, and "penguins are normally birds" with $r_3 = (b|p)$. Consider the OCF κ_{pen} from Table1 , which will act as our running example for the following sections (where we will also elaborate the other OCFs shown in Table 1). The ranking function κ_{pen} accepts the belief base $\mathcal{R}_{pen} = \{r_1, r_2, r_3\}$, i.e., $\kappa_{pen} \models r_i$ for all $1 \leq i \leq 3$. For example, $\kappa \models r_1$ because $\kappa(bf) = 0 < 1 = \kappa(b\overline{f})$.

Note that, although it appears to be simplistic at first sight, this example models an important case of uncertain commonsense reasoning. Birds typically fly but there are exceptions to this rule. This cannot be represented adequately with classical (propositional or first-order) logic [37]. In contrast, conditionals are quite perfect to model rules with exceptions, and ranking theory provides a particularly convenient framework to implement reasoning with such rules.

In most approaches to belief change, the new information is given by a propositional formula or by a set of propositional formulas. In the general framework developed by Kern-Isberner [26], change operations taking conditionals or sets of conditionals into account are provided, based on the central principle of conditional preservation. The principle of conditional preservation is defined both for quantitative probabilistic conditionals and for purely qualitative conditionals [26, 28]. In

Table 1 Ranking functions for the penguin example under different belief change operations, where—denotes that the ranking function is not defined on this world

	Ψ	Ψ°								
		contraction	ignorance	revision	marginalization	focussing	tunnel	conditionalization	fading out	
ω	$\kappa_{pen}(\omega)$	$\kappa_c^\circ(\omega)$	$\kappa_i^\circ(\omega)$	$\kappa_r^\circ(\omega)$	$\kappa_m^\circ(\omega)$	$\kappa_f^\circ(\omega)$	$\kappa_t^\circ(\omega)$	$\kappa_n^\circ(\omega)$	$\kappa_{d,1}^\circ(\omega)$	$\kappa_{d,2}^\circ(\omega)$
$b\,f\,p$	2	2	1	2	1	2	1	1	3	4
$b\,f\,\overline{p}$	0	0	0	2	0	0	0	–	0	0
$b\,\overline{f}\,p$	1	4	1	1	1	1	1	0	1	1
$b\,\overline{f}\,\overline{p}$	1	1	1	3	0	0	0	0	1	1
$\overline{b}\,f\,p$	4	4	3	4	2	2	1	3	4	4
$\overline{b}\,f\,\overline{p}$	0	0	0	0	0	0	0	–	0	0
$\overline{b}\,\overline{f}\,p$	2	5	2	2	2	1	1	1	2	2
$\overline{b}\,\overline{f}\,\overline{p}$	0	0	0	0	0	0	0	–	0	0
	Ex. 1	Ex. 2	Ex. 3	Ex. 4	Ex. 5	Ex. 6	Ex. 7	Ex. 8	Ex. 9	

this paper, we will only use the following special case for ranking functions and for changes taking only a single conditional $(B|A)$ into account. κ will denote the prior OCF of a change and κ° the posterior OCF.

Definition 1 (*c-change* [29]) A belief change from κ to κ° by $(B|A)$ fulfills the *principle of conditional preservation* and is called a *c-change* with $(B|A)$ if there exist integers $\kappa_0, \gamma^+, \gamma^-$ such that

$$\kappa^\circ(\omega) = \kappa(\omega) - \kappa_0 + \begin{cases} \gamma^+ & \text{if } \omega \models AB \\ \gamma^- & \text{if } \omega \models A\overline{B} \\ 0 & \text{if } \omega \models \overline{A} \end{cases} \tag{2}$$

The idea of the principle of conditional preservation is that all worlds having the same behavior with respect to the conditional $(B|A)$ should be treated identically. This is ensured by Eq. (2) governing the change when moving from κ to κ° according to $(B|A)$. The ranks of all worlds ω verifying $(B|A)$ are shifted by adding γ^+, the ranks of all worlds ω falsifying $(B|A)$ are shifted by adding γ^-, and the ranks of all worlds ω where $(B|A)$ is not applicable are not shifted. The constant κ_0 is a normalization constant ensuring that the minimal world rank under κ° is 0, as required for all OCFs. By identifying a (plausible) proposition A with the conditional $(A|\top)$, the c-change approach can also deal with changes induced by propositions.

3 Kinds of Forgetting Operations and Their Interrelationships

Forgetting is a phenomenon of everyday life. We forget to bring milk from the shop, misplace the key of our car, or fail to remember the birthday of a good friend. Here, we consider different forms and contexts of (intentional) forgetting and provide high-level, declarative specifications of the respective forgetting operation by just describing what the posterior belief state Ψ° is expected to fulfill after forgetting. As a proof of concept, we instantiate all considered forgetting operations using conditional beliefs, OCFs, and the general c-change concept outlined in Sect. 2. In most cases, we take a semantic view on the epistemic state Ψ [30], but we also consider cases where forgetting is more on the syntactic level of belief bases R representing explicit (conditional) beliefs. Figure 1 gives an overview of the different kinds of forgetting and their interrelationships for epistemic states and for belief bases presented and discussed in the following sections.

3.1 Contraction, Ignorance and Revocation

The most obvious kind of change that involves intentional forgetting is the direct intention to lose information. For instance, a navigation system might be informed about the (permanent) closure of a street, or laws governing data protection and data security might demand the deletion of information after a given period of time. The operation of contraction, i.e., giving up a (propositional) belief A, is central in the AGM theory [1] of belief change, while there is also a more general perspective [38, 42]). According to the AGM success postulate, a contraction with A results in not believing A afterward:

If Ψ is the prior and Ψ° the posterior belief state of a contraction with A, then $\Psi^\circ \not\approx A$.

Even further, contraction is an operation which typically results in a consistent belief state [32], a property which also holds for AGM contraction. The class of change operations obeying the principle of conditional preservation (see Eq. (2)) can be further specified by the success condition $\kappa^\circ \not\models (B|A)$ of contraction, i.e.,

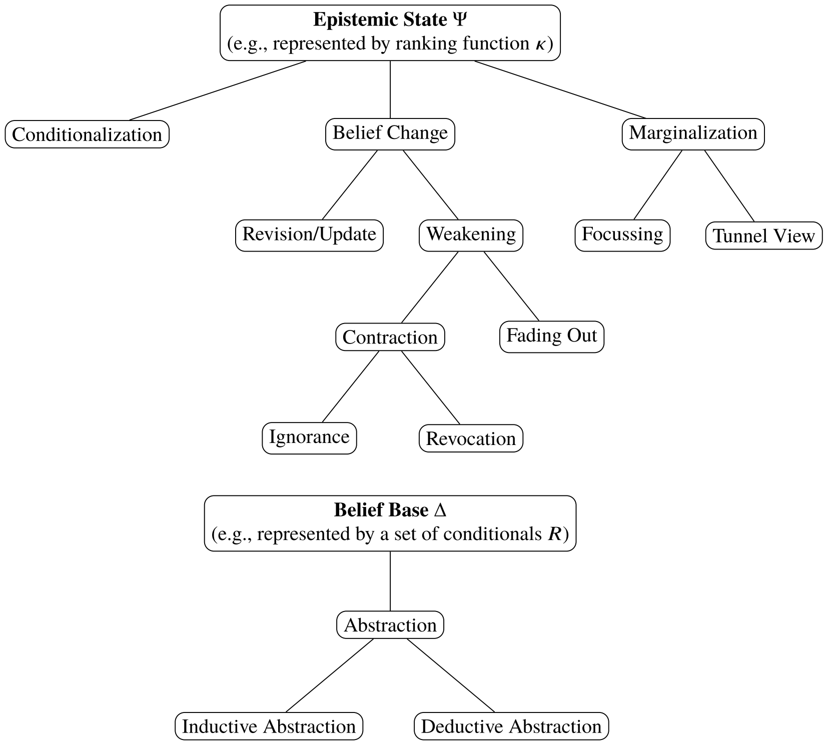


Fig. 1 Kinds of forgetting and their interrelationships for epistemic states ψ and for belief bases Δ

$$\kappa^\circ(AB) \geq \kappa^\circ(A\overline{B}), \tag{3}$$

and use this as a constraint for Equation (2). This leads to the notion of c-contraction [29] in a setting for studying iterated contractions of propositions and conditionals [23, 24].

Definition 2 (*c-contraction* [29]) A change from κ to κ° is called a *c-contraction* with $(B|A)$, if there exist integers $\kappa_0, \gamma^+, \gamma^-$ such that Eq. (2) from Definition 1 holds and the following conditions are satisfied:

$$\kappa_0 = \min\{\gamma^- + \kappa(A\overline{B}), \kappa(\overline{A})\} \tag{4}$$

$$\gamma^- - \gamma^+ \leq \kappa(AB) - \kappa(A\overline{B}) \tag{5}$$

Because κ° should *not* accept $(B|A)$, Eqs. (4) and (5) ensure, together with Eq. (2), that the rankings of worlds are moved such that under κ° the minimum rank of a world falsifying $(B|A)$ is at most the minimal rank of a world verifying $(B|A)$. In this way, it is guaranteed that $\kappa^\circ \not\models (B|A)$, i.e., $(B|A)$ has been forgotten. This illustrates how complex cognitive processes can be described quite easily in the high-level framework of ranking functions. Of course, the processes are significantly simplified but the c-change approach mimics basic features of forgetting by manipulating degrees of (im)plausibility.

Example 2 (continued) Consider $(\overline{f}|p)$, saying that "a penguin usually does not fly". The change from κ_{pen} to κ_c° as given in Table 1 is a c-contraction with $(\overline{f}|p)$, obtainable from Definition 2 by choosing $\gamma^- = 0$ and $\gamma^+ = 3$. Note that $(\overline{f}|p)$ is accepted by the prior belief state κ_{pen}, since $\kappa_{pen}(\overline{f}p) = 1 < 2 = \kappa_{pen}(fp)$, whereas $(\overline{f}|p)$ is not accepted by the posterior belief state κ_c°, since $\kappa_c^\circ(\overline{f}p) > \kappa_c^\circ(fp)$. Thus, the conditional knowledge that penguins usually do not fly has been forgotten.

While in the case of contraction the agent just gives up belief in a piece of information A, the agent could alternatively wish to become deliberately unsure about her beliefs concerning A. Examples for this kind of forgetting occur in particular in case of conflicting information, e.g., where one is unsure about the status of a person because the person is both a student and staff member. More generally, after becoming ignorant in A, neither A nor the opposite of A is believed:

> If Ψ is the prior and Ψ° the posterior belief state of a change to ignorance, then we have $\Psi^\circ \not\approx A$ and $\Psi^\circ \not\approx \neg A$.

Note that this formulation is not the only way of understanding ignorance. In a richer language like a modal logic, one might express ignorance with the agent knowing that she neither believes A nor $\neg A$ [48]. The changes obeying the principle of conditional preservation can be specified further to fulfill ignorance by adding

$$\kappa^\circ(AB) = \kappa^\circ(A\overline{B}) \tag{6}$$

to (2), ensuring both $\kappa^\circ \not\models (B|A)$ and $\kappa^\circ \not\models (\overline{B}|A)$.

Definition 3 *(c-ignoration)* A change from κ to κ° is called a *c-ignoration* with $(B|A)$, if there exist integers $\kappa_0, \gamma^+, \gamma^-$ such that Equation (2) from Definition 1 holds and the following conditions are satisfied:

$$\kappa_0 = \min\{\gamma^- + \kappa(A\overline{B}), \kappa(\overline{A})\} \tag{7}$$

$$\gamma^- - \gamma^+ = \kappa(AB) - \kappa(A\overline{B}) \tag{8}$$

Therefore, a change where the agent becomes ignorant in $(B|A)$ is a special case of a c-contraction with $(B|A)$.

Example 3 (continued) For $(\overline{f}|p)$ from Example 2, the change from κ_{pen} to κ_i° from Table 1 is a c-ignoration with $(\overline{f}|p)$ obtained by choosing $\gamma^- = -1$ and $\gamma^+ = 0$. Since $\kappa_i^\circ(\overline{f}p) = \kappa_i^\circ(fp)$ and therefore $\kappa_i^\circ \not\models (\overline{f}|p)$ and $\kappa_i^\circ \not\models (f|p)$, in the posterior state κ_i°, the agent is fully ignorant with respect to the flying abilities of penguins.

C-ignorations form a strict subclass of c-contractions. For instance, the c-contraction from κ_{pen} to κ_c° with $(\overline{f}|p)$ is not a c-ignoration with $(\overline{f}|p)$, since $\kappa_c^\circ \models (f|p)$, but it is an example of a *revocation*. Revocations with respect to A establish the belief in the opposite of A:

> If Ψ is the prior and Ψ° the posterior belief state of a revocation of A, then we have $\Psi^\circ \approx \neg A$.

Thus, contractions break down into the two disjoint subclasses of ignorations and revocations, formalizing two different types of forgetting a proposition or a conditional, respectively, i.e., becoming indifferent toward this proposition or conditional, or believing its negation.

3.2 Revision/Update

A major motivation for forgetting is when old information is to be replaced by new information, and such changes are the principal objective of revision, or update operations. If we believe (erroneously) that a person is a student, but then receive the information that she is a member of staff, we will *revise* our previous knowledge accordingly. Alternatively, if we receive the new information that a person previously known to be a student is a member of staff, we will *update* our previous knowledge accordingly. Revision reflects new information about a static world, whereas update occurs in an evolving world. In any case, there is a forgetting aspect because previously held beliefs might no longer be available, e.g., whether a person is a student. Revision is one of the central operations of the AGM theory [21], prioritizing the new information A over the existing beliefs. According to the success conditions of both AGM theory and the Katsuno-Mendelzon theory of belief update [25], we can specify revision resp. update as follows:

> If Ψ is the prior belief state and Ψ° is the posterior belief state after revision resp. update by A, then we have $\Psi^\circ \approx A$.

Forgetting is left implicit here. Beliefs conflicting with A have to be forgotten first before A can be adopted as new belief. In the propositional AGM framework, forgetting is made explicit by the Levi identity [21] which states that a revision with A can be obtained by first contracting (i.e., intentionally forgetting) $\neg A$ and then adding A to the remaining beliefs. Normally, if A is consistent, a revision results in a consistent belief state Ψ° [32]. A detailed examination of the notion of revision by (sets of) conditionals in the framework of ranking functions is done by Kern-Isberner [27], leading to the notion of c-revision:

Definition 4 (*c-revision* [27]) A change from κ to κ° is a *c-revision* with $(B|A)$ if there exist integers $\kappa_0, \gamma^+, \gamma^-$ such that (2) and the following conditions are satisfied:

$$\kappa_0 = \min\{\gamma^+ + \kappa(AB), \kappa(\overline{A})\} \tag{9}$$

$$\gamma^- - \gamma^+ > \kappa(AB) - \kappa(A\overline{B}) \tag{10}$$

Because κ° should accept $(B|A)$, Eqs. (9) and (10) ensure, together with Equation (2), that the rankings of worlds are shifted such that under κ° the minimum rank of a world verifying $(B|A)$ is strictly less than the minimal rank of a world falsifying $(B|A)$. Note how Equation (2) is used as a general schema that can be adapted to the needs (in particular, with respect to success) of the considered change operation. This reflects the assumption that forgetting processes (or more generally, reasoning processes) share a common basic strategy that can be adapted to individual needs, depending on the context. The success condition of revision results in the inequality (10), in contrast to what is specified in condition (5) and condition (8).

Example 4 (continued) Changing κ_{pen} to κ_r° in Table 1 is a c-revision with $(p|b)$, saying "birds are usually penguins" which might be motivated by a trip to the Antarctic. In κ_{pen} we believe that a bird usually is not a penguin, i.e., $\kappa_{pen} \models (\overline{p}|b)$, and we also believe that a bird normally is able to fly, i.e., $\kappa_{pen} \models (f|b)$. In κ_r° we believe that birds are usually penguins, i.e., $\kappa_r^\circ \models (p|b)$. As a side effect we do not believe any more that birds are usually able to fly, i.e., $\kappa_r^\circ \not\models (f|b)$. Thus, the revision with $(p|b)$ results in forgetting $(\overline{p}|b)$, but also $(f|b)$.

3.3 Abstraction

Abstraction is one of the most powerful change operations both in everyday life and in science. Suppose an agent who has built up beliefs about bicycles and keeps rules for inferring whether an object is a bicycle. A rule might say, if an object "has a frame, two wheels, and a bike bell", then this object is a bicycle, or if the object "has a frame, two wheels, and there is no bike bell", then this object is a bicycle. In a deductive way, the agent may abstract a new rule: if an object "has a frame and two wheels", then this object is a bicycle. More generally, suppose that we have:

$$\Psi \mathrel{|\!\approx} r_1 : if\ A\ and\ B\ holds\ then\ infer\ C$$

$$\Psi \approx r_2 : if\ A\ and\ \neg B\ holds\ then\ infer\ C$$

Abstracting from the rules r_1, r_2 yields a new rule r_{new}:

$$\Psi^\circ \approx r_{\text{new}} : if\ A\ holds\ then\ infer\ C$$

This abstraction corresponds to a deductive inference. Here, a particular kind of forgetting arises on the level of rules: The inference of C from A does not depend on the status of B; thus, in r_{new} the detail B is forgotten. Moreover, the agent might forget the rules r_1 and r_2.

As another example, suppose an organization provides support for various types of printers. In the past, the support was separated by manufacturers $M = \{$`hp`, `canon`, `epson`, ...$\}$ of the printer. For each manufacturer $m \in M$, the organization maintained a rule stating how to fix a printer problem X. E.g., if the printer has the problem X and the manufacturer is m, then solution C is the right one. But an analysis showed that in most cases the solution C was applied, independently from the printer's manufacturer. Thus, the set of rules could be abstracted to one rule: If a printer has the printer problem X, then solution C is the right solution. This variant of abstraction corresponds to an inductive inference. Suppose that for Ψ we have:

$$\begin{aligned}
\Psi &\approx r_1 : if\ A\ and\ B_1\ holds\ then\ infer\ C\\
\Psi &\approx r_2 : if\ A\ and\ B_2\ holds\ then\ infer\ C\\
&\qquad\ldots\\
\Psi &\approx r_n : if\ A\ and\ B_n holds\ then\ infer\ C
\end{aligned}$$

In Ψ°, the agent might abstract the rules $r_1, \ldots, r_n$ to a new rule r_{new}:

$$\Psi^\circ \approx r_{\text{new}} : if\ A\ holds\ then\ infer\ C$$

In this case of abstraction, the details $B_1, B_2, \ldots, B_n$ are forgotten within the new rule. Moreover, the agent might even forget the rules $r_1, \ldots, r_n$. Actually, this kind of forgetting addresses both syntactic (via changing rule bases) and semantic (via changing Ψ) issues, in particular for the inductive case. We address both in the following. First, a rule like "if A and B holds then infer C" can be represented as a conditional $(C|AB)$. The rules considered in this case form a conditional belief base R, and the posterior belief base (after abstraction) is denoted by R°. Given this, we take a look at the two styles of abstraction, starting with deductive abstraction:

> Let Ψ be the prior belief state accepting the belief base $R = \{(C|AB), (C|A\overline{B})\}$, and Ψ° the posterior belief state after deductive abstraction. Then $\Psi^\circ \models (C|A)$, and the explicit beliefs have changed so as to include $(C|A)$, i.e., $(C|A) \in R^\circ$.

OCFs κ are a perfect model for deductive abstraction since they have the property of *Proof by case differentiation* [35]: If $\kappa \models (C|AB)$ and $\kappa \models (C|A\overline{B})$, then $\kappa \models (C|A)$. Thus, ranking functions comply with deductive abstraction without any need for change, i.e., we can choose $\kappa^\circ = \kappa$. For the syntactic change from R to R°, this means that a prior ranking function κ accepting $\mathcal{R}$ will also accept $(C|A)$. Modeling the forgetting of the explicit conditional beliefs can thus be achieved by setting

$$R^\circ = \big(R \setminus \{(C|AB), (C|A\overline{B})\}\big) \cup \{(C|A)\}.$$

Things are different for inductive abstraction:

> Let Ψ be the prior belief state accepting the belief base $R = \{(C|AB_1), \ldots, (C|AB_n)\}$, and Ψ° the posterior belief state after inductive abstraction. Then we have $\Psi^\circ \models (C|A)$, and the explicit beliefs have changed so as to include $(C|A)$, i.e., $(C|A) \in R^\circ$.

Since the B_i need not be exclusive nor exhaustive, the above property of case differentiation cannot be applied. This means that an accepting ranking function κ for R will, in general, not fulfill the abstraction property, i.e., κ will not accept $(C|A)$. A change operation which has the desired property of changing κ to a state κ° such that $\kappa^\circ \models (C|A)$ can be obtained by revising κ by $(C|A)$ or by $R \cup \{(C|A)\}$ (see Sect. 3.2 for more on revision). On the level of belief bases, this abstraction results in a change from R to R° containing the conditional $(C|A)$. This can result in

$$R^\circ = (R \setminus \{(C|AB_1), \ldots, (C|AB_n)\} \cup \{(C|A)\}$$

where the explicit beliefs $(C|AB_1), \ldots, (C|AB_n)$ are forgotten.

3.4 Marginalization, Focussing and Tunnel View

The blinding out of information can also be seen as a form of forgetting. One form of this process is the removal of certain aspects represented in the language. This has become known as marginalization. Examples of marginalization can be found in situations where a decision is made by taking only certain aspects into account. Marginalization is a central technique, most prominently known from probability theory. It reduces the signature so that certain signature elements are no longer taken into account. For $\Sigma' \subseteq \Sigma$, with $\Psi|_{\Sigma'}$ we denote the restriction of Ψ such that $\Psi|_{\Sigma'} \mathrel{|\!\approx} A$ iff $\Psi \mathrel{|\!\approx} A$ for all $A \in \mathcal{L}_{\Sigma'}$. Then, for the marginalization to $\Sigma' \subseteq \Sigma$ we have:

> If Ψ is the prior and Ψ° the posterior belief state of a marginalization, then we have $\Psi^\circ = \Psi|_{\Sigma'}$.

Thus, the forgetting aspect of marginalization is the reduction of the signature, which might be temporary in many cases. The result of a marginalization in the view of common sense is the forgetting of details that are dealt with by some part of the signature. For conditional logics, marginalization and many other relationships among these logics have been investigated in detail [6, 7, 39, 41], providing the following definition and proposition for OCF-based semantics.

Definition 5 Let κ be a ranking function over Σ and $\Sigma' \subseteq \Sigma$. The *marginalization of* κ *to* Σ', denoted by $\kappa|_{\Sigma'} : \Omega_{\Sigma'} \to \mathbb{N}$, is given by

$$\kappa|_{\Sigma'}(\omega') = \min\{\kappa(\omega) \mid \omega \in \Omega_\Sigma \text{ and } \omega \models \omega'\}. \tag{11}$$

Proposition 1 *Let* κ *be a ranking function over* Σ *and* $\Sigma' \subseteq \Sigma$. *Then for all* $(B|A) \in (\mathcal{L}_{\Sigma'}|\mathcal{L}_{\Sigma'})$, $\kappa|_{\Sigma'} \models (B|A)$ *iff* $\kappa \models (B|A)$.

Note that in the setting of mathematical category theory [33], $\kappa|_{\Sigma'}$ is obtained by applying a so-called *forgetful functor* since all information regarding $\Sigma \setminus \Sigma'$ being available in κ is forgotten in $\kappa|_{\Sigma'}$ [6, 7].

Example 5 (continued) The marginalization of κ_{pen} from Table 1 to $\Sigma' = \{b, p\}$ leads to the ranking function $\kappa|_{\Sigma'}$ over the smaller signature Σ' given by:

$$\kappa|_{\Sigma'}(bp)=1, \qquad \kappa|_{\Sigma'}(\overline{b}p)=2, \qquad \kappa|_{\Sigma'}(b\overline{p})=0, \qquad \kappa|_{\Sigma'}(\overline{b}\overline{p})=0.$$

In a straightforward way, the marginalized ranking function $\kappa|_{\Sigma'}$ can again be extended to a ranking function over the original signature Σ by setting $\kappa^\circ_m(\dot{b}\dot{p}\dot{f}) = \kappa|_{\Sigma'}(\dot{b}\dot{p})$, as shown in Table 1. Thus, we have two OCFs (1) $\kappa|_{\Sigma'} : \Omega_{\{b,p\}} \to \mathbb{N}$ and (2) $\kappa^\circ_m : \Omega_{\{b,p,f\}} \to \mathbb{N}$, where (2) extends (1) to a larger domain. Therefore, we observe here two forms of forgetting by marginalization: (1) the signature element f is forgotten in $\kappa|_{\Sigma'}$, and (2) in κ°_m the differences induced by the signature element f are forgotten, i.e., $\kappa^\circ_m(Af) = \kappa^\circ_m(A\overline{f})$ for every formula A.

Now let us think about a physician who examines a patient with a rare allergy. The physician has to be careful what medication to administer. This is focussing[1]: the process of (temporary) concentration on *relevant* aspects of a specific case. The physician, while being focussed on specific evidence, blinds out other treatments being not relevant for the given case. Thus, we can say:

> The operation of focussing on A first determines all relevant signature elements $\Sigma' \subseteq \Sigma$ with respect to the objective A of the focus, and performs a marginalization to obtain $\Psi^\circ = \Psi|_{\Sigma'}$.

Focussing defined this way is based on marginalization but crucially involves the aspect of relevance. Typically, this change of beliefs of an agent is only temporary.

Example 6 (continued) Consider κ_{pen} from Table 1 and assume that the focussing happens in a situation where it is not relevant whether a penguin is a bird or not. We would select $\{p, f\}$ as the relevant elements of the signature and perform a marginalization of κ_{pen} to $\Sigma'' = \Sigma \setminus \{b\} = \{p, f\}$, leading to the OCF $\kappa|_{\Sigma''}$ given by:

$$\kappa|_{\Sigma''}(fp)=2, \qquad \kappa|_{\Sigma''}(\overline{f}p)=1, \qquad \kappa|_{\Sigma''}(f\overline{p})=0, \qquad \kappa|_{\Sigma''}(\overline{f}\overline{p})=0.$$

---

[1] There is also a different view on focussing, distinguishing generic and evidential knowledge [16].

Before focussing, in the belief state κ_{pen}, we accept the belief $(b|\overline{f}p)$, meaning "a penguin which does not fly is usually a bird". As in Example 5 we can easily extend $\kappa|_{\Sigma''}$ to an OCF over Σ, as given by κ_f° in Table 1. Thus, after focussing and extending, in the state κ_f°, we do not believe that a non-flying penguin is usually a bird, i.e., $\kappa_f^\circ \not\models (b|\overline{f}p)$; a forgetting of $(b|\overline{f}p)$ has occurred.

Technically, focussing is a special form of marginalization since marginalization is used for realizing it. From the commonsense perspective leading the identifications of forgetting in this article, the distinction is given by the *relevance* aspect governing the selection of the respective subsignature elements.

Relevance plays also a crucial role for the kind of forgetting occurring in a tunnel view. A tunnel view denotes a change where also only some beliefs are taken into account but without respecting their relevance sufficiently. In everyday life, there are many situations where reasoning is restricted by a temporary resource constraint. In such a situation, a tunnel view can enable the agent to react faster due to less information load, but this might lead to non-optimal inferences or conclusions. A realization of tunnel view can marginalize out the signature elements that are not part of the tunnel. Even more, in a situation of a tunnel view the agent might not be able to make full use of her mental capacities. This can be modeled by restricting the capabilities of the inference relation $\mid\!\approx$, which we will denote by $\mid\!\approx^r$.

> A tunnel view with the tunnel $T \subseteq \Sigma$ and reasoning limitation r is a change which results in a belief state Ψ° which is Ψ marginalized to T and the agent using the inference relation $\mid\!\approx^r$ such that $\Psi^\circ \mid\!\approx A$ if and only if $\Psi|_T \mid\!\approx^r A$.

Tunnel view is a kind of change which is only temporary. A specific aspect of tunnel view is that the tunneled signature elements are selected without sufficient respect to relevance. While tunnel view might be negatively connoted, from a psychological perspective tunnel view can be seen as part of a protection mechanism against information overflow in a stress situation.

For implementing a restricted inference $\mid\!\approx^r$ in the framework of OCFS and conditionals, we introduce the *collapse to a rank r , the leveling of all ranks greater* than or equal to r. This reduces the agent's capability of differentiated reasoning by reducing in-depth discrimination of possible worlds according to their plausibility.

Definition 6 (*collapse, restricted acceptance,* $\models^r$) Let κ be a ranking function and $(B|A)$ a conditional. The *collapse* of κ to rank r is the ranking function

$$\lceil\kappa\rceil^r(\omega) := \min\{\kappa(\omega), r\}. \tag{12}$$

κ accepts $(B|A)$ under restriction to r, denoted by $\kappa \models^r (B|A)$, if $\lceil\kappa\rceil^r(\omega) \models (B|A)$.

Example 7 (continued) Let κ_{pen}, κ_m° and $\Sigma' = \{b, p\}$ be as in Example 5. The collapse $\lceil\kappa_m^\circ\rceil^1$ is given by κ_t° from Table 1. In the state κ_m° the agent believes that "non-flying penguins are usually birds", i.e., $\kappa_m^\circ \models (b|p)$. Under limited capacities (to rank 1) the agent does not believe this anymore, i.e., $\kappa_t^\circ \not\models (b|p)$. Thus, the restricted capacities by the tunnel view with tunnel $T = \Sigma'$ and $r = 1$ lead to a forgetting of $(b|p)$.

Here, we have used the marginalization k°_m defined on Σ that is obtained from $\kappa|_{\Sigma'}$ defined on the smaller signature Σ' (cf. Example 5). We could also collapse $\kappa|_{\Sigma'}$ to rank 1, yielding the OCF $\lceil \kappa|_{\Sigma'} \rceil^1$ over the subsignature Σ'. Note that extending $\lceil \kappa|_{\Sigma'} \rceil^1$ analogously to $\kappa|_{\Sigma'}$ to the full signature Σ also yields the OCF $\kappa^\circ_t = \lceil \kappa^\circ_m \rceil^1$.

From the commonsense perspective, tunnel view is different from just marginalization because here, the agent's inference capabilities are restricted. In focussing the marginalized elements are selected by relevance, whereas in tunnel view the elements are not selected by (subjective) relevance. Note that relevance here is an external concept that has to be further specified by humans but can be integrated smoothly into the framework of modeling beliefs with ranking functions.

3.5 Conditionalization

Conditionalization restricts our beliefs to a specific case or context. We assume the existence of a conditionalization operator $|$ on Ψ, where $\Psi|A$ has the intended meaning that Ψ should be interpreted under the assumption that A holds. Thus, independently of any particular realization, we may assume $\Psi|A \mathrel{|\!\!\approx} A$ for every A.

If Ψ is the prior state and Ψ° the posterior state after conditionalization, then $\Psi^\circ = \Psi|A$.

Conditionalization is inspired by probabilistic conditionalization where posterior beliefs are determined by conditional probabilities $P(B|A)$, where A represents the evidential knowledge due to this change, i.e., $P^\circ(B) = P(B|A)$. In the framework of OCFs an established realization of a conditionalization is given by Spohn [43, Def. 5]. Let κ be an OCF and A a proposition, then the conditionalization of κ by A is the ranking function $\kappa|A : Mod(A) \to \mathbb{N}$, defined on the models of A as follows:

$$\kappa|A(\omega) = \kappa(\omega) - \kappa(A) \tag{13}$$

The OCF obtained by a conditionalization with A is only defined on $Mod(A)$, thus a change from κ to $\kappa|A$ results in a forgetting of the values of $\Omega \setminus Mod(A)$ under κ.

Example 8 (cont.) We want to shift our belief κ_{pen} into the setting where penguins or non-flying birds are always present, $p \vee b\overline{f}$, with $\kappa_{pen}(p \vee b\overline{f}) = 1$. This corresponds to the conditionalization $\kappa_{pen}|(p \vee b\overline{f})$ of κ_{pen} by $p \vee b\overline{f}$ given by κ°_n in Table 1.

Conditionalization with A is quite a drastic type of forgetting because the models of $\neg A$ are completely forgotten. However, this is usually necessary when reasoning analytically over cases where beliefs regarding one case should not be affected by beliefs about other cases.

3.6 Fading Out

If we use the PIN code of our credit card rarely, the chances that we will not remember the PIN the next time we need it are much higher than in the case of frequent use. This fading out or decay of knowledge occurs in many everyday-life situations. It depends on a number of parameters, e.g., how often we use this credit card, the amount of time since we last used it, the similarity of the PIN code to some other combination of digits important to us, etc.

In cognitive psychology, fading out is a prominent explanation for forgetting. The first evidences go back to a self-experiment of Ebbinghaus [17], whose results are known today as *forgetting curve*. Although this concept strongly influenced cognitive architectures and has been further developed in this area (cf. [2]), there is no approach to model this phenomenon as a variant of belief change in knowledge representation and reasoning. As a step toward modeling this phenomenon we propose to understand fading out as an increasing difficulty to infer the information from the agent's belief state. With inferences, we associate an effort or cost function f depending on the current belief state Ψ (cf. the activation function in ACT-R [4] or SOAR [31]). Then, $\Psi \mathrel{|\!\approx}_f A$ if $\Psi \mathrel{|\!\approx} A$ and the activation value $f(A)$ is above a certain threshold. In KR formalisms, inferences based on thresholds are not very popular because they usually do not comply with logical standards. For instance, in probability theory, beliefs defined by thresholds are not deductively closed, not even closed under conjunction, and hence not very reliable [36]. But weakening of beliefs is possible by other means. First, we define the general specification of fading out:

> A fading out of A is given as a sequence of consecutive posterior belief states $\Psi_1^\circ, \Psi_2^\circ, \Psi_3^\circ, \ldots$ such that there is an n with $\Psi_n^\circ \mathrel{|\!\not\approx} A$.

A specific difficulty of defining a concrete fading-out operation is the requirement of the possibility of recovering/remembering the information A again. In a more refined version of the definition above, one could demand for an increasing effort over the time, i.e., the required effort of inferring A from Ψ_j is at least as high as the effort of inferring A in state Ψ_i for $i \leq j$. One might also expect that the effort for inferring A decreases if A is used in a current reasoning task, thus becoming more relevant. For realizing the possibility of recovering or remembering the information A after a fading out of it, it will be useful to introduce a remembrance operation for this kind of change.

For an approach to model fading out in the framework of ranking functions, we define the degree of strength [22] of a conditional $(B|A)$ with respect to κ as

$$\mathrm{str}(\kappa, (B|A)) = \kappa(A\overline{B}) - \kappa(AB). \tag{14}$$

The strength of a conditional can be positive, 0, or negative, and we have $\mathrm{str}(\kappa, (B|A)) > 0$ iff $\kappa \models (B|A)$.

Definition 7 *(c-weakening)* A change from κ to κ° is a *c-weakening* of $(B|A)$ if there exist integers k_0, γ^+, γ^- such that Eq. (2) and the conditions

$$\kappa_0 = \min\{\gamma^+ + \kappa(AB), \gamma^- + \kappa(A\overline{B}), \kappa(\overline{A})\} \tag{15}$$

$$\gamma^+ - \gamma^- \geq 0 \tag{16}$$

are satisfied. A *proper* c-weakening is given when (16) is refined to $\gamma^+ - \gamma^- > 0$.

It is easy to check that every proper c-weakening indeed reduces the strength of the conditional because $\mathrm{str}(\kappa^\circ, (B|A)) = \mathrm{str}(\kappa, (B|A)) - (\gamma^+ - \gamma^-)$.

Proposition 2 (weakening) *If the belief change from κ to κ° is a c-weakening of $(B|A)$ then $\mathrm{str}(\kappa^\circ, (B|A)) \leq \mathrm{str}(\kappa, (B|A))$, and if it is a proper c-weakening then $\mathrm{str}(\kappa^\circ, (B|A)) < \mathrm{str}(\kappa, (B|A))$.*

For instance, the OCF $\kappa^\circ_{d,1}$ (Table 1) is a proper c-weakening of $(b|pf)$, denoting flying penguins are usually birds, obtained from κ_{pen} by choosing $\gamma^+ = 1$ and $\gamma^- = 0$. Since $\mathrm{str}(\kappa^\circ, (b|pf)) = 1 < 2 = \mathrm{str}(\kappa, (b|pf))$, the strength of the conditional $(b|pf)$ is reduced by 1. A sequence of c-weakenings of $(B|A)$ with sufficiently many proper c-weakenings leads to a fading out of $(B|A)$.

Proposition 3 (fading out) *For an OCF κ and a conditional $(B|A)$ we have:*

1. *There exists a sequence $\kappa, \kappa^\circ_1, ..., \kappa^\circ_n$ of consecutive c-weakenings of $(B|A)$ such that this sequence is a fading out of $(B|A)$, i.e., $\kappa^\circ_n \not\models (B|A)$.*
2. *Any consecutive sequence $\kappa, \kappa^\circ_1, ..., \kappa^\circ_n$ of c-weakenings of $(B|A)$ containing at least $k = \mathrm{str}(\kappa, (B|A))$ many proper c-weakenings is a fading out of $(B|A)$, i.e., $\kappa^\circ_n \not\models (B|A)$.*

Example 9 (cont.) The sequence $\kappa_{pen}, \kappa^\circ_{d,1}, \kappa^\circ_{d,2}$ is a fading out of $(b|pf)$, where every step is a proper c-weakening of $(b|pf)$. Both the initial state κ_{pen} and $\kappa^\circ_{d,1}$ accept $(b|pf)$. In $\kappa^\circ_{d,2}$, the conditional $(b|pf)$ is not accepted and thus forgotten since $\kappa^\circ_{d,2}(pfb) = \kappa^\circ_{d,2}(pf\overline{b})$.

Note that if κ° is obtained from κ by a c-contraction with $(B|A)$ (Sect. 3.1), then the sequence κ, κ° is a fading out of $(B|A)$ in the sense of Proposition 3. For the other direction, the following observation holds:

Proposition 4 *For every ranking function κ and every sequence $\kappa, \kappa^\circ_1, \ldots, \kappa^\circ_n$ of consecutive c-weakenings of $(B|A)$ with $\kappa^\circ_n \not\models (B|A)$ the change from κ to κ°_n is a c-contraction with $(B|A)$.*

Thus, in the framework of c-changes a c-weakening is a generalization of c-contraction, and every sequence of c-weakenings approximates a c-contraction. The ordinal nature of ranking functions allows for a particularly smooth and elegant modeling of weakening and fading out.

Example 10 (cont.) For the sequence $\kappa_{pen}, \kappa^\circ_{d,1}, \kappa^\circ_{d,2}$ (Example 9), the change from κ_{pen} to $\kappa^\circ_{d,2}$ is a c-contraction obtained by choosing $\gamma^+ = 2$ and $\gamma^- = 0$ (Definition 2).

In addition to Propositions 3 and 4, for every κ and every $(B|A)$ accepted by κ there always exists a fading out of $(B|A)$ realizing a c-ignoration with $(B|A)$ (see Sect. 3.1). For instance, the c-contraction in Example 10 is also a c-ignoration. Note that for modeling the pychologically inspired operation of fading out, alternatives to the treatment presented above exist, e.g., activation-based conditional inference [49–51]. For instance, one may introduce a threshold parameter in the framework of ranking functions. Especially for fading out, a more explicit modeling of time is desirable; from such a modeling, also other operators like tunnel view may profit.

4 The Cognitive and Empirical Foundation of IF

Given the formally fine distinction of different forgetting operators introduced previously, the question arises if and how humans do apply such operators and if these principles are transferable to human forgetting mechanisms.

4.1 Architectural Aspects of Cognition

From the perspective of cognitive science, intentional forgetting relies highly on information processing strategies and memory effects. The effects of intentional forgetting (IF) vary depending on the memory system involved. While early research focussed on long-term memory (LTM) using directed forgetting paradigms (e.g., [10, 34]), recent studies have also explored IF in working memory (WM). WM plays a crucial role in selecting, processing, and transferring relevant information while ignoring irrelevant data. Given its limited capacity, not all information can be stored and there is even an indication for a fast removal of some information (e.g., [14]).

Besides the distinction into LTM and WM, memory can be categorized into declarative and procedural systems, each differently affected by IF. Declarative memory stores facts [47]. Procedural memory, on the other hand, involves skills and habits acquired through practice, such as riding a car [12]. Procedural memory is often described by production rules ("if A then B") that can comprise stimulus-response mechanisms from behaviorism and cue-target connections in cognitive psychology. IF effects are well documented for declarative memory, particularly in episodic memory, where individuals can suppress specific information (e.g., [10, 34]).

Intentional forgetting (IF) is typically studied through tasks that test how people respond to previously learned stimuli. A common method uses paradigms where participants repeatedly respond to specific cues while being instructed to remember or forget certain associations. If IF disrupts retrieval, participants are likely to respond slower or make more mistakes when recalling items they were instructed to forget. If IF works through activation, response patterns remain stable, but recall efficiency decreases for forgotten items as the focus shifts to more relevant information.

Two main experimental paradigms have been so far used to study IF [13]: the *list-method paradigm* and the *item-method paradigm*. In the list-method paradigm, participants learn two separate word lists. After learning the first list, one group is told to forget it (through an Oops-paradigm) before learning the second list, while a control group is instructed to remember both. During the memory test, the "forget" group typically recalls fewer words from the first list but performs better on the second list. This suggests that forgetting reduces interference from the first list, allowing for more effective learning of new information.

In contrast, the item-method paradigm involves showing participants individual words, each followed by a cue to either remember or forget it. Memory for all words is later tested, revealing that participants recall fewer words they were instructed to forget compared to those they were told to remember. This paradigm relies more on *encoding-based mechanisms*—participants can suppress an item as soon as they see the "forget" task.

In contrast, IF effects on *procedural memory* are less clear. It is often encoded implicitly and hard to be controlled consciously [15, 45]. However, the findings from the typical research paradigms used in cognitive science so far are not directly transferrable to the formal framework of forgetting operations introduced before. Since the properties of the operations need to be also understood on a cognitive level in order to bring them into applications that should facilitate human handling of information, we developed new paradigms for procedural memory (Alien task and Counting Game).

4.2 IF in Procedural Memory

In a first study to tackle IF effects on procedural memory, the goal was to assess the cost of rule manipulations (i.e., in the form of a revision) compared to a baseline of an additional rule [11]. We used a categorization task based on a fantasy scenario in which participants were presented with images of aliens. The task was to categorize based on appearance by determining from which of four planets they come from. For categorizing them, four rules about their features (e.g., based on the number of eyes, antennae, etc.) were introduced that the participants had to memorize. The rules were always conjunctions of two *if-then* rules, hierarchically categorizing the aliens based on two features. The first feature allowed a categorization into two of the four planets (e.g., *If an alien has stripes, it comes from either planet A or planet B*) and the second feature then allowing to uniquely identify the right answer (e.g., *If an alien has 3 eyes, it comes from planet A*). Then, participants were presented with an image of an alien. Figure 2a shows an example of such a stimulus. After a first set of tasks to practice the rules, a rule was either changed or an additional rule was added. When adding a rule, it was either following the same hierarchy as the existing rules (*within-condition*; e.g., also affecting planet A and B), or broke the hierarchy (*across-condition*; e.g., affecting planet A and D). Overall, we collected data from 165 participants. The data is openly avaiblable on GitHub.[2]

We estimated the cognitive load caused by the revision or the addition of a rule by measuring the impact on categorization correctness and response time. Changes to the set of rules had a substantial impact on performance compared to the training phase with respect to accuracy ($M = 0.9$, $MAD = 0.1$ in training phase; $M = 0.82$, $MAD = 0.18$ for revision; $M = 0.82$, $MAD = 0.14$ for across; $M = 0.79$, $MAD = 0.11$ for within). We didn't find a significant difference in correctness between revision and addition; however, with respect to the response time, we found that the application of the rules after a revision was significantly faster than for an added rule ($mean = 3508ms$ compared to $mean = 3838ms$, MWU: $U = 2295$, $p = 0.04$). This indicates that each rule adds a substantial overhead, suggesting that updating rules is in fact beneficial over simply learning additional rules. Additionally, we found interactions between the old rule and the new rule: When both rules were in accordance (i.e., would lead to the same categorization of an alien), they improved performance significantly compared to situations without interaction. Conflicts, however, didn't significantly impact performance. This indicated that the revision of rules does not erase the old rule perfectly, but leaves artifacts behind, that can even be utilized to improve performance.

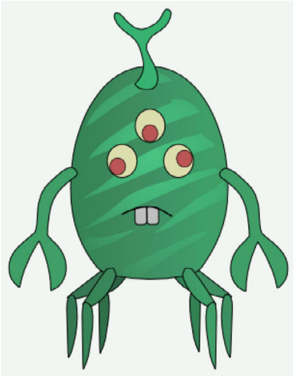
Overall, the Alien task allowed us to gain insights into effects of rule manipulations in procedural memory. We could show that they show similar effects to declarative memory, and can be manipulated in a controlled fashion. However, the scenario is not suited to cover *all* forgetting operators introduced before, leading to the development of the *Counting Game*.

Counting Game

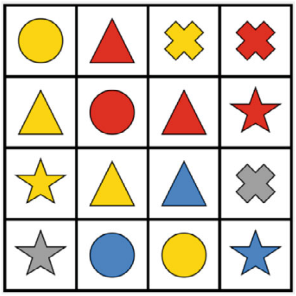
The goal of the study is to measure and compare cognitive costs of different types of forgetting. In order to achieve that, we developed an abstract experimental paradigm based on shapes and colors called the *Counting Game*. The game consists of *trials*—defined by a grid of 16 objects that vary in their *shape* (circle, triangle, cross, star) and *color* (red, yellow, blue, gray), and a winner, i.e., the color that has the largest number of objects. The "competing" colors are red, yellow and blue, while gray elements are optionally present and used as filler objects. For example, Figure 2b shows a trial where yellow is the winner with 6 objects. A valid trial must have only one winning color (which cannot be gray) and at most half of the total number of elements can be of the same color, in order to avoid complete domination of one color.

In the study, participants are presented with multiple trials whose winners they need to determine as fast as possible. There is a scoring system that awards them points for fast and accurate answers. For each correct answer they receive 10 points. Additionally, there is a bonus timer counting down from 20 seconds. For each correct

[2] https://www.github.com/brand-d/cogsci-2022-rulebasedcat.



(a) Example of an alien presented as a stimulus in the *Alien Task*.



(b) A representation of the Counting Game showing a trial whose winner is yellow. A rule "Red triangles count twice" changes the winner to red.

Fig. 2 Example Alien and counting game tasks

response provided in that time frame of 20 seconds, 1 bonus point per second left on the timer is awarded (e.g., a correct answer with 2 seconds left on the timer receives 10 + 2 points). Incorrect answers do not get any points, regardless of how fast they were provided.

When starting the game, the baseline rule is that each object, regardless of color and shape, counts as one. We introduce three types of rules that change how specific objects are counted:

1. *Count-twice* rules state that an object of a certain color and shape will count twice from now on.
   e.g., "Blue triangles count twice".
2. *Count-as* rules state that an object of a certain color and shape will now count as a different color.
   e.g., "Yellow stars count as red".
3. *Not-count* rules state that an object of a certain color and shape does not count anymore
   e.g., "Blue crosses do not count."

We implement six forgetting operators (Fig. 3) by introducing rules that change how certain objects are counted and thus manipulating the individual's rule system within the game. Figure 3 describes the implementation and shows example content for each operator. Each forgetting type needs a different amount of rules to be adequately tested, the maximum number being four for marginalization and abstraction. In order to make the rounds comparable in length, we added filler rules (of type *count-twice* or *count-as*) where necessary.

In the study, a *Phase* refers to a distinct stage within a round where a specific rule configuration is active. A round starts with *Phase 0* and progresses through multiple phases, where whenever a rule is introduced, updated or removed a new phase begins.

Fig. 3 Implementation description and example contents for each forgetting operator. Filler rules are typed out in *italics* and are included to "fill" the rounds that need less than 4 rules to implement a forgetting operator

	Implementation	Content
Contraction	Intentional removal of a specific rule from a rule set, resulting in a state where the contracted rule is no longer accepted. Implemented with one *count-twice* rule that is explicitly withdrawn later.	Blue stars count twice. Blue stars count as one again. *Red triangles count twice.* *Red triangles count as one again.*
Revision	Updating a rule set with new information that is prioritized over existing rules. Implemented with two *count-as* rules, for the same objects, with different target colors, implicitly forcing the individual to forget the first rule.	Red circles count as yellow. Red circles count as blue. *Red circles count as red again.* *Yellow crosses count twice.*
Marginalization	Reducing the set of variables considered, resulting in a simplified state. Implemented with four *not-count* rules affecting same colored objects, essentially removing one competing color.	Red triangles do not count. Red stars do not count. Red crosses do not count. Red circles do not count.
Abstraction	Generalizing specific rules, resulting in a more broadly applicable rule. Implemented with four *count-as* rules with same initial and target colors. The generalized rule states that one color fully counts as another.	Red circles count as yellow. Red stars count as yellow. Red crosses count as yellow. Red triangles count as yellow.
Conditionalization	Restricting rules by omitting ones that do not satisfy a given precondition. Implemented using specific *count-as* rules for gray objects, namely *if* there are gray shapes, they will count as a different color.	Gray crosses count as blue. *Red stars count twice.* *Red stars count as one again.* *Yellow triangles count twice.*
Fading Out	Gradual forgetting process where information becomes increasingly difficult to retrieve over manipulations. Implemented using one *count-as* rule followed by filler rules that make it more difficult to retrieve the first rule.	Blue stars count twice. *Red crosses count twice.* *Yellow triangles count twice.* *Red crosses count as one again.*

That means *Phase 1* begins after applying the first rule, *Phase 2* after the second rule, and so on. During every phase participants were provided with the same trials in a randomized order. With this we can examine how participants adapt to new rules and how prior rules influence the application of new ones.

For all rules there are certain trials that are *critical*. Those are trials whose winner changes after a rule is applied. For example, the trial shown in Fig. 2b is critical for the rule "Red triangles count twice", as its application changes the winner from yellow to red. If an individual provides an accurate answer to a critical trial, it indicates a correctly applied rule.

We developed a framework that generates the experimental tasks, that is available on GitHub.[3]

[3] https://www.github.com/saratdr/CountingGameFramework.

In total, we had six participant groups—one for each forgetting operator. In order to ensure that no unwanted effects or bias were introduced, each participant had a randomly assigned set of colors and shapes. That means that a rule "$color_1$ $shape_1$ count twice" might be about yellow circles for one person and blue stars for another. Naturally, the general structure of the rounds and rules is otherwise preserved.

We obtained data from $N = 97$ participants (age $19 - 73$, 42% female) recruited on Prolific.[4] The experiment was performed online as a web experiment. As compensation, the participants received 9 GBP after completing the experiment. All participants were native English speakers.

After giving consent to have their experimental data stored, participants were introduced to the game and its rules—the possible shapes and colors the element can have and their goal to find the winner among the red, yellow and blue elements as quickly as possible. Afterward, the scoring system was explained to them, indicating that they aim to get as many points as possible. Then, they had the chance to do a practice round with a few tasks and rules in order to get accustomed to the format. When finished with the instructions, the participants started with the study.

Analysis

This analysis aims to evaluate individuals' ability to intentionally forget information. We analyzed the accuracy and response time (RT) for relevant rules and trials in each Counting Game round.

We identified four participants, whose average accuracy scores were more than two standard deviations away from the mean, as outliers and therefore excluded them from the analysis. The final number of participants is 93. The overall mean accuracy was 82.94% (median: 100%, $SD = 37.62\%$). The mean accuracy of trials critical for at least one rule was 74.29%, in contrast to 93.48% for non-critical trials. The median RT across all trials was 1.70s. Critical trials had a longer median RT (2.21s) than non-critical ones (1.96s), suggesting increased cognitive effort when applying rules.

Figure 4 shows the mean accuracy and median response time for critical and non-critical trials for each forgetting operator across relevant phases. The consistent accuracy performance on non-critical trials shows that rules selectively impacted critical trials, as intended. In most cases, critical trials drop in accuracy, suggesting a difficulty in applying rules. We investigate each operator separately. Participants were able to contract a previously presented rule with a success rate of 84.17%. The relatively stable response time indicates that the process does not introduce a burden on the cognitive load. Revision, on the other hand, shows a larger increase in response time when a rule is introduced; however, participants seem to adapt to the actual revision process in Phase 2, with an accuracy of 74.0%. Marginalization and abstraction both show a progressive increase in difficulty as more rules are introduced over the phases and indicate that once a competing color is fully removed

[4] https://www.prolific.com.

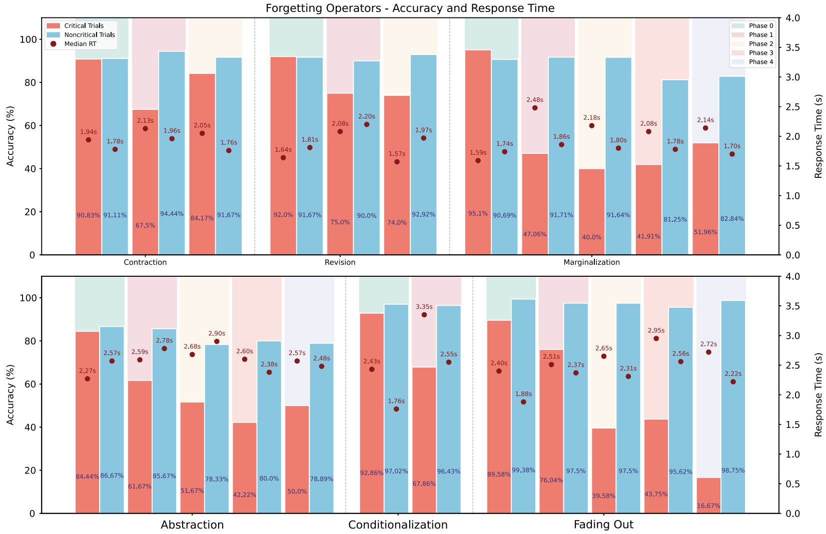


Fig. 4 Mean accuracy and median response time for each forgetting operator. Each depicted phase contains two bars—one for critical and one for non-critical trials

(marginalization) or a generalized, abstract rule can be used to replace all rules (abstraction), the accuracy starts increasing, however with only ca. 50% of success. The response times in both cases indicate the high cognitive costs of maintaining and applying multiple rules, which do not decrease during the marginalization or abstraction phase, even though the accuracy improves. Conditionalization imposes substantial cognitive demands, with accuracy dropping to 67.86% and response time increasing to 3.35s, indicating that selective application of a rule is rather complex for participants. The consistency of the accuracy in non-critical trials shows that they were able to forget of the rule when not applicable, however, with a higher response time indicating increase in cognitive load when processing rule conditions. Finally, fading out illustrates that participants fail to maintain an active rule across the phases. RTs indicate that this was not due to increased cognitive effort, but rather a gradual decline in ability to retain and apply the rule. These results provide a starting point for understanding the cognitive demands of different forgetting operators. However, in order to isolate specific effects of forgetting, we must first establish the inherent complexity of this paradigm's rules and tasks themselves.

5 Conclusions

In this chapter, we identified and elaborated a broad range of belief changes involving forgetting. From a commonsense inspired computer science point of view, we presented high-level formalizations of forgetting operators and their interrelationships. As a proof of concept of these formal foundations, we employed conditional beliefs and ranking functions for introducing realizations of all considered forgetting operators and for illustrating their formal properties.

Beyond their formal derivation, we investigated how these forgetting operators align with human cognitive mechanisms. To this end, we developed multiple different testing paradigms that allowed us to investigate their effects on procedural memory. Our empirical results support the applicability of these operators, demonstrating how intentional forgetting can be further studied and modeled.

With our *Alien task* we were able to investigate the effects of manipulations in procedural memory. As a first step toward understanding and measuring the costs of IF, we could test *revision* of rules in a categorization task. Our results showed that changes to rules in procedural memory leave behind artifacts, which, however, could improve performance in cases where the old and the new rules aligned. Furthermore, the impact of revision on accuracy did not differ significantly compared to the addition of new rules, but the application of revised rules was significantly faster. However, while the paradigm is well suited to investigate effects in procedural memory in a fashion that is comparable to classical declarative study paradigms, it is not suited to cover all forgetting operators introduced in this paper.

The *Counting Game* paradigm allows for investigating cognitive costs of concept changes in general, e.g., rule updating, like addition, modification and deletion of information, through simple rule manipulations. Specifically, an empirical investigation of cognitive costs of forgetting operators is of special interest to us. In order to be able to isolate the effect of forgetting operators, we will first focus on the complexity of the experimental tasks themselves in order to establish what factors need to be accounted for and omitted in future effects and trend analyses. Once that is established, we will be able to create appropriate tasks that will test forgetting operators and enable us to derive cognitive costs that can be then compared to any formal logical framework of the operators.

Finally, a future goal and next step is to derive best practices for applying forgetting operators in formal systems and organizations. Given that some operators are more cognitively demanding, the challenge lies in analyzing which are best suited for specific situations and contexts to optimize human-system interaction, reducing cognitive load and increasing efficiency.

Acknowledgements We would like to thank Tanja Bock, Tina Frenzel, Paulina Friemann, Jonas Haldimann, Steven Kutsch, Luise Mevius, Lara Paoletti, and Hannah Tschannen for supporting our research project during the first or the second project period. The work reported here has been supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), project numbers 427257555 and 318378366.

References

1. Alchourrón, C., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic, 50*(2), 510–530. https://doi.org/10.2307/2274239
2. Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford University Press.
3. Anderson, J. (2007). *Kognitive psychologie*. Spektrum Akademischer Verlag. https://doc1.bibliothek.li/aab/flma119287.pdf
4. Anderson, J., Byrne, M., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*(4), 1036–1050.
5. Beierle, C., Bock, T., Kern-Isberner, G., Ragni, M., & Sauerwald, K. (2018). Kinds and aspects of forgetting in common-sense knowledge and belief management. In *Joint German/Austrian Conference on Artificial Intelligence (KI)*, LNAI, vol. 11117, pp. 366–373. Springer (2018).https://doi.org/10.1007/978-3-030-00111-7_31
6. Beierle, C., & Kern-Isberner, G. (2009). Formal similarities and differences among qualitative conditional semantics. *International Journal of Approximate Reasoning, 50*(9), 1333–1346. https://doi.org/10.1016/j.ijar.2009.04.006
7. Beierle, C., & Kern-Isberner, G. (2012). Semantical investigations into nonmonotonic and probabilistic logics. *Annals of Mathematics and Artificial Intelligence, 65*(2–3), 123–158. https://doi.org/10.1007/s10472-012-9310-1
8. Beierle, C., Kern-Isberner, G., Sauerwald, K., Bock, T., & Ragni, M. (2019). Towards a general framework for kinds of forgetting in common-sense belief management. *KI - Künstliche Intelligenz, 33*(1), 57–68. https://doi.org/10.1007/s13218-018-0567-3
9. Beierle, C., & Timm, I. J. (2019). Intentional forgetting: An emerging field in AI and beyond. *KI - Künstliche Intelligenz, 33*(1), 5–8. https://doi.org/10.1007/s13218-018-00574-x
10. Bjork, R. (1970). Positive forgetting: The noninterference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior, 9*(3), 255–268. https://doi.org/10.1016/S0022-5371(70)80059-7
11. Brand, D., Dames, H., Puricelli, L., & Ragni, M. (2022). Rule-based categorization: Measuring the cognitive costs of intentional rule updating. In J. Culbertson, A. Perfors, H. Rabagliati, V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Meeting of the Cognitive Science Society* pp. 2810–2817.
12. Cohen, N., & Squire, L. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science, 210*(4466), 207–210. https://doi.org/10.1126/science.7414331
13. Dames, H., Brand, D., & Ragni, M. (2022). Evidence for multiple mechanisms underlying list-method directed forgetting. In J. Culbertson, A. Perfors, H. Rabagliati, V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*. https://escholarship.org/uc/item/46921378
14. Dames, H., & Oberauer, K. (2022). Directed forgetting in working memory. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001256
15. Dames, H., Ragni, M., Kiesel, A., & Pfeuffer, C. (2022). *Hard to forget: No directed forgetting of stimulus-response associations*. https://psyarxiv.com/6hzys
16. Dubois, D., & Prade, H. (1997). Focusing vs. belief revision: A fundamental distinction when dealing with generic knowledge. In *Proc. ECSQARU-FAPR'97*. LNCS, vol. 1244, pp. 96–107. Springer.
17. Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur experimentellen psychologie*. Verlag von Duncker & Humblot.
18. Eichhorn, C., Kern-Isberner, G., & Ragni, M. (2018). Rational inference patterns based on conditional logic. In S. McIlraith & K. Weinberger (Eds.), *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18)*, pp. 1827–1834. AAAI Press. https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16834

19. Eiter, T., & Kern-Isberner, G. (2019). A brief survey on forgetting from a knowledge representation and reasoning perspective. *KI - Künstliche Intelligenz, 33*(1), 9–33. https://doi.org/10.1007/s13218-018-0564-6
20. de Finetti, B. (1937). La prévision, ses lois logiques et ses sources subjectives. Annales de l'Institut Henri Poincaré **7**(1), 1–68, Engl. transl. *Theory of probability*. J. Wiley & Sons (1974).
21. Gärdenfors, P., & Rott, H. (1995). Belief revision. In *Handbook of logic in artificial intelligence and logic programming*, (Vol. 4), pp. 35–132. Oxford University Press.
22. Goldszmidt, M., & Pearl, J. (1996). Qualitative probabilities for default reasoning, belief revision, and causal modeling. *Artificial Intelligence, 84*(1–2), 57–112. https://doi.org/10.1016/0004-3702(95)00090-9
23. Haldimann, J., Beierle, C., & Kern-Isberner, G. (2021). Syntax splitting for iterated contractions, ignorations, and revisions on ranking functions using selection strategies. In W. Faber, G. Friedrich, M., Gebser, & M. Morak (Eds.), *Logics in artificial intelligence—17th European Conference, JELIA-2021*. LNCS, vol. 12678, pp. 85–100. Springer. https://doi.org/10.1007/978-3-030-75775-5_7
24. Haldimann, J., Kern-Isberner, G., & Beierle, C. (2020). Syntax splitting for iterated contractions. In D. Calvanese, E. Erdem, & M. Thielscher (Eds.), *Principles of knowledge representation and reasoning: Proceedings of the 17th international conference, KR 2020*, pp. 465–475. https://doi.org/10.24963/kr.2020/47
25. Katsuno, H., & Mendelzon, A. (1991). Propositional knowledge base revision and minimal change. *Artificial Intelligence, 52*, 263–294.
26. Kern-Isberner, G. (2001). *Conditionals in nonmonotonic reasoning and belief revision*, LNAI, Vol. 2087. Springer.
27. Kern-Isberner, G. (2004). A thorough axiomatization of a principle of conditional preservation in belief revision. *Annals of Mathematics and Artificial Intelligence, 40*(1–2), 127–164.
28. Kern-Isberner, G. (2018). Axiomatizing a qualitative principle of conditional preservation for iterated belief change. In *Proceedings KR 2018*, pp. 248–256. AAAI Press.
29. Kern-Isberner, G., Bock, T., Sauerwald, K., & Beierle, C. (2017). Iterated contraction of propositions and conditionals under the principle of conditional preservation. In *GCAI-2017. EPiC series in computing*, Vol. 50, pp. 78–92. EasyChair. https://doi.org/10.29007/3Q8L
30. Kern-Isberner, G., Bock, T., Sauerwald, K., & Beierle, C. (2019). Belief change properties of forgetting operations over ranking functions. In*PRICAI 2019: Trends in artificial intelligence*. LNCS, Vol. 11670, pp. 459–472. Springe. https://doi.org/10.1007/978-3-030-29908-8_37
31. Laird, J. (2012). *The soar cognitive architecture*. MIT Press.
32. Levi, I. (1977). Subjunctives, dispositions and chances. *Synthese, 34*(4), 423–455.
33. Mac Lane, S. (1972). *Categories for the working Mathematician. Springer*. https://doi.org/10.1007/978-1-4612-9839-7
34. MacLeod, C. (1998). Directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches*, pp. 1–57. Lawrence Erlbaum Associates Publishers.
35. Makinson, D. (1994). General patterns in nonmonotonic reasoning. In D. Gabbay, C. Hogger, & J. Robinson (Eds.), *Handbook of logic in artificial intelligence and logic programming*, vol. 3, pp. 35–110. Oxford University Press.
36. Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo, Ca: Morgan Kaufmann.
37. Ragni, M., Kern-Isberner, G., Beierle, C., & Sauerwald, K. (2020). Cognitive logics—features, formalisms, and challenges. In *ECAI 2020*, pp. 2931–2932. IOS Press. https://doi.org/10.3233/FAIA200459
38. Sauerwald, K. (2022). *Semantics of belief change operators for intelligent agents: Iteration, postulates, and realizability, dissertations in artificial intelligence*, vol. 352. IOS Press. https://doi.org/10.3233/DAI352
39. Sauerwald, K., Beierle, C., & Kern-Isberner, G. (2024). Propositional variable forgetting and marginalization: Semantically, two sides of the same coin. In *FoIKS 2024, Proceedings*. LNCS, Vol. 14589, pp. 144–162. Springer. https://doi.org/10.1007/978-3-031-56940-1_8

40. Sauerwald, K., Ismail-Tsaous, E., Ragni, M., Kern-Isberner, G., & Beierle, C. (2025). Sequential merging and construction of rankings as cognitive logic. *International Journal of Approximate Reasoning, 176*, Article 109321. https://doi.org/10.1016/J.IJAR.2024.109321
41. Sauerwald, K., Kern-Isberner, G., Becker, A., & Beierle, C. (2022). From forgetting signature elements to forgetting formulas in epistemic states. In *Scalable uncertainty management 2022*. LNCS, Vol. 13562, pp. 92–106. Springer.https://doi.org/10.1007/978-3-031-18843-5_7
42. Sauerwald, K., Kern-Isberner, G., & Beierle, C. (2020). A conditional perspective for iterated belief contraction. In *ECAI 2020*, pp. 889–896. IOS Press. https://doi.org/10.3233/FAIA200180
43. Spohn, W. (1988). Ordinal conditional functions: A dynamic theory of epistemic states. In *Causation in decision, belief change, and statistics, II*, pp. 105–134. Kluver Academic Publishers.
44. Squire, L. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory, 82*(3), 171–177.
45. Tempel, T., & Frings, C. (2016). Directed forgetting benefits motor sequence encoding. *Memory & Cognition, 44*(3), 413–419. https://doi.org/10.3758/s13421-015-0565-8
46. Timm, I. J., Staab, S., Siebers, M., Schon, C., Schmid, U., Sauerwald, K., Reuter, L., Ragni, M., Niederée, C., Maus, H., Kern-Isberner, G., Jilek, C., Friemann, P., Eiter, T., Dengel, A., Dames, H., Bock, T., Berndt, J. O., & Beierle, C. (2018, September 24-28). Intentional forgetting in artificial intelligence systems: Perspectives and challenges. In F. Trollmann & A. Turhan (Eds.), *KI 2018: Advances in artificial intelligence—41st German Conference on AI, Berlin, Germany, Proceedings*. LNAI, Vol. 11117, pp. 357–365. Springer. https://doi.org/10.1007/978-3-030-00111-7_30
47. Tulving, E. (1985). How many memory systems are there? *American Psychologist, 40*(4), 385–398. https://doi.org/10.1037/0003-066X.40.4.385
48. van Ditmarsch, H., Herzig, A., Lang, J., & Marquis, P. (2009). Introspective forgetting. *Synthese, 169*(2), 405–423. https://doi.org/10.1007/s11229-009-9554-4
49. Wilhelm, M., Howey, D., Kern-Isberner, G., Sauerwald, K., & Beierle, C. (2023). Activation-based conditional inference. *Journal of Applied Logics*, *10*(2), 221–248. https://collegepublications.co.uk/downloads/ifcolog00058.pdf
50. Wilhelm, M., Howey, D., Kern-Isberner, G., Sauerwald, K., & Beierle, C. (2021, September 28). A brief introduction into activation-based conditional inference. In C. Beierle, M. Ragni, F. Stolzenburg, & M. Thimm (Eds.). *Proceedings of the 7th Workshop on Formal and Cognitive Reasoning co-located with the 44th German Conference on Artificial Intelligence (KI 2021)*. CEUR Workshop Proceedings, Vol. 2961, pp. 4–8. CEUR-WS.org. http://ceur-ws.org/Vol-2961/paper_1.pdf
51. Wilhelm, M., Howey, D., Kern-Isberner, G., Sauerwald, K., & Beierle, C. (2022). Integrating cognitive principles from ACT-R into probabilistic conditional reasoning by taking the example of maximum entropy reasoning. *The International FLAIRS Conference Proceedings, 35*. https://doi.org/10.32473/FLAIRS.V35I.130674

Examining Human and AI Interactions in AI-Based Application Systems—Realizing Local and Remote Experiments in Industry 4.0 Settings

Marcus Grum

1 Introduction to Human AI Interaction

In business context, process tasks are often realized under high competitive pressure [36]. As a new opportunity to be compatible as business, Artificial Intelligence (AI) seems to be attractive to accelerate task realizations across various task types, reduce costs and minimize failures at mindless work [32]. So, human workers tend to use AI or cooperate with AIs. For this reason, as AIs are recognized as individual process participants capable of using their actuators to manipulate their environment and their sensors to perceive their task-related environment—in this context, we can speak of AI-based cyber-physical systems (AI-based CPS)—we find them in a growing number of teams consisting of AI-based CPS and human employees [16].

However, for humans, relying on an AI-based system is a trustworthy, complex issue, since AIs recognize complex pattern based on data and compose technique-inherent logic, so that its analyses and decisions are not always immediately understandable but rather non-transparent for people [9]. Further, if AI-based CPS are designed to continuously learn and adapt to its environmental changes, AIs might drift and shift behavior involuntarily or surprisingly from human worker perspective [6]. This puts trustworthiness of AIs, their reliability and AI autonomy design in focus, especially when dealing with autonomous AI-based systems [39].

On the other side, in organizational context, business transforms permanently and demands for continuous adaptation, too. For instance, one can find regularly amended laws, market and price variations, reorganizations of teams, modifications in the course of processes, changes in the environmental factors of individual process steps as well as required alterations at individual process task realizations [14]. There is therefore not only a need for continuous adaptation for AIs, but also for human

M. Grum (✉)
University of Potsdam, Potsdam, Germany
e-mail: grum@uni-potsdam.de

J. O. Berndt et al. (eds.), *Intentional Forgetting with Intelligent Systems*, Progress in IS,
https://doi.org/10.1007/978-3-032-17621-9_8

process participants to align with business needs at realization of a certain task. This goes along with learning new routines and forgetting outdated routines.

For motivation, if it was possible to use the capability of continuously learning AI-based systems to autonomously detect new complex process patterns and forget outdated process patterns by intention, there might be a chance to guide a teaming of human and AI to be aligned with business needs [12]. The AI so becomes a management instrument [17]. So, as a research resolution, in a first step, understanding the effect of learning and forgetting AI-based CPS on learning and forgetting human workers is essential. Then, in a next step, the designing of instruments is attractive, which support the provision of learning and forgetting AI-based CPS that fit the individual needs of human workers—the latter are supposed to learn or forget due to the well-designed influence of the previously mentioned AIs and is aligned with business needs. The objective of intended studies so refers to the exploratory examination of evolving AI on human forgetting and vice versa. For this reason, the following research question will be focused.

What effects do forgetful AI-based CPS have on forgetful people?

In a design-science-oriented manner, this research article contributes with an artifact of an *AI Interaction Instrument (A3I)*, which is to be designed to model interactions of learning and forgetting human workers with learning and forgetting AI-based process participants and vice versa. It functions as the basic research instrument to examine the quantified effect of forgetful AI-based CPS on forgetful people. With modifications, it also enables examining effects of forgetting people on forgetful AIs.

The remaining part structures as follows. The second section presents foundational terms and definitions to characterize cooperation with AIs in Industry 4.0 context. The third section frames the research conducted as design-science research and clarifies requirements on the research artifact of A3I. In the fourth section, the artifact design is presented. The last sections address its live demonstration in experiments, discuss its evaluation and conclude the article.

2 Foundation of Cooperation with AIs in Industry 4.0 Context

2.1 AI-Based Application Systems

Researchers, practitioners, everyday users and the literature often have a different understanding when mentioning AI and application systems. Even different disciplines have an heterogeneous understanding when speaking about AI [14]. The following thus tries to establish a common understanding for the research realized. Relevant terms are described in detail below.

Application Systems. In general, an *application system* terms a system that contains all programs that are developed, introduced and used as application software for a specific operational application example, including the technology, data and IT infrastructure, consisting of hardware, software, data, storage technology, communication and network, as well as the persons responsible for it [30, 33]. An application system can use an AI at a certain point of a process, which is referred to as *point-wise AI utilization* [14].

AI-Based Application Systems. An application system can be classified as an *AI-based application system* if that system supports the user's tasks with the help of hardware, software and/or logical elements (including organizational elements) to perform its tasks on the foundation of AI. Here, the application system considers AI as a foundation of an application house that builds on it. The AI represents the basis for on-building system components being triggered by the AI [14], because these have been attached onto the AI ground floor. Here, AI structures represent the process, because the AI comes up with a processual behavior [14]. As the AI-triggered systems can either refer to further AI-based systems or non-AI-based systems, a joint application system is constructed that shows different levels of cognition.

AI-based CPS. If the AI-based application system is brought into a permanent loop of (1) using its own sensors and data input to perceive its environment, (2) processing its perception on the basis of AI, and (3) carrying decisions out via its actuators or connected workflows to manipulate its environment—this is perceived by the AI-based application system again and the loop mentioned is closed—one can speak from an *AI-based Cyber-Physical System* (AI-based CPS) [16]. This type of system is particularly interesting for the research carried out because, firstly, it enables the modeling of complex interaction patterns between humans and AI based on this closed loop and, secondly, it enables continuously learning AI in live situations.

Interim Conclusion. By now, these kinds of systems have been demonstrated and stacked to cooperating AI-based CPS: These approaches have demonstrated learning and unlearning AIs in transforming production contexts [14]. However, the learning and unlearning AI-based CPS have not been considered as instrument to guide forgetting of human workers, yet. A research gap becomes apparent, here.

2.2 Learning and Forgetting AIs

Learning. The state of knowing the expected or successful patterns and the ability to request desired patterns is achieved by *learning* procedures. AI typically gets into this state by carrying out training processes. For instance, Artificial Neural Networks (ANNs) are initialized with random weights, first. Then, due to training algorithms, such as backpropagation [42, 43], PROP [40], Quickprop [7], conjugated gradients [19, 47], and L-BFGS [4], weights are as long modified as a desirable quality criteria or pre-defined condition has been reached. If the AI design and training was well performed, then the AI is able to request desired patterns because it knows the expected or successful behavior.

The *behavioral adaption*, for instance in changed process patterns (e.g., production routines), is preceded by knowing the expected or successful patterns and the ability to inhibit unwanted patterns. Feldman [8] argues that routine dynamics are primarily concerned with learning, whereas Hedberg [18] includes unlearning and forgetting. While new information generates new options in routine performance, prior knowledge and experience can restrain variability therein [22] and interfere with adapting to the new routine as employees tend to relapse into the old routine. Older routines can also interfere with new knowledge as cognitive capacity is allocated for routine performance and the consolidation of new information. Intentionally inhibiting the recall of undesired tasks and objectives can free cognitive capacity [1, 41] and ease the transition. This paper, therefore, concentrates on researching these operations instead of learning. Two concepts can be used to explain the elimination or suppression of older knowledge: unlearning and forgetting [29].

Unlearning. On the routine level, *unlearning* is either substituting, overwriting, disregarding or stopping to show an unwanted behavior [48]. Enforcement and enactment are sufficient to overwrite it, if executed long enough [20, 35]. Fiol and O'Connor [10] identify three steps of unlearning: destabilizing and discarding an old routine as well as the subsequent establishment of a new one. In AI context, once an AI has learned something, unlearning is achieved in continuing training or refining the AI. In particular in continuously learning contexts, unlearning AIs can be observed [6]. For instance, due to environmental changes, data drifts or variations in learned relations, once learned behavior cannot be requested any more. In AI context, often, this is referred to as *catastrophic forgetting* [26, 28].

Forgetting. The unlearning concept has been criticized from the psychological perspective [23, 44]. The term *forgetting* is preferred as it also connects to broader explanations of the underlying cognitive aspects, which is in particular attractive when facing AIs as artificial cognitive systems. Forgetting is generally defined as inhibiting the recall of previously learned information [46] and addresses both accidental and intentional aspects. The intentional aspect of forgetting includes active suppression and sorting out knowledge without a functional equivalent [2, 41, 52, 53]. The individual is therein aware about unwanted, habituated routines and actively aims at inhibition. This accounts for AIs as well as human employees. However, intentional forgetting AIs require an AI to be aware about unwanted, habituated routines, or rather require an AI to reflect on itself. Please be aware of the differentiation of *amnesic AIs*, which show a partial or complete inability to remember past experiences or store new memories after the causative event, such as an accident, a management intervention, or AI design decision. However, forgetting rather focuses on a natural and daily progress that disregards the inabilities mentioned. This type of concept is therefore particularly interesting for the research carried out.

Interim Conclusion. Although learning and forgetting AI-based CPS were implemented and brought to experimentation [15, 16], experiments have not examined the effect of AIs on human forgetting. Further, by now, self-reflecting AIs justifying to actively inhibit unwanted, habituated routines is a research objective. By now, human knowledge managers are still required for being responsible to make an AI intentionally (at least from the manager's perspective) forget and guide its knowledge development [12]. Here, concrete AI-human-adequate knowledge management instruments are to be worked out and researched, which includes experimental procedures to proof functioning of these instruments. Thus, a research gap becomes apparent, here.

However, the experimental procedures mentioned go hand in hand with the design of human-machine interfaces bringing evolving AIs and forgetful human in an interaction.

2.3 Human-Machine Interfaces and AIs

The literature describes the factors of *AI transparency* [9, 50], *AI autonomy* [39], and *AI reliability* [5, 31, 49] that influence the nature of the human-machine interface and thus the relationship between humans and AI. These factors are described in more detail below.

AI Autonomy. AI autonomy determines the extent to which the AI supports the user in the task to be performed and how much responsibility the user assumes [39]. According to Peres et al., AI autonomy in Industry 4.0 can be divided into five degrees [39]. These range from scenarios with purely supportive activity, in which the user assumes full responsibility and the AI does not implement any independent decisions (level 1), to a scenario of partial autonomy, in which the AI implements its decisions fully autonomously, the user can intervene in the implementation of the decision (grade 4), through to a scenario of full autonomy in which no user needs to be involved in the AI's decision-making process and the AI assumes full responsibility (grade 5).

AI Transparency. Transparency in the context of AI is a complex, multi-factorial construct with several dimensions such as the accessibility and interpretability of information about the AI's decision-making process [9]. Meanwhile, several techniques have evolved to rise transparency, such as sensitivity analyses, feature visualization, saliency visualization, attribution and knowledge extraction [14]. However, the breakthrough in an all-embracing whitening the AI's or rather ANN's blackbox is still missing. Thus, an AI system combination is particularly interesting for the research carried out, because partly transparent outcomes can be used for reasoning about the AI's decisions.

AI Reliability. Often, reliability is referred to as the circumstance that a test, survey or measurement can be reproduced and the same model or outcome can be generated again and again [3]. In case of AI context, an AI is reliable if it is able to reproduce an outcome—even in changing and dynamic contexts. One can speak about an equality of the measurement results when used at different times. In this research, however, AI reliability is referred to as a reliableness of a human to rely on an AI's outcome, because the AI's outcomes are correct under all circumstances, such as changing and dynamic contexts or different times. It differs from AI accuracy in that decimal deviations of the AI outcome from the target labels are not taken into account—the focus is on the final classification from the end user's point of view, which is either correct or incorrect [5]. Knowing that there are cases in which the evaluation of an AI can be right or wrong, but also more complex cases in which evaluations are not one hundred percent clear, for example because the evaluation can be assigned to several classes, the design of an AI reliability metric is complex and case sensitive.

AI Trust and Frustration. Since the interaction between humans and AI can influence the physical, psychological and social characteristics of humans (so-called human factors), it is also interesting for the present work to determine which human factors are influenced by the design of the AI in the application. The literature shows that the design of human-machine interfaces has a major influence on human factors, such as *AI trust* and *AI frustration* [24, 49]. Thus, as these factors are involved at the cooperation with AIs in business context, their meaning in transforming business environments, which includes the need for learning of new routines and forgetting of outdated routines, is essential.

Interim Conclusion. However, an AI's autonomy and transparency design has not been quantified and researched in transforming business environments. Thus, their effect on learning new routines as well as forgetting outdated routines represents a research gap. A further research gap becomes apparent when considering an AI's autonomy, its transparency design and reliability as management instrument to align an AI-human teaming with business needs of permanently transforming contexts.

Having the objective of intended studies in mind, which refers to the exploratory examination of evolving AI on human forgetting and vice versa, transforming business environments bring complexity into the object of investigation: Is the AI evolving due to environmental changes? Is this evolution to be characterized as beneficial or (intentional or catastrophic) AI forgetting? One might assume human workers to forgive AI faults if the AI's decision was made transparent. But what about correct AI behavior in transformed business situations not being understood by human parties? Is AI being unjustifiably accused here and restricted in its autonomy? It could be that repeated AI competency tests provide human workers with indications of the AI's reliability, which makes experimenting with the A3I an important step in researching the sustainable design of cooperation between humans and AI.

3 Methodological Researching AI Interaction

This research is realized by following the Design-Science Research Method (DSRM) by Peffers [37, 38]. So, the design of a research artifact is worked out based on a research problem first. The research problem and the research gap were presented in the first sections. The artifact of the A3I is then designed and brought to demonstration. Finally, it is evaluated to what extend the initial research problem has been tackled with the A3I research artifact successfully.

Requirement Collection. As part of a design-science-oriented, behavioral research, and for the reason of having a software artifact, the following requirements were collected. The A3I research artifact development was guided by these requirements. Further, this collection has functioned as quality gate for the artifact developed. Thus, the following will support the connection of future research, too.

1. The experiment shall enable a *node-independent device configuration*, so that it can run on a laptop, raspberry or any further device, such as the devices of the Potsdam Zentrum für Industrie 4.0 (ZIP4.0).
2. The experiment shall support an *online and on-site experiment realization*, so that test persons can take the experiment in arbitrary local and remote configurations.
3. In order to guarantee, that the observed effects are not effected by prior knowledge, a new routine or rather *production process* needs to be learned in a *first experiment phase*. This is referred to as test person *training* and can distinguish from the real experiment environment because of pedagogic elements, such as explanations, tutorials, step-by-step guides, supporting elements and exercises.
4. In order to assure, that the learned production process is mastered successfully, the experiment needs to have a performance assessment within the real experiment environment in a *second experiment phase*. This is referred to as *reference scenario* (without AI support). Here, the as-is performance or routine behavior will be identified. It will be used for comparing the performance with the scenario having an AI support (cooperation focus) as well as with the scenario having an AI breakdown (forgetting focus).
5. The standard work routine needs to be modified in a *third experiment phase* bringing an AI to the original production process (the one that was mentioned in Req. 4). For instance, a company introduces an AI to support a production process. When human test persons use this AI and perform the new production process several times, a new human work routine is implemented and a new routine behavior is established.
6. The new work routine (having AI support) will be eliminated due to arbitrary reasons, such as an AI breakdown, in a *fourth experiment phase*. By this, test persons are forced to come back to the original work routine (without AI support). Here, it is tested, to what extend the very same behavior and performance can be identified, as it has been observed in the second experiment phase. Deviations that occur here could be related to the fact that the standard or rather reference

work routine has been forgotten because too much reliance has been placed on the AI by the test persons.
7. As the research demands for providing different degrees of AI transparency, in particular, the *combination of multiple AIs* is attractive, because the cascade of AIs and step-wise AI-based decisions can be clarified (transparency condition). It needs to be checked, if these decisions lead to confusion if not clarified (intransparent condition).
8. As the research demands for providing different degrees of AI reliability, in particular, the reliable configuration of *multiple AIs* is attractive, because the cascade of AIs and step-wise AI-based decisions can be modified in regard with the experiment design effectively.
9. An *external service supplier*, such as *Prolific*, should (1) guarantee a representative selection of test persons, (2) realize secured and correct test person payment, (3) enable an ad hoc test person extension, such as for smaller pretests and extensive experiment runs.

4 Design of an Instrument for Examining Human and AI Cooperation

In accordance with the design-science cycle, the A3I artifact was developed. Its design is clarified subsection wise by clarifying the Design of Experiment (DoE) and relevant experiment variables (first subsection), its environmental experiment layout (second subsection), the production process that was simulated in the experiment (third subsection), the detailed experiment process (fourth subsection), as well as the software design in the fifth subsection. Please be aware that individual AI-based systems have been developed [13] and published as Open Source beforehand [15, 16]. These are included in this research article as *AI-CBR* and *AI-ANN*. The first cares about the most valuable ANN identification, which is a fruit type identification on the basis of order information, and the second cares about fruit type and quality detection on the basis of camera information.

4.1 Variable Operationalization and Design of Experiments

Aiming for examining human AI interactions and the role of forgetting at both, human knowledge carriers as well as AI-based knowledge carriers, the following three types of input variables are focused.

Transparency. The transparency of the AI is operationalized as the control by means of the explication level, which transparently explains the AI's behavior, its decisions or rather its justification why a certain outcome was generated by the AI. For the early research conducted, here, it refers to two forms, namely (1) *low transparency* representing the displaying of basic information, such as the concrete AI decisions, and (2) *high transparency* representing the displaying of additional information to characterize (a) the selected AI knowledge base (in case of the AI-CBR system for fruit type identification) and (b) the recognition results (in case of the AI-ANN system for fruit type and quality detection).

Autonomy. The autonomy of the AI is operationalized as the control by means of user integration. For the early research conducted, here, it refers to two forms, namely (1) *low autonomy* representing the process continuation with AI results only after user approval, and (2) *high autonomy* representing the process continuation with AI results after 10s.

Reliability. The reliability of the AI is operationalized as the control by means of consequent outcome correctness. For the early research conducted, here, it refers to two forms, namely (1) *low reliability* representing the process of flipping AI results by simulation control, so that, e.g., the identified *banana* fruit type (this might be correct or false) is changed to *orange* or *apple* or the identified *good* fruit quality (this might be correct or false, too) is changed to *bad*. Further, one can find the form (2) *high reliability* representing the process of keeping AI results as they are generated by the AI. Please be aware that this reliability interpretation does not guarantee to flip or remain correct AI results. To be keen on external validity, true AI results are to be used, here, instead of displaying by logic derived correct or incorrect results.

Mediating Variables. Beside these three variables, additional variables are surveyed, such as *trust in AI*, *AI frustration* or *AI acceptance*. In order to avoid the tendency toward the center, these have been operationalized by a caliper with 6 numerical units. As these variables characterize the AI utilization, these variables are planned to be considered as mediating variables on forgetting. Since their effect on forgetting is to be researched. the forgetting variable be considered as output or rather as dependent variable as follows.

Forgetting. The human forgetting because of forgetting AIs or rather AI-based CPS is operationalized as the unintended behavioral change of humans due to an AI's intervention, which mostly goes along with an observable performance decrease. Thus, the forgetting needs to be evaluated (comparable to a school test) in a well-defined task setting having an ideal or best solution. In the experiment case researched, this task and the corresponding performance decrease are observably by the click routine at a software-based work area, here referred to as AI-based CPS. As this kind of forgetting can be indicated by multiple attempts, the experimentation tool to be created in this research intends to offer multiple approaches for measuring forgetting at different experiment stages. Assuming to have a pure, software-supported production process being operated via the software-based work area, ...

- ...the forgetting can be operationalized by the *average number of clicks*—if it is higher in the fourth experiment phase (after test persons are used to the AI) than before the AI has been introduced (measured in the first experiment phase), a performance decrease can be detected, because the initially learned click routine was forgotten.

Interacting AI Systems	AI Reliability a	AI Reliability b	AI Reliability c	AI Reliability d
AI-CBR	*high*	*high*	*low*	*low*
AI-ANN	*high*	*low*	*high*	*low*

(a) Possible experiment runs w.r.t. AI reliability.

X

Transparency	Autonomy *low*	Autonomy *high*
high	Scenario 1 ↪ *1a, 1b, 1c, 1d*	Scenario 2 ↪ *2a, 2b, 2c, 2d*
low	Scenario 3 ↪ *3a, 3b, 3c, 3d*	Scenario 4 ↪ *4a, 4b, 4c, 4d*

(b) Prossible experiment runs w.r.t. AI autonomy and AI transparency.

Fig. 1 Derivation of experiment scenarios required

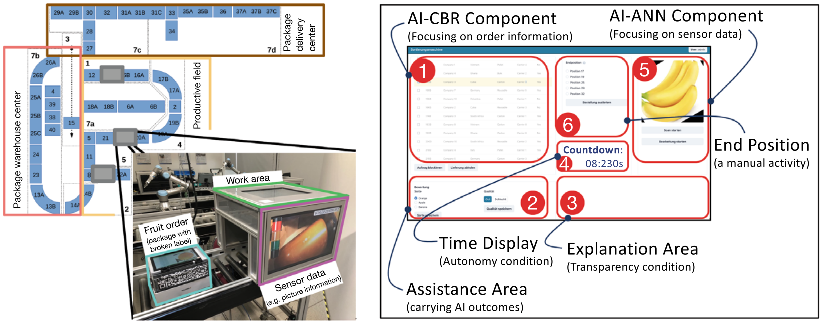
- ...the forgetting can be operationalized by the *click types* available—if more incorrect click type pattern are produced in the fourth experiment phase (after test persons are used to the AI) than before the AI has been introduced (measured in the first experiment phase), an indicator for the forgetting of the original click routine can be identified.
- ...the forgetting can be operationalized by the *average time* each click routine takes—if more time is required for producing click type pattern in the fourth experiment phase (after test persons are used to the AI) than before the AI has been introduced (measured in the first experiment phase), an indicator for the forgetting of the original click routine can be identified.

In addition to these initial attempts to measure forgetting, numerous further approaches are attractive to be researched. However, these three measuring attempts were selected as attractive first research step.

Deriving Possible Experiment Runs. Considering the three input variables focused in this research, Fig. 1 brings their forms together.

Figure 1(a) clarifies the four possible combinations of the *AI reliability* and the two types of AIs involved in the experiments, namely *AI-CBR* and *AI-ANN*. As these cooperate with the test person, its detailed interaction design will be clarified in Sect. 4.3. For realizing the experiments described, in total 5.760 AIs have been developed and released. Further information about the AI designs, the systems trained and published for being re-used can be found at the software repository called *AI-CPS* and corresponding public Docker's Hub images and publications [13]. Fig. 1(b) clarifies the four possible combinations of the *AI transparency* and *AI autonomy*. Having a look on the cross-product of these three variables, the twelve possible combinations of relevant experiment runs can be identified. These are labeled with 1a-1d, 2a-2d, 3a-3d and 4a-4d.

Experiment Planning. The minimum number of required experiment runs was determined as follows: Although we were interested in the multiple R2 value in the first step and a sample size of N >= 50 + 8(k) was sufficient in accordance with [11], we were interested in beta weights in the second step. According to Green, this required that N >= 104 + k. Here, N stands for the sample size and k for the number of independent variables. So, assuming to observe a medium effect of the



(a) Physical experiment environment.

(b) Local and remote work area.

Fig. 2 Experiment environment design.

independent variables on the dependent variable (R2=0.7 and beta=0.2), we aimed to complete at least 107 experiments (focusing on beta weights). In order to achieve a full factorial experiment design and bivariate variables, 12 combination possibilities required about 9 experiments per combination (107/12≈8.92), which in total amounted to about 108 experiment runs. As test subjects were supposed to participate in the experiment once, the experiment had a between-subject design in regard to the experiment conditions presented [27].

4.2 Experiment Environment Design

Since experiments were planned to be realized by the Industry 4.0 environment laboratory of Potsdam University, titled with *Zentrum Industrie 4.0 Potsdam (ZIP4.0)*, its current layout was extended. Fig. 2(a) clarifies this by presenting its spatial cyber-physical system arrangement with blue rectangles. As former experiments have shown, it is important to enable several test persons in parallel. So, the layout was designed to realize three test persons in parallel (see gray rectangles in the figure). Each of the three workplaces so was designed to provide the same components—these can be seen in the zoomed picture of Fig. 2(a).

All together, Fig. 2(a) clarifies that the experiment environment of each test person was designed to have two types of components: The first type refers to *the individual working machine*, which provides private components only accessible by the individual test person. The second type refers to *the joint experiment environment*, that provides public components that are accessibly by all test persons. Here, all of the individual test person environments are at an intersection. The corresponding five concrete components are clarified in the following.

- **Package warehouse field:** *(see orange color)*
  It is part of the joint experiment environment and provides the packages of fruit orders for all test persons. Since an order only can be processed by one test person, a certain complexity and coordination effort comes up for the individual test persons and the joint experiment environment.
- **Productive field:** *(see yellow color)*
  It is part of the joint experiment environment and provides the three physical and local workplaces for the in-place experiments.
- **Package delivery field:** *(see violet color)*
  It is part of the joint experiment environment and receives the packages of fruit orders of all test persons as these have been processed. Since an order can have different states, such as different fruit types and quality levels, various specialized end positions can be headed. Thus, an additional complexity and coordination effort comes up for the individual test persons and the joint experiment environment.
- **Fruit order:** *(see turquoise color)*
  The fruit packages ordered from the warehouse are processed by local test persons at the physical instance of the individual working machine. They are transported from the warehouse to the work area via the conveyor belt system.
- **Sensor data:** *(see purple color)*
  The fruit packages arriving are scanned and photographed multiple times at the individual working machine. This is part of the production process being described in Sect. 4.3.
- **Work area:** *(see green color)*
  The individual working machines of local as well as remote test persons presents the same, software-based work area.

As the work area can be considered as the key instrument to provide human and AI interaction possibilities and measure forgetting of human test persons, its detailed components are displayed in Fig. 2(b). These clarified in the following. Here, the numbered list entries refer to the numbers being displayed in the figure.

1. **AI-CBR Component:** *(for managing orders)*
   The orders of the central warehouse are displayed and requested via the first area.
2. **Assistance Area:** *(for inserting human and AI results)*
   If AIs support the production process, their decisions are displayed in the assistance area (see highlighted second area). If the AIs are not part of the production process, the very same decisions are to be carried out by the human test persons.
3. **Explanation Area:** *(for clarifying AI decisions)*
   The third area clarifies the AI decisions if these are to be made transparent (transparency condition). In the non-transparent condition, this area remains empty.
4. **Countdown:** *(human chance to intervene AIs)*
   In the fourth area, the remaining time is displayed, if an AI is allowed to act autonomously in the autonomy condition. Human test persons have a chance

to intervene as long as the countdown is active. If the time is run out, the AI proceeds automatically. In the non-autonomous condition, human test persons need to confirm each AI outcome manually. As the latter condition has no time pressure, this remains empty.

5. **AI-ANN Component:** *(presenting pictures)*
   The fifth area shows visual package information, which is taken by a video camera. So, for instance, test persons can watch the package arriving. These live insights drastically increase the immersion for remote test persons. Further, this area displays pictures from the package' content, such as nice fruit or rotten fruit deliveries.
6. **End Position:** *(for instructing package paths)*
   When the productive activities, namely the *fruit type identification* and its *quality assessment*, are finalized (with or without AI support), test persons are supposed to instruct relevant conveyor belts to send the package to one of many specialized end positions.

4.3 The Production Process as Industry 4.0 Application Context

In one experiment run, test persons do have to process a pre-defined number of orders per experiment phase. One order is processed by realizing the production process once. But what is the production process about? The following describes the simulated scenario each test person is faced at the experiment. It makes plausible the process context in which the A3I brings test subjects into interaction with AI.

Simulation Context. In general, the Industry 4.0 production process refers to the *trouble shooting procedure* or rather second level support (in regard with ITIL) of packages having a damaged QR code. If the QR labeling is damaged, the digital and automated package delivery machinery cannot handle the corresponding packages and the Industry 4.0. logistic system is disturbed. However, the human worker or rather test person tries to identify the type of fruit being delivered on the basis of order information, such as the supplier details, information about the delivery's origin, or contracted forwarders. Then, the human worker tries to assess the fruit quality based on a look on the fruits. Finally, based on the previous two tasks, the human worker picks the correct end position, so that the digital and automated package delivery machinery can proceed. Among the set of end position, i.e., one position represents banana packages having a certain cooling requirement. Other end positions do differ in cooling requirements. For instance, apple deliveries do need more space and air circulation but they do not need a cooling. Further end positions store rotten fruit.

Production Process. Figure 3 presents the detailed process model as UML's activity diagram It clarifies the processing of one trouble case and is explained in the following.

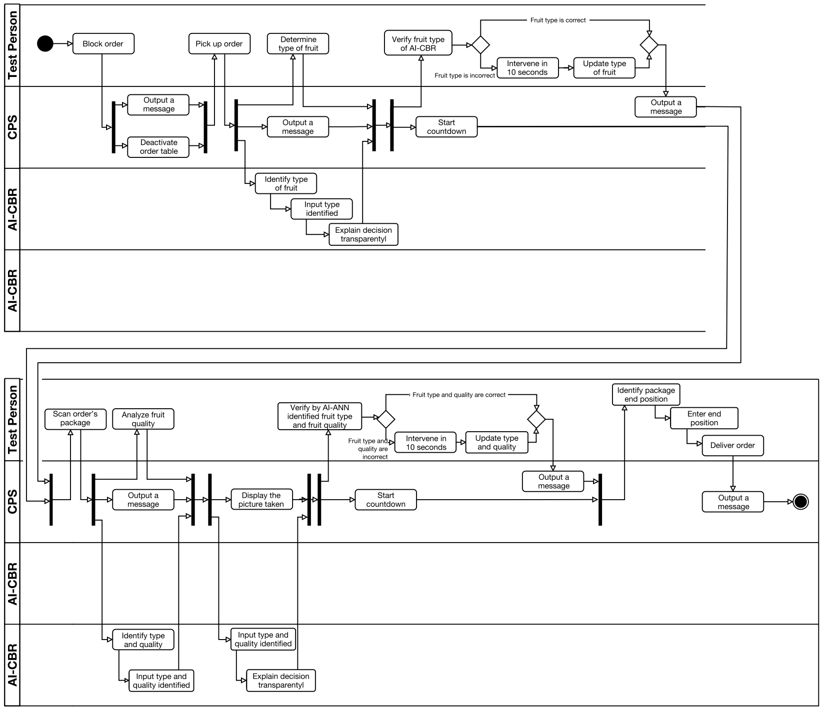


Fig. 3 Production process being realized in experiments multiple times per test person

In the figure, one can see four horizontal swim lanes. Each group activities for the (1) human *test person*, (2) the *cyber-physical system* providing the work area for functioning as human-machine interface (compare with Fig. 2), (3) the AI using CBR-based mechanisms for identifying the fruit type on the basis of order information only, and (4) the AI using ANN-based mechanisms for identifying the fruit type and assessing the fruit's quality on the basis of camera information.

The process model clarifies the **first activity** of the test person. It needs to block the corresponding order and initiate the package's pick up, so that the conveyor system transports the corresponding order package from the warehouse to the work area. The CPS displays corresponding system messages.

As **second activity**, the fruit type needs to be identified, which will be overtaken or rather supported by the AIs in selected simulation scenarios. If an AI has supported the fruit identification, the fruit type recognized needs to be verified by the human test person. Depending on the current scenario, either this verification needs to be realized within 10 seconds, or there is an unlimited amount of time for confirming the AIs decisions (autonomy condition). The CPS functions as the central control instance and displays relevant system messages.

Then, in a **third activity**, the order needs to be scanned and analyzed for fruit quality detection. If an AI has supported the quality identification, the quality level recognized needs to be verified by the human test person, again. Depending on the current scenario, either this verification needs to be realized within 10 seconds, or there is an unlimited amount of time for confirming the AIs decisions (autonomy condition). The CPS functions as the central control instance caring about timings. Further, it displays relevant system messages and functions as HMI.

Finally, in the **last activity**, the package's end position needs to by identified and entered manually by the human test person. On the basis of this selection, relevant conveyors are activated (compare with blue rectangles of Fig. 2a) to transport the test person's package from the work area (compare with gray rectangles of Fig. 2a) to the correct place at the delivery center (compare with violet rectangle of Fig. 2a). Again, the CPS functions as the central control instance and HMI. The production process is finished by an output message of the CPS.

Since the process model shows the activity of both AIs, the interaction of the AIs can be clarified: While the first system called *AI-CBR* cares about the identification of the fruit type on the basis of order information, an ANN model will be selected that is specialized on the fruit type identified. For verification, the second system called AI-ANN will (1) re-evaluate the fruit type and (2) assess the quality of the fruit on the basis of camera information. Due to this chained AI utilization, mistakes of the first AI might lead to reasonable mistakes of the second AI—at least if the specialized AI does not know the corresponding fruit type. This kind of variety enables valuable debugging information and transparent decision clarification (transparency condition).

4.4 Experimental Setting for Researching Forgetting Systems

The experimental setting clarifies, how the production process (cf. Sect. 4.3) will be used to enable human and AI interaction modeled. But how is human forgetting because of forgetting AIs modeled? The following describes the well-designed sequence of simulated scenarios each test person is faced at the experiment realization. It makes plausible the experiment phases (compare with requirement collection of Sect. 3) and describes the transformation of business process contexts in which the A3I brings test subjects into interaction with AI under varying circumstances.

Course of the Experiment. For modeling the knowledge flows and interactions among human and AI-based process participants, the Neuronal Modeling and Description Language (NMDL) has been selected as superior candidate for this purpose [14]. Using NMDL's *ProcessView*, Fig. 4 clarifies the sequence of experiment tasks and phases. While some phases and tasks are the same for all scenarios, some show content depending on the scenario selected for the current experiment run, which will be explained in the following.

In the figure, one can see the sequence of tasks that have been visualized by green rectangles and have been grouped to phases visualized by white rectangles. These have been connected by arrows representing the control or rather task or phase flow.

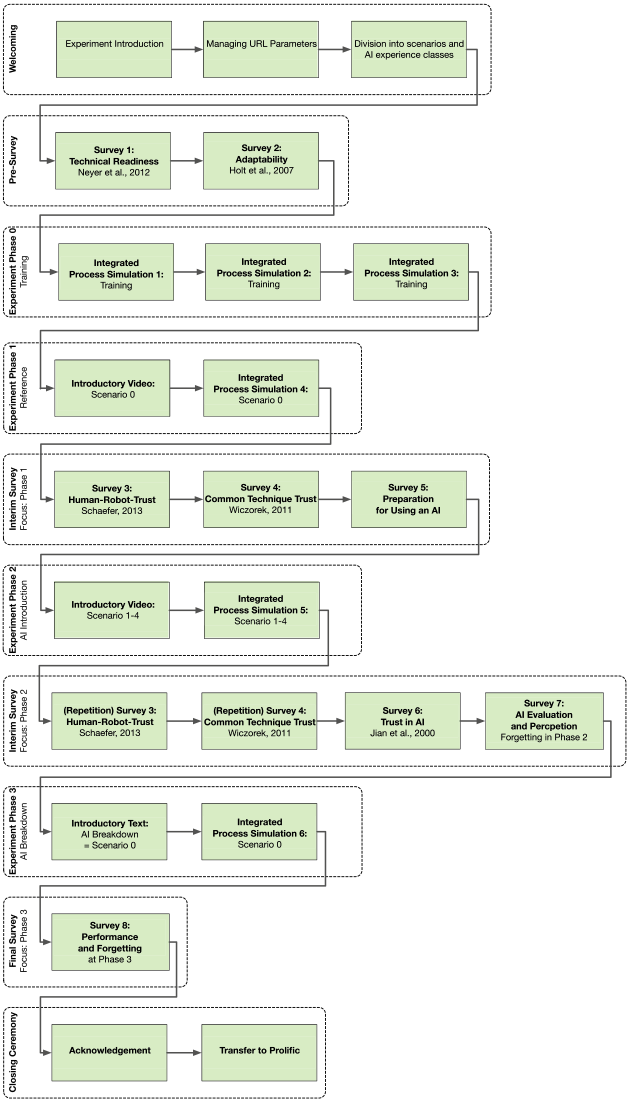


Fig. 4 Experiment process

Dashed lines cluster scenarios w.r.t. their role in the experiment. Thus, these are clarified in the following.

The **welcoming phase** guaranteed the warm welcoming of test persons. W.r.t. test person management, relevant *URL parameters* were set up by background mechanisms (cf. Req. 9), so that a representative test person set was created and the required experiment runs were started. The experiment implementation operationalized the selection of the current experiment condition by random variables according to the

Design of Experiment (compare with Fig. 1) and the experiment planning. These were mapped to test person-specific URL parameters, then.

In the **pre-survey phase**, test persons were faced with questions by [34] to characterize the initial level of *technical readiness*. Further, questions by [21] were presented to conduct the test person's *adaptability* These were presented to research the general technical attitude of a test person to have an effect on forgetting because of AIs.

Four **experiment phases** were realized then. Each was composed by the same structure: First, an introduction to the productive process of the current scenario was given that was to be realized by the test person next. This referred to the process that was designed in (Sect. 4.3) and realized under the experiment condition (compare with Fig. 1). Second, the production process was realized numerous times by using the software-based working area and corresponding environment as they were designed in Sect. 4.2. Here, relevant performance metrics were used to measure the test person's performance and enable the identification of forgetting.

The forgetting measurement was realized by providing experiment phases for multiple times. Each time has had a different role. The *Experiment Phase 0* enabled the training in three separate (equivalent) sessions, so that the test persons got familiar with the productive process. Here, failures were welcome and additional explanations were presented in form of a tutorial. The *Experiment Phase 1* enabled the training realization under experiment conditions. It was realized on the experiment day and was used for measuring the performance baseline. Please be aware of the following: Up to here, only manual work was realized by test persons. There was no AI support, yet. An AI support was integrated in *Experiment Phase 2*. Here, the behavioral change and performance change were measured, so that deviations from the baseline or reference situation could be derived. A particular analysis focus was set on the identification of a performance improvement because of AI support. The *Experiment Phase 3* forced the AI removal, so that test persons were supposed to return to the manual process. As the behavioral change and performance change were measured, analyses could focus on a comparison of the behavior observed at this phase and the reference phase. This comparison should lead to insights on the relevant process of forgetting because of an AI.

Each experiment phase was followed by an individual **interim/final survey phase**. Since *Experiment Phase 1* only used robots as non-AI-based systems, questions focused on *human robot trust* according to Schaefer [45] and *common technique trust* according to Wiczorek [51]. Since the *Experiment Phase 2* and the scenarios 1-4 integrated AIs, questions additionally taken focused on *trust in AI* according to Jian et al. [25] and *AI perception*. The questions after AI breakdown of *Experiment Phase 3* focused on the test person's perception of *performance and forgetting*.

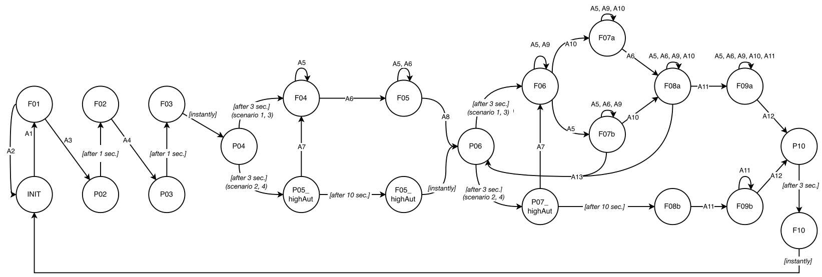


Fig. 5 Software states

4.5 *Software Design for Human and AI Interactions*

Since the experiment was designed to be used in several scenarios (cf. the experiment conditions of Fig. 1), and the productive process (cf. Fig. 3) was to be used in several phases within the experiment (cf. Fig. 4), a software application has been based on a *state machine design* that is presented in a first subsection. It integrates standard components and individual implementations, that have been presented in the second subsection. Examples refer to the well-known survey tool called *LimeSurvey*, the external service provider for test person management called *Prolific*, and a collection of programming libraries being used for individually implemented software components.

The Design as State Machine. The state machine of the software-based application can be found in Fig. 5. While the *states* of this machine map the current progress of the productive process (cf. Sect. 4.3), the *transitions* from state to state are either realized by the test person's interactions with the software's work area or by timed conditions, such as the expiry of 10 seconds. In a pretest, this period was deemed sufficient to be able to intervene in an AI decision. However, it implements the design of the *AI's autonomy* (compare with autonomy condition issued in Sect. 4.1). Please find these states as circles and actions as directed arcs in Fig. 5. The corresponding state and action descriptions can be found in Table 1.

Data Model Design. For realizing the A3I artifact, a combination of numerous components with individual data sets was set up. The integrative data model is shown in Fig. 6. It is explained in the following.

For acquiring, selecting and paying test persons, the external service provider *Prolific* was chosen. At this platform, test persons have been acquired as long as a pre-defined amount of money is payed. With Prolific's support, a representative test person set is set up and its demographical data is collected (cf. on the left side of Fig. 6). Each test person selected for experimentation is equipped with an individual *prolificID*, which is transferred to the *LimeSurvey* tool via an URL-based parameter.

Commonly, LimeSurvey is used as a tool for realizing simple web-based surveys. However, since it enables the modeling of complex control flows and conditioned survey elements, this tool has been extended to provide a web-based platform for

Table 1 Collection of the state machine's states and actions

ID	Description	Type
INIT	A new job has been started	State
F01	A job has been selected	State
P02	The job is blocked	State
F03	The order is blocked	State
P04	The AI-CBR evaluates the fruit type	State
F04	A fruit type has been selected	State
F05	A fruit type has been saved	State
P05HA	The countdown is active	State
F05HA	The countdown has expired without intervention and the fruit type selected by the Ai-CBR has been saved	State
P06	The scan is active, the AI-ANN evaluates the fruit type and fruit quality	State
F06	A fruit quality and fruit type has been selected	State
F07a	A fruit quality has been saved	State
F07b	A fruit type was saved	State
F08a	A fruit quality and fruit type has been saved	State
P07HA	The countdown is active	State
F08b	The countdown has expired without intervention and the fruit type and quality selected by the AI-ANN has been saved	State
F09a	An end position has been selected	State
F09b	A final position has been selected	State
P10	The order is being delivered	State
F10	The order has been delivered	State
A1	The test person selects an order	Action
A2	The test person deselects the selected order	Action
A3	The test person blocks the selected order	Action
A4	The test person picks up the blocked order	Action
A5	The test person selects a type of fruit	Action
A6	The test person saves the currently selected type of fruit	Action
A7	The test person stops the active countdown by intervention	Action
A8	The test person starts the scan	Action
A9	The test person selects a fruit quality	Action
A10	The test person saves the currently selected fruit quality	Action
A11	The test person selects an end position	Action
A12	The test person delivers the order	Action
A13	The test person starts a new scan	Action

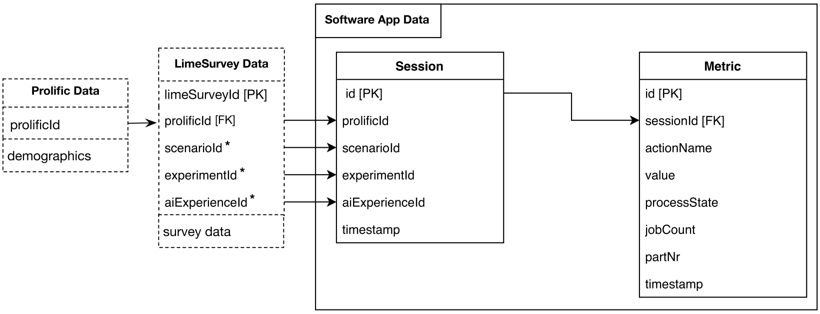


Fig. 6 Data model for integrating Prolific, LimeSurvey and further software metrics of local CPS docker containers

providing arbitrary complex survey elements according to the experimentation of forgetting (cf. Fig. 4). Prolific data and LimeSurvey data are integrated by transferring the *prolificID* to LimeSurvey via URL parameter. Here, it is stored as external variable and used as *foreign key* next to LimeSurvey's individual *primary key*.

For assigning test persons to the relevant experiment conditions, the variables *scenarioID* and *aiExperienceId* are randomized according to the experiment design (cf. the four numbered scenarios at Fig. 1a and the four with letters labeled conditions at Fig. 1b). These are initialized when the LimeSurvey-controlled experiment is started via the web. So, the DoE was implemented and controlled at the LimeSurvey.

Following the experiment process (cf. Fig. 4), the LimeSurvey tool controls the provision of the software-based work area (cf. Fig. 2) and its manifestation, such as having an AI support or not. Thus, it implements the current experiment phase, which is indicated by the variable called *experimentId*. Since the LimeSurvey tool routes relevant variables to the software-based work area, the software-based work area can be embedded in LimeSurvey as required by the current experiment condition. Further variables to fully characterize the joint experiment setting are presented at Table 2.

The test person's performance and work pattern data is gathered via the software-based work area software as follows: The software-based work area generates an event by each interaction with the test person. By linking the current session with its *sessinID* as foreign key, the test person's metric data can be implemented. For instance, this refers to a click on a certain button, the insertion of a user's decision or its inactivity. This implements the forgetting operationalization presented in Sect. 4.1. Based on the data collected, here, the test person's performance or rather software use pattern can be analyzed, such as a correct or false click routines, required time for realizing activities, correctness and efficiency of realizing the production process (cf. Fig. 3). Further, it can be analyzed to what extend test persons were able to restore learned processes by comparing work patterns and performances at different experiment phases (compare with the requirements presented in Sect. 3).

Table 2 Collection of URL parameters to configure the current experiment run

URL Parameter	Description	Default
timeout	It controls the time in minutes until the experiment is continued	6
maxJobs	It controls the number of tasks the test person must complete before the experiment continues	6
debug	It controls whether debug parameters are displayed	FALSE
scenarioId	It controls which AI scenario the app will implement. It ranges from 1 to 4	1
prolificId	If any string is specified, a session is created in the backend and metrics are activated. This parameter is generated by the external service provider called *Prolific*	none
aiExperienceId	It controls which AI experience the app will implement. It ranges from 1 to 4	1
experimentId	It indicates the part of the experiment that currently uses the app. It is linked to the backend session. Further, it is only valid with prolificId	none

Processing Node Design. To implement the data model described, and enable both, the local as well as the remote experimentation (cf. Req. 2), Fig. 7 clarifies the instances involved.

The test person takes the experiment from its individual *Webclient* or Internet browser of choice (see on the bottom right of the figure). Here, local test persons use the browsers provided at the individual machine's local work area (compare with Fig. 2a) and remote test persons use the browsers of an individual device of choice, such as their laptop, desktop or smartphone.

Relevant experiment data are provided via the Prolific's *Test Person Management*, which provides the relevant *URL of Experiment* if a test person is part of the representative test person set composed. The *LimeSurvey Controller* is requested by a HTTP(S) request via an *Apache2 Webserver*. The controller mentioned manages the access to the original *LimeSurvey Tool* whose data is stored in a *MySQL Data Base Management System (DBMS)*.

The LimeSurvey tool implemented the experiment control flow (cf. Sect. 4.4). Here, the specified software-based work area is provided as container at relevant LimeSurvey pages. If requested at LimeSurvey, the corresponding container stack is requested at an *Apache2 Webserver* via HTTP(S), which redirects to the *Web-Based Work Area*. The latter is provided via a virtual machine instance that runs a *Docker Compose Container Network* (cf. Fig. 2). As the app's components are containerized by node-independent Docker images, the app can be run on numerous systems having a different hardware characteristics.

As Fig. 7 clarifies, the web application implements numerous components for running the *Web-Based Work Area*. For instance, the NGINX-based *Ingress Controller* is responsible for making the ports of the containers available via different

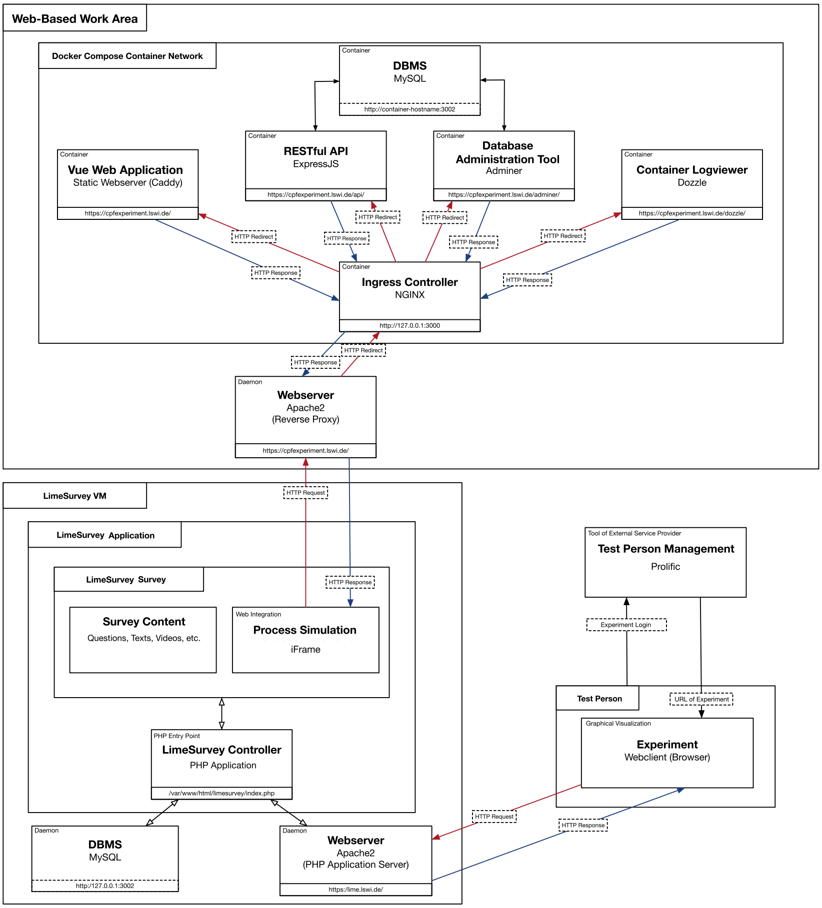


Fig. 7 Software-based system architecture for node-independent device configuration

URL paths. Each container has its own path through which a request can be correctly assigned to a container.

The *Vue Web Application* represents the graphical user interface (GUI) of the work area in the production process (cf. Sect. 4.3). It is responsible for making the interactive production process available to the test persons via the browser. The developed web application was compiled into a static website using the functions of VueJS. This procedure optimizes the web application for efficient use in the browser. The *Caddy* tool is used to provide this optimized version.

The *REST API* is responsible for providing the front end with an interface to the database and the functions of the AIs. On the other hand, it provides static content. This includes the order data and the fruit images of the orders.

The *MySQ DBMS* is responsible for storing and managing the data collected in the simulations. All database operations can only be carried out by other components via the REST API or the *Database Administration Tool*, which is in this case the *Adminer* tool.

The *Container Logviewer* Dozzle is another tool that facilitates the development process, as it makes information about the installed containers within the DCCN and the log entries of these containers available via the browser.

5 Demonstrating the AI Interaction Instrument

In the demonstration phase of the DSRM, the A3I artifact was brought to reality—it was used in a real experiment realization, which contains all, two *experiment master* perspectives and one *experiment participant* perspective. Each is clarified subsection wise.

Master Perspective on Remote Experiments. The first perspective is related to the researchers using the A3I as research tool to realize remote experiments. These parties are responsible for compiling a representative list of online experiment test persons, managing their access and realizing test person payment. The operational participant recruitment is realized via Prolific by following the researcher's strategic requirements on test persons: Among dozens of attractive filter criteria, such as demographics (age, sex, geographic information), languages (first language, fluent languages, primary language), education and work, only German as fluent language has been selected as requirement. This has lead to a test person set with about 9.135 attractive test persons. Further information, that was to be specified by the researchers, referred to the average reward per hour. This was determined to be £9.52, since it was at the top level of Prolific's recommendation. Having realized a pretest with about 40 test persons, the median time of 60 minutes was determined and entered in the Prolific survey specification. Further, the AI3 was fine-tuned due to pretest feedback. Finally, the maximum number of 150 possible test subjects was determined with the available total budget. When having specified the experiment's requirements and having started the experiment, Prolific has sent attractive test persons to the A3I artifact, which enabled the parallel experimentation via the following hyperlink:https://lime.lswi.de/index.php/225396?lang=en-informal&prolificId=MaxMustermann_ID&prolificStudy=MaxMustermann_Study&prolificSession=MaxMustermann_Session.

At Prolific, participating test subjects were shown live in an overview so that each individual test subject could be approved for payment, rejected due to invalid experiment execution, approved for re-execution or automatically deregistered by Prolific (see Fig. 8). Examples of an invalid test execution are the idling of the test, the changing of experiment workstation or having a lunch break instead of concentrating on the execution of the production process because, for example, the time measurement was disturbed.

Master Perspective on Local Experiments. The second perspective is related to the researchers using the A3I as research tool to realize local experiments in Potsdam's ZIP4.0. Although these managing parties have the very same responsibilities

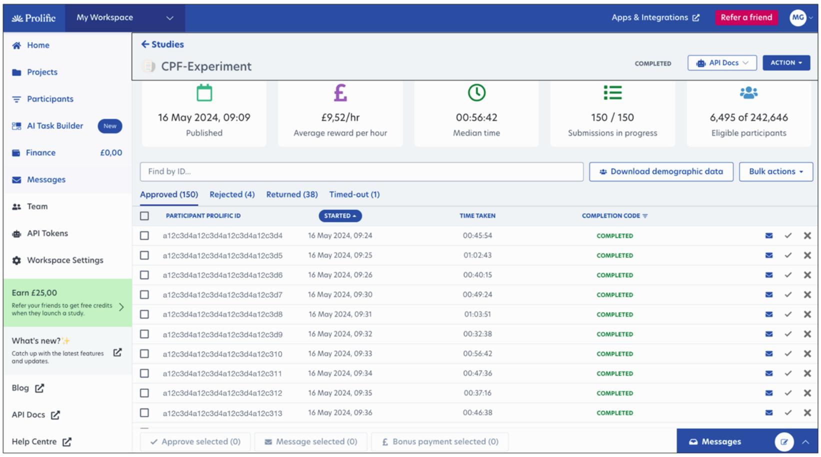
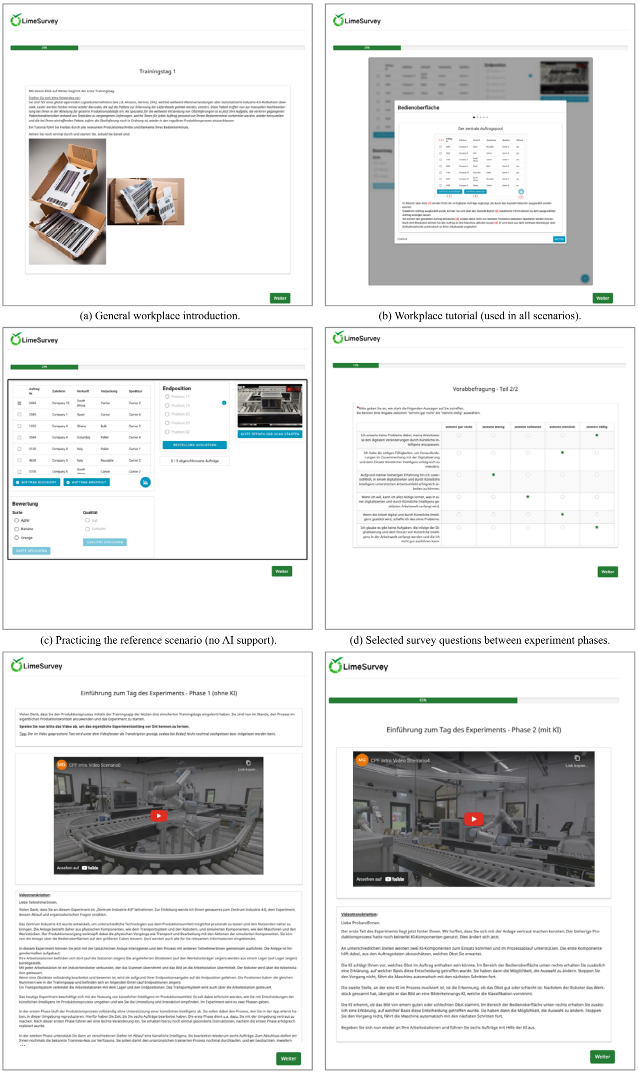


Fig. 8 Managing access to the A3I experimentation tool

as shown in the remote experimentation, namely compiling a representative list of local experiment test persons, managing their access and realizing test person payment, the access provisioning is realized via emails manually. Here, the very same hyperlink was used.

Participant Perspective. The third perspective is related to the test persons using the A3I as research tool to be part of the experiment. These parties are payed for realizing the production processes provided by the experiment. Figures 9 and 10 clarify extracted test person views that are displayed in individual test person browsers when the previously mentioned hyperlink was used. Their sub-figures are clarified in the following.

Figure 9a provides the general scenario context: test persons are part of the second level support team caring about packages with broken QR code in a logistic center. In a step-by-step routine, participants are taught to deal with the work area in a tutorial (see Fig. 9b) in *Experiment Phase 0*. Please note, the experiment phase labeling is in accordance with the designed course of experiment (see Fig. 4). However, Fig. 9c shows the work area when participants were faced without AI support (as reference in *Experiment Phase 1* as well as in *Experiment Phase 3* as AI breakdown). Before *Experiment Phase 1*, this work area was trained and the reference process was practiced. Before *Experiment Phase 3*, the very same work area was augmented with AI support. Between the experiment phases, relevant information has been surveyed, as Fig. 9d shows. After having trained the production process, the reference production process (*Experiment Phase 1*) was introduced by a video (see Fig. 9e). As Fig. 9f shows, the integration of an AI to the production process (*Experiment Phase 2*) was introduced by a video, too. The video transcripts were shown below the videos, so that test persons were enabled to read the text, hear the text and see relevant items. This should appeal to different types of learners.

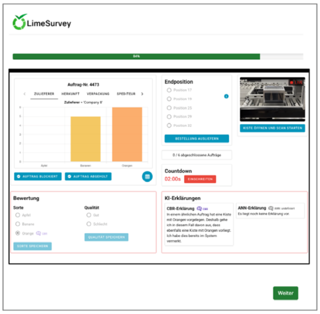


(a) General workplace introduction.
(b) Workplace tutorial (used in all scenarios).
(c) Practicing the reference scenario (no AI support).
(d) Selected survey questions between experiment phases.
(e) Introduction to scenario 0, the reference scenario (without AI support).
(f) Introduction to scenario 1/2/3/4, the AI-supported scenarios.

Fig. 9 A3I experimentation instrument (part 1)

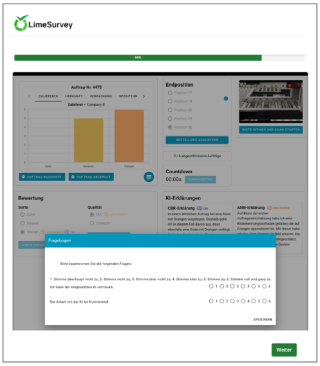
Fig. 10 presents further details of the test person view. Their sub-figures are clarified in the following:

In the *Experiment Phase 2*, the AI-CBR system was visualized as shown in Fig. 10g. Since the AI shows an explanation on the bottom right (transparency condition) and a countdown is indicating the AI's autonomous proceeding on the center right (autonomy condition), this figure shows the most complex visualization at sce-



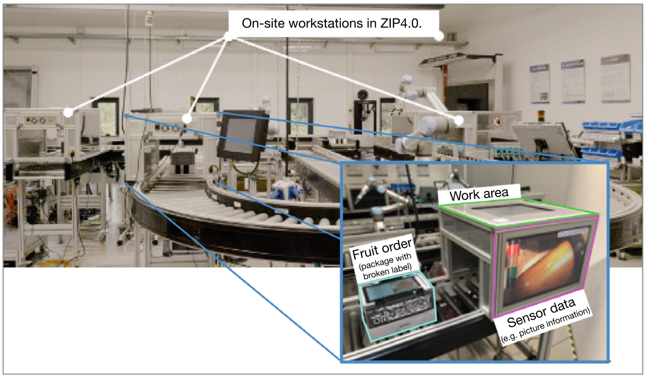
(g) AI support at fruit type identification (using AI-CBR).

(h) AI support at fruit quality identification (using AI-ANN).



(i) Surveying the AI perception during the working session.

(j) Introduction to AI breakdown (scenario without AI support).



(k) Integration of physical on-site and virtual remote workstations to enable local and online experimentation.

Fig. 10 A3I experimentation instrument (part 2)

nario 4. After the AI-CBR system has finished working, the AI-ANN system starts working, which is shown in Fig. 10h. Thus, the ANN's explanation is displayed in addition. After having finalized the productive process, the test person's intermediate impression was conducted directly via the work area (see Fig. 10i). This enabled the analysis of the course of *AI trust* and *AI frustration*. Fig. 10j shows, how the AI breakdown scenario was introduced (*Experiment Phase 3*). One AI breakdown has been explained as maintenance work. The on-site workspace of three test persons is shown in Fig. 10k. While the work area has been the same for remote and on-site test persons, physical movements of the current fruit order and corresponding sensor data (pictures taken from the current fruit delivered) have been displayed at the video screen on the top right of the work area. This enabled remote and on-site test persons to have the same information. A pretest indicated this to be essential for remote workers.

6 Discussion About A3I Instrument-Based Experiments

As last phase of the design-science-oriented research cycle, the DSRM envisages the evaluation of the A3I artifact and clarification of the research contribution. This addresses the question, to what extend the artifact has proven to be useful in the demonstration case and goes along with communicating the meaning of research insights.

Evaluation of Requirement Fulfillment. In a first evaluation step, the requirements that have been collected before the A3I artifact design was set up are checked. For this, the following clarifies the requirement fulfillment by using the same requirement numbering as Sect. 3 shows.

Req. 1 (*node-independent device configuration*) can be satisfied, because the (1) software-based work area, the (2) AI-CBR system, the (3) AI-ANN system, the (4) survey tool LimeSurvey, as well as (5) their required sub-systems have been set up as a Docker containers, which can be initialized and run on any computing device, such as laptops, raspberries and servers. This clarifies the *node-independent experiment provision* and its software required for providing the experiment. As the experiment simply can be accessed via one hyperlink carrying the relevant external information as URL parameters, the *node-independent participation* can be realized on any computing device, such as laptops, raspberries and desktops.

Req. 2 (*online and on-site experiment realization*) can be fulfilled, because test persons can either participate at the experiment *remotely* via their laptop, smartphone or desktop machine. Alternatively, test persons can participate at the experiment *locally* in ZIP4.0, because the very same experiment tools are provided at its processing cubes—the so-called machines (compare with Fig. 10k). Req. 3 (*first experiment phase*) can be satisfied, because three repetitive training sessions have enabled the pedagogic valuable learning of the *production process* by experiment participants. (cf. *Experiment Phase 0* in Fig. 4) Since the real productive environment was provided in an extra experiment phase (cf. *Experiment Phase 1* in Fig. 4), Req. 4 (*second experiment phase*) has been fulfilled. The variation of the standard production process is the realized by a following experiment phase (cf. *Experiment Phase 2* in Fig. 4). Here, the AIs are provided in addition, so that Req. 5 *third experiment phase* has been satisfied.

Req. 6 *fourth experiment phase* has been fulfilled, because the experiment has provided the original production process phase after making test persons comfortable in working with AIs (cf. *Experiment Phase 3* in Fig. 4). Req. 7 (*combination of multiple AIs*) is satisfied by the A3I artifact due to the provision of two cooperating AIs, namely (1) the AI-CBR and (2) the AI-ANN. Since the AI-ANN builds on the AI-CBR decision, failure cascades can either confuse human workers if not clarified (intransparent condition) or can make AI behavior plausible if clarified (transparency condition). Here, each AI was designed to be reliable or not. So, Req. 8 has been fulfilled, too. Since Prolific has been used for setting a representative selection of test persons, realizing test person payment and managing the number of total experiment participants ad hoc, Req. 9 (*external service provider*) has been fulfilled.

Research Contributions and Practicability for Practitioners. The article has presented a software-based work area design for *manifesting different kinds of AI autonomy*. This was exemplified in Industry 4.0 context and one production process. A software vendor could bring the very same AI autonomy operationalization to alternative production processes or to software of different domains. For instance, an AI-based trading system could come up with invest options and carry them out autonomously, if a human expert does not intervene in a certain period of time.

Further, the article has presented a design for a work area manifesting *different kinds of AI transparency*. This was exemplified with two types of cooperating AIs, namely an AI-CBR and an AI-ANN system. The very same kinds of transparency increasing explanations can be provided by AI-supported software. For instance, a software vendor could display an AI's reasons of coming up with a certain decision to increase trust of its users in the cooperating AI provided via the software's work area. This might reduce AI frustration in addition.

The article has not only examined a software-based work area providing different manifestations of cooperating AIs interacting with human workers. This new kind of AI-supported work area has been embedded in an experiment proceeding that enables the examination of different experiment conditions, influence variables, AI manifestations and AI interaction sets. For instance, an ERP supplier could use the very same experiment routines and survey tools developed to identify the end user's trust and frustration levels when facing the user with an AI-based ERP system. So, AIs having different ways to interact with an user at the ERP's selected work areas can be analyzed with statistical instruments. Best ways to design cooperating AIs in ERP systems can be identified by this.

Coming back to the original objective of intended studies, namely the exploratory examination of evolving AI on human forgetting and vice versa. The A3I has been proven to be an adequate tool to carry out experiments on evolving AIs, forgetting humans and transforming business environments. Each of which can be designed explicitly and integrated into the A3I. As its collects any kind of data specified in advanced and provided as survey question, it so suits as an adequate experiment tool to research quantified effects of the socio-economical team of AI, human and their cooperation context.

Limitations. Although requirements on the A3I research artifact have been fulfilled, the demonstration proofed the technical functioning of the A3I research instrument and its complex components, only. To what extend this artifact is useful to measure the quantified effects of forgetful AI-based CPS on forgetful people, which corresponds to the initial research question, has not been addressed in this article. However, data has been gathered and its analyses are published next.

Outlook. The literature shows the importance in designing AI autonomy and transparency levels carefully. Thus, in particular, researching the AI design, its manifestations and ways to interact with software users are very meaningful. With the A3I developed, a tool for systematically researching these objects of investigation has been provided. So, after having completed data analysis on the Industry 4.0 context provided in this article, and being able to quantify the effect of forgetful AI-CPS on forgetful people, the A3I re-use and adaptation to further production processes, software contexts and alternative domains are attractive.

In particular, once effects of forgetful AI-CPS on forgetful people have been identified, the way of integrating learning and forgetting AIs to software can be recognized as instrument to guide and manage learning and forgetting human workers so that the joint team of human and AIs are aligned to business needs. A second research focus on their purpose-related design should be established, then.

Acknowledgements The author wishes to thank C. Thim, N. Gronau, W.M. Roling and A. Kluge for the valuable feedback on process models, J. Gonnamann for the introduction video recording used in experiments and N. Franz for the valuable support at implementation. The scientific project is sponsored by the German Research Foundation DFG (ID GR 1846/21-2), the Federal Ministry for Economic Affairs and Climate Protection based on a resolution of the German Bundestag (BMWK: KK5272103MS3) and the Gebauer GmbH bringing AI to ERP systems.

References

1. Altmann, E. M., & Gray, W. D. (2002). Forgetting to remember: The functional relationship of decay and interference. *Psychological Science, 13*(1), 27–33. https://doi.org/10.1111/1467-9280.00405
2. Bjork, R. (1998). *Intentional forgetting in perspective: Comments, conjectures, and some directed remembering* (pp. 453–481). Lawrence Erlbaum.
3. Bühner, M. (2011). *Einführung in die Test-und Fragebogenkonstruktion* (Vol. 4033). Pearson Deutschland GmbH.
4. Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. Y. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing, 16*(6), 1190–1208. https://www.researchgate.net/publication/2837734_A_Limited_Memory_Algorithm_for_Bound_Constrained_Optimization
5. Cappellani, F., Card, K. R., Shields, C. L., Pulido, J. S., & Haller, J. A. (2024). Reliability and accuracy of artificial intelligence chatgpt in providing information on ophthalmic diseases and management to patients. *Eye, 38*(7), 1368–1373.
6. Cossu, A., Spinnato, F., Guidotti, R., & Bacciu, D. (2024). Drifting explanations in continual learning. *Neurocomputing, 597*, Article 127960.
7. Fahlman, S. (1989). Faster learning variations on back-propagation: An empirical study. In D. Touretszky, G. Hinton, & T. Sejnowski (Eds.), *Proceedings of the 1988 connectionist models summer school* (pp. 38–51). San Mateo, Morgan Kaufmann.
8. Feldman, M. S. (2000). Organizational routines as a source of continuous change. *Organization Science, 11*(6), 611–629. https://doi.org/10.1287/orsc.11.6.611.12529

9. Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2020). Towards transparency by design for artificial intelligence. *Science and Engineering Ethics,26*(6), 3333–3361.
10. Fiol, C. M., & O'Connor, E. (2017). Unlearning established organizational routines-part i. *The Learning Organization, 24*(1), 13–29.
11. Green, S. B. (1991). How many subjects does it take to do a regression analysis. *Multivariate Behavioral Research,26*(3), 499–510. PMID: 26776715.
12. Grum, M. (2020, July 6-8). Managing human and artificial knowledge bearers: The creation of a symbiotic knowledge management approach. In *Business Modeling and Software Design: 10th International Symposium, BMSD 2020, Berlin, Germany, Proceedings 10* (pp. 182–201). Springer.
13. Grum, M. (2022). *AI-CPS Repository*. https://github.com/MarcusGrum/AI-CPS.git
14. Grum, M. (2022). *Construction of a concept of neuronal modeling*. Springer Nature.
15. Grum, M. (2024). Managing multi-site artificial neural networks' activation rates and activation cycles. In B. Shishkov (Ed.), *Business Modeling and Software Design* (pp. 258–269). Springer Nature Switzerland.
16. Grum, M. (2024). Researching multi-site artificial neural networks' activation rates and activation cycles. In B. Shishkov (Ed.), *Business Modeling and Software Design* (pp. 186–206). Springer Nature Switzerland.
17. Grum, M., Thim, C., & Gronau, N. (2022). Aiming for knowledge-transfer-optimizing intelligent cyber-physical systems. In A. L. Andersen, R. Andersen, T. D. Brunoe, M. S. S. Larsen, K. Nielsen, A. Napoleone, & S. Kjeldgaard (Eds.), *Towards sustainable customization: Bridging smart products and manufacturing systems* (pp. 149–157). Springer International Publishing.
18. Hedberg, B. (1981). *How organizations learn and unlearn* (Vol. 1, pp. 3–27). Oxford University Press.
19. Hestenes, M. R., & Stiefel, E. (1952). Methods of conjugate gradients for solving linear systems. *Journal of Research of National Bureau of Standards, 49*(6), 409–436.
20. Hislop, D., Bosley, S., Coombs, C. R., & Holland, J. (2014). The process of individual unlearning: A neglected topic in an under-researched field. *Management Learning, 45*(5), 540–560.
21. Holt, D. T., Armenakis, A. A., Feild, H. S., & Harris, S. G. (2007). Readiness for organizational change: The systematic development of a scale. *The Journal of Applied Behavioral Science,43*(2), 232–255.
22. Howard-Grenville, J. A. (2005). The persistence of flexible organizational routines: The role of agency and organizational context. *Organization Science, 16*(6), 618–636. https://doi.org/10.1287/orsc.1050.0150
23. Howells, J., & Scholderer, J. (2016). Forget unlearning? How an empirically unwarranted concept from psychology was imported to flourish in management and organisation studies. *Management Learning, 47*(4), 443–463. https://doi.org/10.1177/1350507615624079
24. Investigation of Frustration. (2019). *ACM*. https://doi.org/10.1145/3340764.3344915
25. Jian, J. Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics, 4*(1), 53–71.
26. Kemker, R., McClure, M., Abitino, A., Hayes, T., & Kanan, C. (2018). Measuring catastrophic forgetting in neural networks. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 32). PKP Publishing Services Network.
27. Keren, G. B., & Raaijmakers, J. G. (1988). On between-subjects versus within-subjects comparisons in testing utility theory. *Organizational Behavior and Human Decision Processes, 41*(2), 233–247.
28. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., et al. (2017). Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences, 114*(13), 3521–3526.
29. Klammer, A., & Gueldenberg, S. (2019). Unlearning and forgetting in organizations: A systematic review of literature. *Journal of Knowledge Management, 23*(5), 860–888. https://doi.org/10.1108/JKM-05-2018-0277

30. Laudon, K. C., Laudon, J. P., & Schoder, D. (2010). *Wirtschaftsinformatik: Eine Einführung*. Pearson Deutschland GmbH.
31. Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors,46*(1), 50–80. PMID: 15151155.
32. Mazurova, E., & Standaert, W. (2024). Implementing artificial intelligence across task types: Constraints of automation and affordances of augmentation. *Information Technology & People, 37*(7), 2411–2440.
33. Mertens, P., Bodendorf, F., König, W., Picot, A., Schumann, M., & Hess, T. (2005). *Grundzüge der Wirtschaftsinformatik* (Vol. 9). Springer.
34. Neyer, F. J., Felber, J., & Gebhardt, C. (2012). Entwicklung und validierung einer kurzskala zur erfassung von technikbereitschaft. *Diagnostica,58*(2), 87–99.
35. Nguyen, N. (2017). The journey of organizational unlearning: A conversation with William H. Starbuck. *The Learning Organization, 24*(1), 58–66.
36. Omowole, B. M., Olufemi-Philips, A., Ofadile, O., Eyo-Udo, N. L., & Ewim, S. E. (2024). Barriers and drivers of digital transformation in smes: A conceptual analysis. *International Journal of Frontline Research in Multidisciplinary Studies, 5*(2), 019–036.
37. Peffers, K., Tuunanen, T., Gengler, C. E., Rossi, M., Hui, W., Virtanen, V., & Bragge, J. (2006). The design science research process: A model for producing and presenting information systems research. *1st International Conference on Design Science in Information Systems and Technology (DESRIST), 24*(3), 83–106.
38. Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Management Informations Systems, 24*(3), 45–78.
39. Peres, R. S., Jia, X., Lee, J., Sun, K., Colombo, A. W., & Barata, J. (2020). Industrial artificial intelligence in industry 4.0—Systematic review, challenges and outlook. *IEEE Access, 8*, 220121–220139. https://doi.org/10.1109/ACCESS.2020.3042874
40. Riedmiller, M., & Braun, H. (1993). A direct adaptive method for faster backpropagation learning: The RPROP algorithm. in *Proceedings of International Conference on Neural Networks* (pp. 586–591). San Francisco, CA. citeseer.ist.psu.edu/riedmiller93irect.html
41. Roediger, H. L., Weinstein, Y., & Agarwal, P. K. (2010). *Forgetting: Preliminary considerations* (pp. 1–22). Psychology Press.
42. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088), 533–536. https://doi.org/10.1038/323533a0
43. Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 1: Foundations*. MIT Press.
44. Salvato, C., & Rerup, C. (2010). Beyond collective entities: Multilevel research on organizational routines and capabilities. *Journal of Management, 37*(2), 468–490. https://doi.org/10.1177/0149206310371691
45. Schaefer, K. (2013). *The perception and measurement of human-robot trust*. http://purl.fcla.edu/fcla/etd/CFE0004931
46. Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review, 112*(3), 610.
47. Shewchuk, J. R. (1994). *An introduction to the conjugate gradient method without the agonizing pain* (p. 1). Technical Report, Carnegie Mellon University, Pittsburgh, PA, USA.
48. Tsang, E. W. (2008). Transferring knowledge to acquisition joint ventures: An organizational unlearning perspective. *Management Learning, 39*(1), 5–20. https://doi.org/10.1177/1350507607085169
49. Tschopp, M., Ruef, M., & Monett, D. (2022). *Vertrauen Sie KI? Einblicke in das Thema Künstliche Intelligenz und warum Vertrauen eine Schlüsselrolle im Umgang mit neuen Technologien spielt* (pp. 319–346). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-63117-1_16
50. Weller, A. (2017). *Challenges for transparency*. CoRR *abs/1708.01870*. http://arxiv.org/abs/1708.01870

51. Wiczorek, R. (2011). *Entwicklung und evaluation eines mehrdimensionalen fragebogens zur messung von vertrauen in technische systeme*.
52. Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology, 55*, 235–269.
53. Wixted, J. T. (2005). A theory about why we forget what we once knew. *Current Directions in Psychological Science, 14*(1), 6–9.

Planning to Forget: A Simulation of Cognitive Workload in Emergency Response Teams

Jan Ole Berndt, Lukas Reuter, Anna-Sophie Ulfert-Blank, Benedikt Graf, Conny H. Antoni, Thomas Ellwart, and Ingo J. Timm

1 Introduction

Quick and adequate responses to emergency situations are as crucial as they are stressful for everyone involved. Emergencies can arise in various domains and contexts, including social, economic, as well as environmental crises like shooting rampages, epidemics, financial breakdowns, major fires, or oil spills. Response teams, such as fire fighters, have to assess the specific situation accurately, allocate resources, and

J. O. Berndt (✉) · I. J. Timm
German Research Center for Artificial Intelligence (DFKI), Trier, Germany
e-mail: jan.ole.berndt@dfki.de

I. J. Timm
e-mail: ingo.timm@dfki.de

L. Reuter
Center for Informatics Research and Technology, Trier University, Trier, Germany
e-mail: reuter@uni-trier.de

A.-S. Ulfert-Blank
Industrial Engineering and Innovation Sciences, Human Performance Management Group, Eindhoven University of Technology, Eindhoven, The Netherlands
e-mail: a.s.ulfert.blank@tue.nl

B. Graf
Work and Organizational Psychology, Göttingen University, Göttingen, Germany
e-mail: benedikt.graf@uni-goettingen.de

C. H. Antoni · T. Ellwart
Business, Work and Organizational Psychology, Trier University, Trier, Germany
e-mail: antoni@uni-trier.de

T. Ellwart
e-mail: ellwart@unitrier.de

I. J. Timm
Center for Informatics Research and Technology, Trier University, Trier, Germany

J. O. Berndt et al. (eds.), *Intentional Forgetting with Intelligent Systems*, Progress in IS,
https://doi.org/10.1007/978-3-032-17621-9_9

initiate action to prevent damage under time pressure. Time is often limited in emergency situations, which leads to a high amount of cognitive workload for such teams as they attempt to identify and implement the best possible response measures.

Cognitive workload can be understood as the amount or oversupply of information that has to be processed for making a decision [16, 26]. Having limited *information processing capacity* (IPC), humans (as well as computers) suffer from *information overload* if their cognitive workload exceeds their IPC [14, 33]. This can lead to mistakes due to poor decision-making, particularly in stressful situations. Previous research shows that cognitive workload can be reduced by subdividing complex problems into more manageable tasks and assigning them to specialist team members [12, 14]. In a team of fire fighters, for instance, there are team members responsible for actually fighting a fire, those for rescuing injured persons, those for closing off routes for public traffic, and various more tasks which require different competences. Effectively, this specialization is a form of intentional forgetting, as team members can neglect, ignore, or remain unaware of information that is irrelevant to their specific tasks [48]. Specialization can increase a team's IPC as each member can concentrate on their specific responsibilities. In fact, research shows that a smart allocation of work roles and tasks can reduce workload and increase efficiency [8, 47].

Nevertheless, it remains a challenge to plan adequate role configurations for emergency response teams in advance; i.e., to forget intentionally and goal directed. In addition to real-world training exercises, computer simulation can provide insights into team performance in different situations. While simulation is a well-established method for designing and evaluating processes in industrial applications, simulation for planning and assessing teamwork in critical situations requires an additional level of complexity. It has to include models of human decision-making and cognitive capacities. For that reason, this chapter introduces a novel agent-based model of human decision-making with limited information processing capacity. This model is designed to represent human IPC in teamwork situations in order to enable simulations of information overload.

The objective of this chapter is to develop the aforementioned simulation model and to evaluate whether it is suitable for simulating team performance under stress in various configurations between role specialization and generalization of team members. This approach is an interdisciplinary endeavor with a computer science and artificial intelligence focus on designing and evaluating the simulation model. Based on psychological and simulation research, the model extends the Beliefs-Desires-Intentions intelligent agent architecture with cognitive workload and heuristic decision-making. Subsequently, the chapter demonstrates how to apply that model in an agent-based social simulation of fire fighting in teams which represents a psychological experiment covering the same task. Both the laboratory experiment and the simulation are presented and discussed in detail. They show how information overload leads to mistakes and how these can be reduced through alternative team setups, implementing intentional forgetting. Moreover, the agent-based model is validated by fitting social simulation results to the experiment data which leads to parameter values that match well-known findings from psychological research. This demonstrates the agent decision-making architecture's capability of plausibly representing the effects of cognitive workload in teams as observed in the experiment, making it potentially usable for planning intentional forgetting through re-organization. Finally, the chapter closes with a concluding summary and an outlook on future research.

2 Cognitive Workload and Decision-Making in Teams

Modeling and simulating human decision-making, IPC, as well as cognitive workload requires a thorough understanding of psychological research. Consequently, this section first discusses psychological foundations of human cognition and forgetting in teamwork and their application to emergency response situations. Moreover, it presents related work in the field of social simulation to identify promising approaches to model decision-making using artificial agents. It discusses their application potential for emergency response simulation as well as shortcomings which motivate the design of a new approach to cognitive workload modeling.

2.1 Psychological Foundations: Information Processing in Teams

In psychology, cognitive workload is defined as the amount of information an individual is required to process at a given time to make a decision or perform a task [16]. Information processing includes different steps, such as perceiving new information input or stimuli, comparing this information with existing knowledge, retrieving information from long-term memory, and making action decisions [30, 51]. In a bush fire emergency, for example, a fire chief has to process information about the location of fires, predict their development, check for available fire fighting capacities, as well as assign particular tasks to fire fighting teams. In order to perform this task effectively, a certain amount of information needs to be processed via the limited IPC of the individual decision-makers. The limited IPC stimulates an overload condition when the workload rises too high and therefore exceeds the IPC of an individual. The resulting state is called information overload [16].

Information overload has been associated with low performance, poor decision-making, as well as high occurrence of errors [16, 51]. Studies in cognitive psychology show that high amounts of workload impede the perception of stimuli as well as memory retrieval capabilities [17, 33]. Under these conditions, the required knowledge to analyze the situation and decide on appropriate action can become incomplete [44]. This is particularly critical in emergencies when situation assessments and reactions have to be determined quickly and under stress. In such situations, humans utilize *heuristics* to make decisions [19]. Instead of meticulously planning optimal courses of action, heuristics provide cognitive shortcuts to find feasible solutions.

For instance, the aforementioned fire chief might simply allocate the nearest firefighters to each fire. While this reduces travel times and spares decision efforts, the resulting composition of assets and capabilities to extinguish those fires can be suboptimal. Hence, this approach, although highly time efficient, can lead to biases and error [18]. For example, decision-makers applying the mentioned take-the-best (i.e., take-the-nearest) heuristic are prone to stop comparing options too early and thus decide for erroneous action.

Nevertheless, the problem of limited individual IPC can be mitigated in appropriately organized teams. Teams can be understood as collective information-processing systems [24]. Subdividing joint tasks and assigning them to team members with particular roles reduces the amount of information each member must process [14]. This is effectively a form of *intentional forgetting*. A fire chief who is only responsible for a specific region can ignore any events which do not affect this region and delegate the handling of these events to other decision-makers. Consequently, the distribution of responsibilities in such teams (i.e., its role structure) determines the cognitive workload for its members [27]. In psychology, these distributions are called *team cognitions* [25]. They describe the extent to which knowledge is shared (generalist knowledge) or distributed (specialist knowledge) between team members [12, 14]. To reduce cognitive load and avoid information overload, teams should be predominantly specialized.

However, it is often impossible to form fully specialized teams because coordination requires a minimal amount of shared competences and awareness of other members' responsibilities [34]. In fact, this shared knowledge enhances trust within the team and improves team performance as too much specialization will also lead to mistakes due to mis-coordination. A fire chief ignoring fires in some regions still needs to know who is responsible for fighting them to make sure the emergency is kept under control. Thus, it is necessary to strike a balance between specialization and generalization. The challenge of striking such a balance is to guarantee smooth team functionality while keeping individual workload sufficiently low to avoid critical mistakes. This is the challenge to be met when planning for intentional forgetting in a team context.

2.2 Agent-Based Social Simulation: Modeling and Simulating Decision-Making in Emergency Response Teams

The aforementioned cognitive aspects of teamwork and individual decision-making are analyzed by psychologists in small group experiments. In addition, agent-based social simulation has become the de facto standard to analyze and plan emergency responses [40]. This includes simulations of evacuation and first responder actions by authorities as well as the RoboCup Rescue competition [23]. Agent-based models used for these simulations represent individuals in terms of agents. These agents can interact with each other as well as with their shared environment. This makes them particularly suitable to simulate small teams, larger organizations, and entire populations to evaluate the interdependent behaviors of their members.

Existing models of emergency scenarios can be categorized by their respective purpose which directly relates to the modeling techniques being used. For instance, models for evaluating disaster response plans focus on *demographic properties of a population* with daily routines in order to analyze the effects of early warning strategies on evacuation movements [11]. Their purpose is to evaluate rescue plans and warning strategies according to how fast a population can be evacuated or how severely a disaster will affect it [10, 43]. For such objectives, it is frequently sufficient to abstract from individual preferences or cognitive capabilities of potential victims and rescuers represented as agents in the model.

By contrast, several other models focus on *organizational structures* to represent teamwork between emergency responders and analyze their effectiveness [38]. Agent teams define roles of their members as well as communication channels. Analogous to psychological team research, these roles define which knowledge each team member has to process and allow for evaluating different team structures [2, 42, 54]. However, these evaluations largely focus on the impact of communication structures while leaving task distributions out of account. An exception to this is provided by Timm et al. who simulate specialist and generalist team performance in a firefighting scenario [49]. While that work shows benefits of specialist team cognitions over generalist ones, it focuses on collective information processing without specifically modeling individual cognitive capacities.

On an individual level, agent-based models sometimes use cognitive or deliberative architectures to represent decision-making processes, subjective behaviors, workload, and motivations. One of the most common examples is the *Adaptive Control of Thought—Rational (ACT-R)* architecture of human cognition [3]. ACT-R is based on theories from cognitive psychology and neuroscience. It distinguishes between declarative and procedural knowledge of an agent; the former consisting of facts that can be retrieved and the latter comprising production rules that manage information flows between different components of the model. This architecture allows for detailed representations of an individual agent's perception, recognition, and knowledge processing capabilities. It is especially well suited for modeling limitations of these capabilities like limited IPC and mental workload [17]. Consequently, ACT-R has been applied to emergency situations for modeling reaction times under pressure or distraction as well as for evaluating emergency procedures in the transport and critical infrastructure sectors [6, 13, 55].

An alternative option for modeling human decision-making is the *Beliefs-Desires-Intentions (BDI)* architecture for practical reasoning [39, 52]. This architecture represents mental states of agents, including knowledge (beliefs), goals (desires), as well as planned courses of action (intentions). Moreover, it provides decision-making algorithms which define how these states are processed by an agent in order to come to decisions about planned action for reaching its goals given its current knowledge. In distributed artificial intelligence, BDI agents are particularly useful for modeling teamwork and defining knowledge structures in those teams [41]. In social simulation, BDI agents are thus used for applications focusing on knowledge representation, planning, emotions, and cooperation [1]. This holds specifically for emergency simulations in which the BDI approach has been applied to joint planning by rescuer agents, to modeling personal preferences of residents facing bushfires, as well as to representing emotions and social bonds between evacuees [29, 45, 50].

Given those successful applications, both ACT-R and BDI agents are generally suitable for modeling cognitive workload and decision-making of emergency response teams. However, neither of their existing variants is fully capable of representing the aforementioned concepts of team cognition and limited IPC for planning and evaluating intentional forgetting at the same time. As ACT-R focuses mainly on an individual decision-maker's limited ability to retrieve information from memory [4], this architecture excels at modeling individual forgetting. However, it keeps teamwork largely out of account. While existing extensions of ACT-R (as well as other cognitive architectures) for social simulation do provide modules for communication and collective activities of several agents [46], it remains difficult to explicitly model a team's role structure and agent knowledge about distributed tasks. By contrast, existing BDI approaches assume perfect agent capabilities for memory retrieval and decision-making [52]. Deriving from artificial intelligence research, these agent models are only limited by the available computing power instead of simulating cognitive limitations of humans and the resulting heuristic decision-making. Nonetheless, BDI provides a well-established framework for cooperative problem-solving in teams with extensive formalizations of shared knowledge and joint objectives for coordinated action [52, 53].

To combine the advantages of both approaches, the next section presents a BDI-based model which integrates cognitive limitations to represent workload and its impact on decision-making in teams for agent-based social simulation. That model utilizes the formal framework of BDI to make it compatible with cooperative problem-solving approaches and with the authors' previous model of team cognitions [41]. However, it extends the existing BDI decision-making procedure with a model of IPC and cognitive workload that has been derived from psychological experiments and successfully evaluated in ACT-R models. Due to the well-defined reasoning process of BDI, it is easy to identify those steps where this limited cognition impacts an agent's decisions. This makes the resulting architecture and its implications on teamwork in emergency response situations easy to apply and understand. Hence, the approach presented in this chapter shows how to tap into the expressive and explanatory power of BDI for modeling teamwork in social simulation while mixing in mechanisms of limited cognitive capacity from models more true to the actual processes underlying human cognition. This provides the technological basis for simulating team configurations and their impact on cognitive workload in order to identify the need for and plan the implementation of forgetting in teamwork.

3 A Model of Cognitive Workload and Decision-Making

This section extends Wooldrige's formulation of the BDI decision-making process [52] with a model of cognitive workload based on the discourse agent for representing team cognitions in distributed artificial intelligence [41]. As depicted in Fig. 1, the BDI process consists of a repeatedly executed sequence of seven functions. First, an agent generates percepts *Per* by observing the state *Env* of its environment: *perceive Environment* : $\{Env\} \rightarrow \{Per\}$. Based on this information, the agent updates its belief base B about the environment according to those percepts by adding any new information, removing obsolete facts or replacing outdated knowledge: *revise Beliefs* : $\{Per\} \times \{B\} \rightarrow \{B\}$. After this step, the agent has internalized the required information for deliberating about its goals and making decisions about appropriate action to achieve them.

In contrast to previous work, the agent can also revise its desires based on its updated knowledge. While, e.g., Wooldridge uses the set of desires D solely as a temporary collection of potential goals to generate intentions from, desires remain persistent in this variant unless the environment has changed significantly. Here, a desire describes a state of the environment the agent wants to be reached. When revising its desires, the agent unifies that goal condition with its beliefs to remove all fulfilled desires or add any new objectives it may consider to reach: *revise Desires* : $\{B\} \times \{D\} \rightarrow \{D\}$. Subsequently, the agent uses its revised beliefs and desires to reconsider its intentions I. An intention is a tuple of a desire and a plan to reach the desired state: $I \subseteq D \times \Pi$, where Π is the set of all possible sequences of actions an agent can take. Being the set of all currently pursued courses of action to reach its goals, the agent must check whether its intentions are still valid, given the observed environment changes: *reconsider Intentions* : $\{B\} \times \{D\} \times \{I\} \rightarrow \{I\}$. Figure 1 depicts a single-minded commitment strategy [39]. The agent iterates over all of its current intentions and drops any of them which are either fulfilled or deemed impossible. An intention is fulfilled if the associated desire is reached (whether by its own or another agent's activities or by the environment's dynamics). Moreover, an

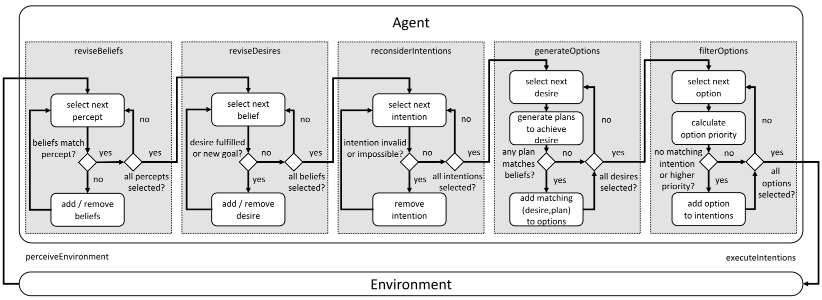


Fig. 1 BDI agent decision-making procedure

intention becomes impossible if the chosen plan to reach the goal state is no longer applicable.

After reconsidering its intentions, the agent potentially has some currently pursued goals left. It will add further intentions to them in accordance with its updated beliefs and desires by generating and filtering options for further action. To that end, the agent generates plans for each of its desires and combines these to tuples in a set $O \subseteq D \times \Pi$ of optional new intentions. If there are several possible ways to reach a goal, the agent can create different options for the same desire. If no plan can be generated for a desire, that goal cannot be reached by any means at the agent's disposal. However, the agent will also consider only those tuples as options which match its current beliefs; i.e., which contain a plan that is applicable to the current situation: *generate Options* : $\{B\} \times \{D\} \rightarrow \{O\}$. While the resulting options are all individually achievable, O can still contain contradictory goals and courses of action. Therefore, the agent must prioritize them together with its current intentions to filter out inconsistencies and select only the most important ones as its next intentions. Since option priorities depend on the current situation, the filtering process takes the agent's beliefs into account, as well: *filter Options* : $\{O\} \times \{B\} \times \{I\} \rightarrow \{I\}$. The result is a consistent set of intentions in which there is one intention at most for each desire and none of the associated plans conflict with each other. Then, the agent can perform one or more actions $\{a_1, \ldots, a_n\} \subseteq A$ from these plans: *execute Intentions* : $\{I\} \rightarrow 2^A$.

As pointed out in the preceding section, BDI agents are limited in the described functionalities only by the computational power at their disposal. Humans, on the other hand, have limited information processing capacity. This necessitates an extension of the BDI model to enable the analysis and design of intentional forgetting in teams by means of social simulation. From an IPC perspective, cognitive workload can affect the BDI process at different steps. For instance, information overload can influence the perception and comprehension steps (*perceive Environment*, *revise Beliefs revise Desires*, *reconsider Intentions*), thus hampering a person's *situation awareness* [15]. The result will be suboptimal decisions due to missing or outdated knowledge. Alternatively, actions may not be executed as planned in stressful situations which leads to undesired effects. However, the most immediate location of mistakes and limited cognitive capacity is the step of deliberating about potential decisions: *generate Options* [18, 41, 44]. Hence, this chapter focuses on the effects of cognitive workload on that part of the process.

Existing cognitive models of decision-making emphasize human limitations in retrieving required information from memory [4]. From a decision-making perspective, such limitations lead to incomplete or uncertain information for planning appropriate courses of action. As pointed out in the preceding section, human decision-making under uncertainty often uses heuristics as a shortcut to feasible albeit potentially suboptimal solutions [19]. More precisely, the probability of retrieving required information decreases as the workload increases [4]. This leads to an increasing probability for deciding heuristically as the information retrieval worsens under increasing amounts of cognitive workload. Consequently, individual performance of team members can be modeled by choosing between precisely planned activities and heuristically selected actions, depending on the cognitive workload a person faces. If that person has enough information processing capacity available, they will attempt to optimize their decisions, whereas they will tend to heuristic decision-making as their workload increases.

In order to include the described concept into the BDI process, this chapter proposes to decide whether to apply optimal planning methods or use a heuristic by a workload-dependent stochastic choice in an agent's *generate Options* function (Eq. 1). Both methods produce a set of matching plans for each desire the agent currently has: $plan : \{B\} \times \{D\} \rightarrow \{O\}$ and $heuristic : \{B\} \times \{D\} \rightarrow \{O\}$. However, optimal planning guarantees to return the best solution at the cost of high cognitive demands while heuristics are cognitively frugal at the expense of guaranteed optimality. The decision for either of them depends on the available IPC which is directly inverse to the current cognitive workload.

In psychological science, the amount of workload is given by the number of stimuli to be processed. In BDI terms, stimuli are perceptions which lead to beliefs and desires. Any perceptions that demand agent decisions change the set of beliefs and subsequently the set of desires D in the agent's *revise Desires* function. That set also determines the amount of options the agent has to generate and decide about. Therefore, the set of desires directly reflects the current cognitive workload such an agent is facing, making the probability $P(optimize|D)$ of planned or heuristic decision-making dependent on D (Eq. 2).

$$generate\,Options(B, D) = \begin{cases} plan(B, D) & \text{with } P(optimize|D) \\ heuristic(B, D) & \text{with } 1 - P(optimize|D) \end{cases} \tag{1}$$

$$\text{where} \quad P(optimize|D) = \frac{x}{1 + \left(\frac{|D|}{a}\right)^b} \quad \text{with} \quad x = \begin{cases} 1 & \text{if } |D| \leq load_{max} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

If a person's workload exceeds their maximum cognitive capacity $load_{max}$, that person is in a state of information overload. In that case, activities are always chosen heuristically as controlled by the factor x in Eq. 2. Otherwise, the probability of using a heuristic increases monotonically with increasing workload using a sigmoid curve with turning point a and slope b. This probability function is empirically corroborated by Estes [17], who shows that logistic regression models fit well to relations between the amount of processed information and the corresponding subjective workload. As long as a person has only few tasks, any additional one increases their perceived workload only marginally, leaving enough IPC to plan their actions optimally. The increase intensifies if that person's workload approaches their cognitive limit, leading to them extending their use of heuristics as decision-making shortcuts. At high levels of existing workload, any additional task will only slightly further increase the use of heuristics as they are already the person's predominant decision method. The next section shows how this model can be applied to simulate the effects of workload and role configurations in emergency response teams.

4 Simulating Workload of Emergency Response Teams

This section applies the BDI-based IPC and workload model to a simulated emergency response task. To that end, a, results from a laboratory experiment with human participants are compared with those from an agent-based social simulation of exactly the same setting. The purpose of this combined interdisciplinary empirical and simulation study is to evaluate the model's feasibility for simulation of teamwork under cognitive load. If the model can artificially replicate the laboratory experiment, it can provide a tool for designing team configurations that implement forgetting by reducing cognitive load through re-organization.

Consequently, the research question for the laboratory experiment is whether different team cognitions (generalists vs. specialists) in the simulated emergency response task actually lead to performance differences due to different cognitive load for the team members. If this is the case, the experiment is designed to furthermore measure the extent of this difference and provide reference data for testing the agent-based cognitive load model in a simulation study. In that simulation, agents simply replace the human participants in the experiment, facing the very same tasks with the same independent variables and being evaluated in their activities according to the very same dependent variables (i.e., team performance indicators). If the hypothesis of the model being adequate to simulate team performance under pressure is valid, the simulation study should be able to match the experiment results. Thus, the following section first describes the experimental setting and procedure for generating reference data. The subsequent section shows the simulation configuration and compares the simulation and experiment results to evaluate the aforementioned hypothesis.

4.1 Serious Game Environment and Laboratory Experiment Procedure

The laboratory experiment is based on a computer-based simulation task of firefighting which utilizes the *Networked Fire Chief (NFC)* serious gaming and simulation system [36, 37]. The NFC has been successfully used before to train teamwork of emergency responders as well as to analyze cognitive workload and evaluate team compositions according to their effectiveness in laboratory experiments [20–22, 32, 35, 49]. Research strongly suggests that the NFC is well suitable for generating cognitive workload in a controlled decision-making environment and for measuring its implications on situation awareness and teamwork. In particular, information processing, working memory, and their impact on decision performance has been meticulously studied using the NFC [20, 21]. Therefore, this chapter adapts the NFC to generate reference data on the relationships between cognitive workload and mistakes in decision-making. The original version of the NFC simulation system does not allow for controlling the user input. Hence, to apply the BDI workload model, the setting of the NFC is replicated by an agent-based simulation model in Repast Simphony.[1]

The NFC is designed to simulate bush fires in Australia [35, 37]. It features a grid-based map with various kinds of possible environment types (e.g., grassland, woods, housing). Grid cells can catch on fire and spread the fire to neighboring cells. To contain and fight those fires, fire trucks and other vehicles are available. These can be utilized by participants to extinguish fires if they have water available, which can be refilled in lakes. The participants can control these vehicles using a computer mouse. Routes are planned by drag-and-drop actions on the vehicles whereas firefighting and refilling at a vehicle's current position are done by clicking on a vehicle. Team performance in fighting fires can be measured by a scoring system based on the amount of burnt or unburnt grid cells over time as well as by evaluating detailed protocols of each participant's activities. In order to compare the performance of different teams in a controlled setting, it is possible to reuse particular maps and to specify scheduled events that introduce new fires at specific times and grid cells.

For the experiment described here, a particular subset of the outlined features is used. This subset is specifically designed to control cognitive load and to measure its effects on team performance. To that end, the map consists of 40×40 grid cells that display only grassland, houses, lakes, and fire trucks as shown on the left hand side of Fig. 2. Participants simultaneously control fire trucks in three-person teams. Every team member has the same map displayed on their own computer screen and is able to see the actions of their team mates on that map. The aim of the task is to minimize damage caused by wildfires as well to accurately follow the preset rules of asset usage depicted on the right hand side of Fig. 2 and explained below.

Minimizing the damage can be achieved by extinguishing fire on houses and grassland. The map consists of twelve blocks of buildings with eight colored houses each. There are three colors such that there are four blocks of buildings for each color. These blocks are symmetrically distributed over the map with respect to the center point. Grouped around the center are six colored clusters of lakes (in groups of two times three lakes for each color) as well as three colored groups of fire trucks (each containing three trucks). The colors of houses, lakes, and fire trucks correspond to each other and are referenced in the asset usage rules. Fire trucks have a limited water capacity and need to be refilled (using the lakes) after four firefighting actions.

The experiment follows a 2×2 design with *task complexity* (low vs. high) as well as *team cognition* (generalists vs. specialists) forming the independent variables as specified in Table 1 and detailed below. Participants were students from a German university who took part in the experiment for course credit or a payment. The sample size is $N = 177$ participants which were nested in 59 teams of three persons each. The participants were split into two experiment groups, one consisting of 30 teams

[1] https://repast.github.io/.

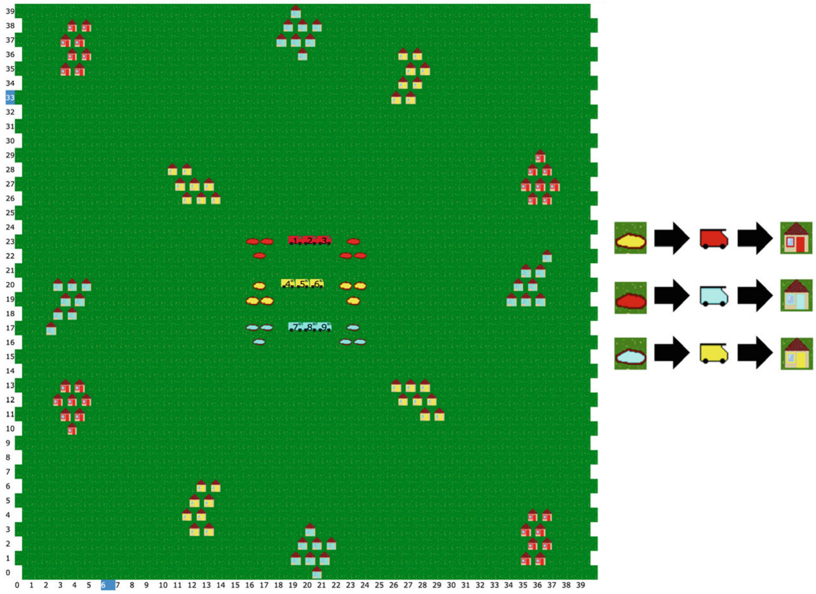


Fig. 2 Emergency simulation scenario (left) and high complexity fire truck usage rules (right)

of generalists and the other comprising 29 teams of specialists. The mean age of the predominantly (71%) female participants was M_{Age} = 23.33 years.

Table 1 Laboratory experiment design overview. The two independent variables of *task complexity* and *team cognitions* are operationalized in their respective two levels by means of different refill rules (cf. Fig. 2), colors of vehicles controlled, and colors of houses where fires can be fought by an individual participant

Task complexity	Team cognitions	Refill rule	Vehicles controlled	Houses responsible for
Low	Generalists	Any lake	3 Trucks of all different colors	Houses of all different colors ($12 \times 8 = 96$)
Low	Specialists	Any lake	3 Trucks of one single color	Houses of one single color ($4 \times 8 = 32$)
High	Generalists	Lakes of specific non-matching colors for each truck	3 Trucks of all different colors	Houses of all different colors ($12 \times 8 = 96$)
High	Specialists	Lakes of a single non-matching color with respect to truck color	3 Trucks of one single color	Houses of one single color ($4 \times 8 = 32$)

The two levels of *task complexity* were defined in the experiment by the set of knowledge rules on allowed asset usage that the participants had to follow to successfully complete the task (cf. Fig. 2). These rules describe allowed combinations of different colored fire trucks, lakes, and houses. For instance, a yellow fire truck might be used to fill up water on a blue lake and extinguish fires on yellow houses—there were no restrictions on which truck to use on grassland. The knowledge rules were introduced to the participants and displayed to the them on the right side of their screen during the entire experiment. While the participants were told to strictly follow the rules, they were able to disregard them by mistake (e.g., by refilling water at a wrong lake). The system allows for these actions to proceed in order to simulate how people might deviate from regulations under cognitive workload and stress.

In the experiment, task complexity was manipulated by means of two different sets of color combinations in the usage rules. For low cognitive complexity, the colors of fire trucks and houses match and there is no restriction at which color of lake a truck can be refilled. This makes it easy to recognize what asset can be used to fight which fire while the nearest lake is always sufficient for maintaining a truck's availability for use. High complexity, on the other hand, uses rules with restrictions for refilling the trucks with water as depicted in Fig. 2. These restrictions use non-matching colors. This increases the cognitive effort of the task because non-matching colors are counter-intuitive and often lead to mistakes unless serious cognitive capacity is allocated to memorizing, retrieving, or looking up the correct combinations due to the so-called Stroop Effect [31].[2]

Furthermore, research indicates that task complexity can also be manipulated via the amount of information assets, sources, or elements to be processed in a task [5, 28]. Therefore, the experiment additionally controls the rate of fires occurring in the environment. This rate directly corresponds to the amount of time pressure the participants face. Slow rates of fire occurrence lead to only few targets for firefighting the participants have to be aware of. Consequently, they have a small number of simultaneous goals to track which leaves time to remember and re-evaluate the rules to follow when making a decision. By contrast, a faster pace increases the number of goals to track at the same time and requires quick decision-making. Selecting an appropriate rate of fire occurrence allows for fine-tuning the cognitive complexity such that the overall task is neither too easy nor too difficult for the participants.

The second independent variable, *team cognition*, distinguishes between teams of generalists and those of specialists. In generalist teams, each participant controls one fire truck of each color. Thus, every participant has the necessary assets available to extinguish fires on houses of any color. However, such a team member must memorize and follow the complete set of usage rules for all of those assets. In low complexity settings, this is easily manageable by color matching whereas high complexity makes this task considerably more difficult. By contrast, specialist teams consist of participants who control only fire trucks of a single color. While the

[2] This effect can be observed in the form of delays in reaction times and increased occurrence of errors due to incongruent stimuli. A common example is the difficulty to correctly read out a color's name which is written in a different font color; e.g., the word *blue* written in *red* letters.

number of trucks controlled remains the same (i.e., three trucks per participant), specialists only have to remember the single usage rule for their specific type of vehicle. Hence, they can forget or do not even have to learn the other rules. This lowers the cognitive effort of their decision-making and narrows down the number of fires they are potentially responsible for fighting (only those four blocks of houses that match the color of their trucks). This setting reflects empirical research on team cognitions [12, 14, 34]: Generalists are more flexible in their actions at the cost of having to process more information. Contrastingly, specialists can focus on their respective expertise at the loss of flexibility which can lead to higher coordination efforts for achieving the team goal.

The performance of these different team cognitions for either complexity level was measured by three dependent variables: a *team performance score* based on successfully extinguished fires, the amount of *fire fighting mistakes*, and the amount of *refilling mistakes*. While the overall score allows for evaluating the general performance of a team, assessing the application of knowledge rules allows for more detailed analyzes of the individual participants' cognitive workload and its impact on their ability to correctly fulfill their tasks. To that end, all firefighting and refill actions were extracted from the log files of each experiment run for each of the teams. According to the asset usage rules, these can be subdivided into correct and incorrect actions for every participant (with a mismatch between vehicle and house colors indicating an incorrect firefighting action and a (mis)match between lake and vehicle colors indicating an incorrect refill action). However, pre-tests had shown that the overall amount of activities differs between participants: Some persons are highly active and fight a lot of fires whereas others are either more reluctant or more considerate which leads to fewer activities. Consequently, the ratio of mistakes relative to the total amount of actions forms the respective dependent variable (for each, firefighting and refilling actions). This balances out the increased possibility of making mistakes for higher activity counts and makes individual scores comparable between the four combinations of independent variable levels outlined in Table 1.

The experiment lasted about 90 min for each team of three participants and was carried out in sequence between January 01, 2017, and January 01, 2018, with one team at a time. The particular setting described above was part of a larger experiment which included an introduction phase, a learning phase, and an experimental phase for each time. First, the participants were randomly assigned to one of the two experiment groups (generalists or specialists; with the entire three-person team being assigned to the same group). They received a short introduction about the background of the study and filled out a demographic questionnaire (introduction phase). This was followed by a presentation of the game functionalities and a 10-minute learning tutorial, in which the participants were able to practice the functions of the computer-based task on the basis of a task sheet (learning phase). Afterward, two task runs were conducted, one for each complexity level with vehicle assignments according to the team's experiment group (experimental phase). Each task run consisted of a short learning period of the rules (30 s); a rules clarification period, in which the rules could be discussed (60 s); and the subsequent 3-minute simulation period, in which the participants actually played the serious game and data for the dependent variables was recorded. Since the participants were made aware of the actual study goal and procedure in the introduction phase, a formal debriefing was unnecessary. However, a final question round was held to receive feedback on the experiment and to clarify any misunderstandings.

From the 59 log file collections (one for each participating team), four contained corrupt data and had to be excluded from further processing. Thus, all results used as reference data for the simulation of teamwork under cognitive load are based on the remaining 29 teams of generalists (87 participants) and 26 teams of specialists (78 participants). This is still a large enough sample size for the mistake measures to produce meaningful results since these are based on individual actions instead of joint team performance.

In general, participants reported to have enjoyed the experiment and they were able to successfully work together in fighting fires. However, they did indeed tend to make mistakes; the amount depending mainly on the task complexity and the rate of fire occurrence. In the low complexity setting, refill mistakes are impossible and firefighting mistakes are easy to avoid by simple color matching. Consequently, low complexity is unsuitable for comparing team cognitions by the induced cognitive load which would lead to mistakes. This setting will be excluded from evaluating the BDI model of cognitive workload. In the high complexity setting, slow occurrence rates still make it easy to avoid mistakes for both generalists and specialists while too rapidly occurring fires lead to both team cognitions to fail at avoiding mistakes. Hence, a middle ground rate of fire occurrence is best suitable in this setting to measure cognitive load and its impacts on individual as well as team performance. This is in line with previous research results on teamwork under stress using the NFC [49].

Moreover, the following simulation of cognitive workload for emergency response team members will particularly use the refill mistakes measure as reference data to evaluate the BDI-based model. This is due to firefighting actions occurring on both burning grassland and houses. When experiment participants use their trucks to extinguish fires on grassland, this will never be a mistake. This makes the variable of firefighting mistakes less robust because a generalist team might assign one member to save houses while all others just help contain fires on nearby grassland. In these cases, the potential to make mistakes becomes underrated for generalists. By contrast, all members of both generalist and specialist teams face the same unavoidable challenge of refilling their trucks in accordance with the rules as long as they use them at all. Hence, the refill mistakes variable has turned out to be the most reliable in representing the impact of cognitive workload on a person's accuracy of rule application. The following section thus describes the application of the BDI-based workload model to the NFC experiment setting and compares the simulation results with the measurements for that variable.

4.2 Simulation Configuration and Results

The goal of simulating individual and team behavior in this chapter is to evaluate how well the BDI-based workload model is suitable for representing cognitive workload and the effects of information overload in a social simulation. Consequently, each player of the serious game (i.e., each experiment participant) is represented by a BDI agent in the simulation. To make decisions, an agent keeps a set of beliefs B about the entire current situation including all fires, locations of fire trucks, and their water fill levels. Each time an agent activates, it perceives the environment and revises its beliefs according to any changes. That is, an agent tracks and updates information about the positions of fire trucks, their water tank status, and positions of fires. Depending on its responsibility for specific (or all) houses and the state of its controlled trucks, an agent can then decide whether any of the observed changes require action to be taken.

Whether an agent needs to take action depends on its desires D. These desires are to fight any existing fires it has the appropriate assets for and to keep its fire trucks filled. When a new fire breaks out (as denoted in its updated beliefs), the agent creates a desire to fight that fire if it has the correct truck to do so. Hence, an agent's workload $|D|$ depends on the number of fires it is currently responsible for fighting. In addition, if a fire truck runs empty, the agent creates a desire to refill it. Thus, its corresponding cognitive workload depends on the number of assets it controls and on their usage. The more fires it decides to fight, the more fire trucks an agent has to refill in the process.

After revising beliefs and desires, the agent drops any existing intentions to fight fires that have been or cannot be achieved. This is the case if a fire has been extinguished, the area has burned down, or all matching fire trucks need refilling. Intentions that are still achievable are kept and the agent will continue working on them. That is, it will not abort any firefighting activities if the respective truck is still filled with water. Also, if a truck is on its way to a fire or to refill at a lake, the agent will proceed with driving the truck there and perform the planned action. Thus, individual agents and agent teams in this model remain committed to their intentions as long as they are appropriate [7, 9].

Subsequently, an agent creates new plans for fighting fires as well as for refilling trucks according to any new desires by means of its *generate Options* function. These plans select a particular truck to use, calculate the shortest path to its destination (burning house/grassland or lake for refilling), and denote the action to be taken there (fire fighting or refilling). Because refilling is a necessary prerequisite for extinguishing fires, that activity always takes precedence when the agent prioritizes its newly generated plans. Moreover, this activity is also subject to the most complicated rules, resulting in humans making the most mistakes when selecting a refill lake for their trucks [49]. In order to model realistic human behavior under high cognitive workloads, agents in this scenario should exhibit the same behavior.

To that end, an agent's refill decisions are either optimally planned or heuristically made. For optimal planning, the function $plan(B, D)$ selects the nearest correct watering place for each refill desire and returns all such optimized plan-desire-pairs as options. Thus, optimal planning always takes the asset usage rules into account and never makes mistakes. For heuristic planning, the function $heuristic(B, D)$ implements the *color heuristic*, which simply selects a lake that matches the color of the fire truck to be replaced. For the high complexity setting of the laboratory experiment, this decision is a mistake. However, this behavior represents the difficulty of matching different colors that is described by the Stroop Effect as outlined above [31]. As the heuristic selects the first obvious (albeit wrong) solution that comes to mind, it can be considered a variant of the take-the-best heuristic [19]. That is, the heuristic stops searching for the best solution as soon as any seemingly feasible course of action is found.

Depending on its current cognitive workload, an agent uses either its heuristic or optimal planning as specified in Eq. 2. As a result, agents facing high cognitive workload due to a large number of incidents within their area of responsibility tend to resort to heuristic decision-making which can lead to mistakes. The parameters of its workload curve control when and how gradually or suddenly an agent will begin facing information overload and thus tend to make mistakes as Fig. 3 depicts. Asymmetric workload curves (Fig. 3, top left) represent model gradually approaching information overload. These agents are either overwhelmed by information early on with a long tail of retaining the possibility for optimal planning or they face slowly increasing information overload until they suddenly reach their maximum IPC. Symmetric workload curves (Fig. 3, top right) lead to a balanced gradual decrease in optimal planning capability. These agents become overwhelmed by information around the turning point of those curves while retaining some capability to plan optimally and with the intensity of that effect being controlled by the curve's slope. Finally, truncated workload curves (Fig. 3, bottom) lead to a sudden drop in planning capability. These agents become suddenly overwhelmed by information to be processed as they reach their maximum IPC, similar to the effects of some asymmetric curves.

In order to evaluate whether an agent's workload curve can be parameterized to yield realistic rates of mistakes in the NFC-based experiment setting, the simulation explores a variety of parameter combinations for that curve. In particular, the turning point of the workload curve is varied between $a = 5$ and $a = 15$, the curve's slope between $b = 2$ and $b = 4$, and the maximum IPC in steps of two between $load_{max} = 5$ and $load_{max} = 19$. Moreover, the agents take either the role of specialists or that of generalists with the respective responsibilities and fire truck assignments as applied in the laboratory experiment. As depicted in Table 2, this leads to a full-factorial experiment design with $11 \times 3 \times 8 \times 2 = 528$ design points for the simulation. For each of these design points, a team of three agents is assembled and run through the same firefighting task that human participants faced in the laboratory experiment. To account for the stochasticity of agent decisions, each of those simulation runs is repeated 300 times (for a total number of $158,400$ simulation repetitions). Then, the relative ratios of refill mistakes are derived from the log files of agent activities in the simulation, statistically aggregated, and compared with the results of human participants.

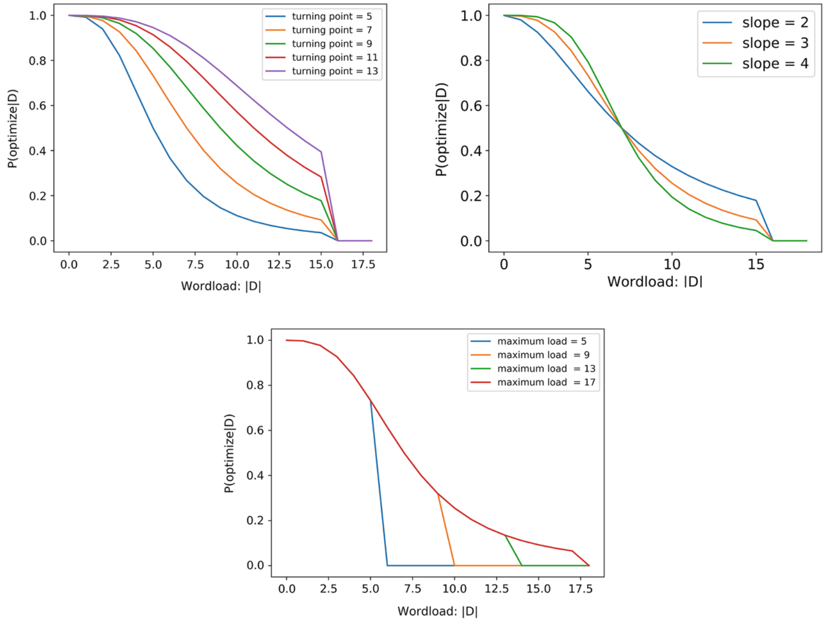


Fig. 3 Examples of cognitive workload curves analyzed in the simulation. Top left: Asymmetric curves with varying turning points a (with maximum IPC $load_{max} = 15$ and slope $b = 3$). Top right: Symmetric curves with varying slopes b (with maximum IPC $load_{max} = 15$ and turning point $a = 7$). Bottom: Truncated curves with varying maximum IPC $load_{max}$ (with turning point $a = 7$ and slope $b = 3$)

Table 2 Simulation experiment design of varying workload curves and team cognitions. All parameter values are combined with each other in a full-factorial design of $11 \times 3 \times 8 \times 2 = 528$ design points with 300 simulation repetitions each

Parameter	Values
Workload curve turning point a	{5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15}
Workload curve slope b	{2, 3, 4}
Workload curve maximum IPC $load_{max}$	{5, 7, 9, 11, 13, 15, 17, 19}
Team cognitions	{`generalists`, `specialists`}
Simulation repetitions	300 per design point

Figure 4 shows the results of the laboratory experiment of the high complexity setting (upper left) together with those of selected simulation design points for the same setup. The y-axis of each plot depicts the ratio of mistakes made by members of specialist and generalist teams when refilling their fire trucks (with 1.0 meaning all refill actions taken are erroneous while 0.0 represents all refill actions conforming to the rules). In the laboratory experiment, human participants in both generalist and specialist teams tended to make a considerable number of mistakes. For generalists,

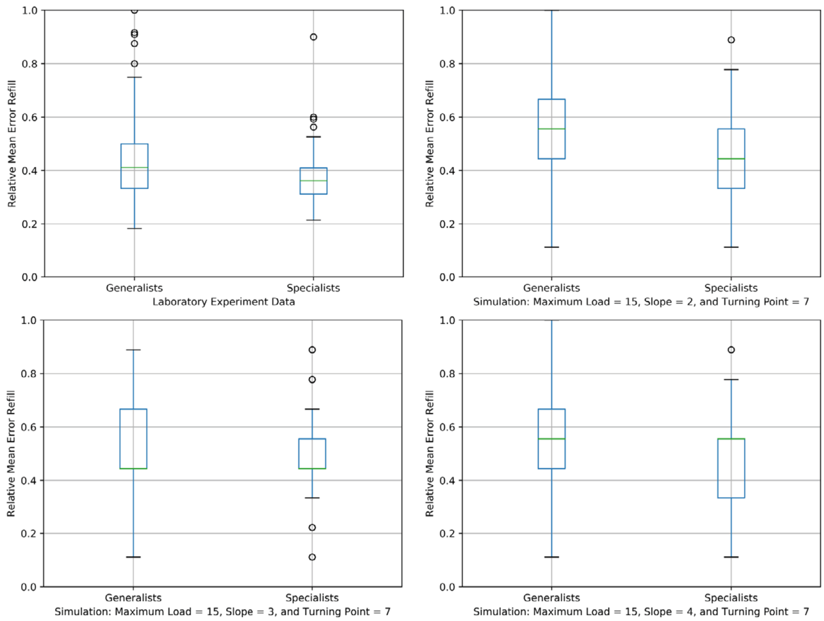


Fig. 4 Fire truck refilling mistakes made by specialist and generalist teams of humans and agents with varying cognitive workload curves (maximum IPC $load_{max} = 15$; turning point $a = 7$; slope $b \in \{2, 3, 4\}$)

41.2% of all refill actions were against the rules on average (median value). Specialists tend to make fewer mistakes (median 36.2%) and the quantile width of their performance is smaller than that of generalists. Nonetheless, in both cases there are outliers toward the higher end of the spectrum. These participants possibly either misunderstood the asset usage rules or they actively ignored them (a mistake rate of 100% seems very unlikely otherwise).

Overall, the laboratory experiment results show that a specialist team cognition leads to better and more predictable performance than a generalist setup, confirming the experiment hypothesis. The amount of observed mistakes indicates a state of information overload for team members in either role. However, specialists tend to cope better with that situation as they have fewer information items to process (responsibilities and rules to follow). This leaves them with slightly higher IPC remaining to avoid making mistakes. These results are backed by existing research on information overload in psychology and show that the NFC task is suitable to distinguish between specialist and generalist team cognitions [14, 16, 17, 49]. Consequently, it can serve as a reference for evaluating the agent-based simulation model.

Indeed, the simulation results shown in Fig. 4 depict a similar effect. These are measurements from the best matching design points out of the 528 examples in the analysis. In these examples, both specialists and generalists suffer from information overload which leads to mistakes. Thus, their workload in the emergency response scenario frequently exceeds their cognitive capacity for accurate planning. However, while the ratio of mistakes in specialist teams is only slightly lower than that of generalists on average, the former tend to vary toward fewer mistakes than the latter. In these cases, the agents are less likely to use heuristics when facing small amounts of workload, which makes their behavior more predictable and less prone to error. The difference between generalist and specialist median mistake rates in the simulation is most notable for a small workload curve slope value of $b = 2$. In this case, the BDI-based cognitive workload simulation model can distinguish between these team cognitions similar to the experiment with human participants. For steeper slopes, the performance of both team cognitions becomes more uniform because mistake rates increase more drastically as an agent's information load approaches the turning point a. Thus, the workload curve's slope controls how gradual or abrupt an agent will begin making mistakes when facing information overload.

In the depicted best-fitting result plots, all workload functions of the BDI agents use a turning point of $a = 7$ and a maximum IPC of $load_{max} = 15$. Interestingly, these values reflect human information processing capabilities known from psychological experiments [33]: On average, a person can memorize and retrieve seven information items at a time. Thus, having the steepest decline in an agent's capability for optimal planning at that mark implements exactly that effect. Additionally, the value of $load_{max}$ corresponds to the IPC being reached at twice that number of information items. This leads to the same symmetric sigmoid shape of the workload curve (cf. Fig. 3) as experimentally found by Estes [17]. Finding both these aspects reflected in the best matching simulation results can be viewed as strong circumstantial evidence for the BDI-based workload model's validity in the NFC experiment.

Nevertheless, there is one obvious flaw in the results shown that is subject to future research: The agents produce consistently higher mistake rates (ranging from median values of 44.4% to 55.6%) and the quantile width of their performance is larger than that of human participants. There are the following three possible reasons for this phenomenon.

1. *Agents that face information overload will always make mistakes when their maximum IPC is exceeded.* This is not necessarily the case for human participants in the experiment. A person may be completely overwhelmed by the situation. Still, that person is capable of making optimal decisions from time to time. To account for this, the workload function can be offset in y-direction by a certain value such that there is always a chance for an agent to plan optimally regardless of cognitive load.
2. *Agents are doomed to predictably make the same mistakes each time they decide heuristically.* While human experiment participants are likely to use heuristics in their decision-making, they may apply several or different heuristics (e.g., sometimes wrongly choose a lake by color, sometimes just the nearest one). These will not necessarily lead to mistakes. In fact, heuristic decision-making has the strength of providing a shortcut to a decision that is correct in most

instances [18]. Hence, switching out the color heuristic for a different one or a set of heuristics can account for this potential problem.

3. *Agents only refill their fire trucks when they have to.* That is, an agent will generate a refill desire only when a fire truck's water tank is empty. This happens particularly often when the agent has to fight a lot of fires. In this situation, the agent faces high cognitive load. Therefore, it is especially prone to making mistakes when its trucks need refilling. A more intelligent strategy is to refill the trucks early when few fires need fighting. This can be implemented by adjusting the conditions under which agents generate refill desires (e.g., by introducing a conditional refill threshold that utilizes the full vehicle capacity when there are many fires and refills early when there are few).

In order to determine which of these possibilities replicates best the behavior of human experiment participants requires more detailed analyses of the existing log files as well as information about their cognitive load. The latter has been gathered in another experiment where secondary tasks and additional questionnaires are utilized to track self-assessments of cognitive workload and free IPC similar to those used by Estes [17]. These will serve to further validate and refine the BDI workload model.

Notwithstanding the aforementioned limitations of the simulation results, the cognitive workload modeling approach presented in this chapter can still produce the effects observed in the laboratory experiment to a large degree. Thus, it is a promising candidate for a simulation-based tool to design emergency response team compositions which enable their members to make correct decisions under stress. There are three reasons for this. Firstly, the model reproduces the effects of information overload as well as their differences between specialist and generalist emergency response teams in a social simulation. Secondly, it is based on psychological theory and its most appropriate parameterization matches findings from psychological research. Finally, it is embedded in the well-understood BDI framework of decision-making which is particularly suitable for modeling teamwork tasks that require concerted activities, joint intention, communication, and knowledge shared between team members [41]. If these results can be generalized, the work presented here is an important step toward using agent-based models and social simulation for designing and evaluating role configurations in teams which alleviate cognitive load; i.e., to plan forgetting processes in an organizational context.

5 Conclusions

This chapter has presented the problem of limited information processing capacity (IPC) of emergency response teams to illustrate the need for forgetting in organizations. These teams have to make accurate situation assessments and decide under stress for appropriate action. If the amount of information team members have to process exceeds their IPC, this can lead to critical mistakes due to information overload.

Psychological research shows that subdividing tasks between specialist team members can prevent capacity overload, enhance team decision-making performance, and help avoid individual mistakes. Since social simulation is the de facto standard for designing and evaluating emergency response plans, the chapter has motivated the need for representing cognitive workload in these simulations to reflect human decision performance. Such a simulation model can then be utilized to better understand, evaluate, and adapt so-called team cognitions of emergency response teams to avoid mistakes. That is, it can provide a tool for planning to forget by means of re-organization. To that end, this chapter has presented a model for integrating cognitive workload into the Beliefs-Desires-Intentions (BDI) agent architecture. That model utilizes an empirically supported workload curve as well as well-researched heuristics to represent cognitive limitations of humans in social simulations.

To test real-world team performance and to generate reference data for the model's evaluation, a psychological laboratory experiment has been conducted. In that experiment, human participants form specialist and generalist teams in a simulated emergency response game in which they control firefighting vehicles. The results show that both teams suffer from information overload when experiencing stress which leads to mistakes. However, specialist teams tend to make fewer mistakes and are more reliable than generalists in the experiment scenario. This shows that the scenario is suitable for producing information overload, stress, and the resulting effects on decision performance as reported by existing research in cognitive and organizational psychology. Moreover, from a practical organization design perspective, it shows the potential of forgetting by re-organization to avoid these effects.

Based on these results, the BDI model has been applied for evaluation to the very same simulated emergency response scenario with agents replacing all human experiment participants. In the simulation, this model reproduces the differences between team cognitions as observed in the experiment: Specialists make fewer mistakes than generalists and their performance tends to be more predictable. The differences between these teams are in the same order of magnitude as those from the laboratory experiment. Furthermore, fitting the parameters of the agent-based model to the reference data results in values for an agent's IPC that match well-established findings from psychological research. This is a strong indicator for the model's validity in the simulated setting. Hence, the model is a potent contender for simulating decision performance of human teams under conditions of information overload, forming an important step toward planned intentional forgetting in organizations.

Nonetheless, additional research is still required to complement and expand on those results. In particular, the modeled agents tend to make more mistakes than human experiment participants for both specialist and generalist teams alike. Therefore, possible adjustments of the agent behavior have been discussed in the chapter that require testing in further simulation experiments. These will reveal implications of agent strategies in the emergency response scenario as well as impacts of the decision heuristics in the model in further detail. While the workload curve represents accurately when an agent is overwhelmed by information, the applied heuristics determine whether and how this will actually result in a mistake. Thus, additional and alternative heuristics should be tested.

Moreover, planning ahead smartly can even avoid information overload in the first place. The BDI architecture easily allows for including this typical human behavior in future simulation studies. On the one hand, this will reduce the number of mistakes agents make and will result in overall more intelligent behavior. On the other hand, planning and controlled action is hampered by incomplete knowledge and limited awareness of the situation in real-world emergencies [15, 44]. To represent this and to avoid unrealistically rational behavior, additional cognitive limitations can be included at other steps of the BDI process. This holds particularly for an agent's perception and knowledge retrieval where findings from other cognitive agent architectures can be integrated [4]. These tasks, however, are subject to the authors' future research.

Acknowledgements
Competing Interests The project *AdaptPRO: Adaptive Process and Role Design in Organizations* (Grant title: Adaptive forgetting through emergent knowledge structures in socio-technical systems. DFG project number: 318341635. Grant numbers: TI 548/-1, EL 269/9-1, and AN 205/6-1) was funded by the German Research Foundation (DFG) within the Priority Program "Intentional Forgetting in Organizations. Mechanisms of forgetting as adaptive performances of organizations to an environment of constantly growing amounts of information" (SPP 1921).
Ethics Approval The ethics committee of Trier University approved the procedures of the research project and the included experiments (06.01.2016, 74/2018, and 75/2022). At the beginning of the experiment reported in this chapter, participants read information about the study, signed a consent form, and read information about their role in the experiment as well as a description of the software they would be working with.

References

1. Adam, C., & Gaudou, B. (2016). BDI agents in social simulations: A survey. *The Knowledge Engineering Review, 31*(3), 207–238.
2. Aldewereld, H., Tranier, J., Dignum, F., & Dignum, V. (2011). Agent-based Crisis Management. In *CARE@AI 2009 / CARE@IAT 2010 International Conference on Collaborative Agents - Research and Development* (pp. 31–43). Springer.
3. Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction, 12*(4), 439–462.
4. Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology, 30*(3), 221–256.
5. Carey, J. M., & Kacmar, C. J. (1997). The impact of communication mode and task complexity on small group performance and member satisfaction. *Computers in Human Behavior, 13*(1), 23–49.
6. Carvalho, P. V. R., de Oliveira, M. V., & Luquetti dos Santos, I. J. (2009). A computerized tool to evaluate the cognitive compatibility of the emergency operational procedures task flow. *Progress in Nuclear Energy, 51*(3), 409–419.
7. Castelfranchi, C. (1995). Commitments: From individual intentions to groups and organizations. In V. R. Lesser & L. Gasser (Eds.), *Proceedings of the First International Conference on MultiAgent Systems (ICMAS 1995)*, (pp. 41–48). The MIT Press.
8. Charbonneau, D., & Dornhaus, A. (2015). When doing nothing is something: How task allocation strategies compromise between flexibility, efficiency, and inactive agents. *Journal of Bioeconomics*, *17*(3), 217–242.

9. Cohen, P. R., & Levesque, H. J. (1990). Intention is choice with commitment. *Artificial Intelligence, 42*(2–3), 213–261.
10. Crooks, A. T., & Wise, S. (2013). GIS and agent-based models for humanitarian assistance. *Computers, Environment and Urban Systems, 41*, 100–111.
11. Dawson, R. J., Peppe, R., & Wang, M. (2011). An agent-based model for risk-based flood incident management. *Natural Hazards, 59*(1), 167–189.
12. DeChurch, L. A., & Mesmer-Magnus, J. R. (2010). The cognitive underpinnings of effective teamwork: A meta-analysis. *Journal of Applied Psychology, 95*(1), 32–52.
13. Deng, C., Cao, S., Wu, C., & Lyu, N. (2019). Modeling driver take-over reaction time and emergency response time using an integrated cognitive architecture. *Transportation Research Record, 2673*(12), 380–390.
14. Ellwart, T., & Antoni, C. H. (2017). Shared and distributed team cognition and information overload: Evidence and approaches for team adaptation. In *Information and communication overload in the digital age* (pp. 223–245). IGI Global.
15. Endsley, M. R., & Garland, D. J. (2000). *Situation awareness analysis and measurement*. Lawrence Erlbaum Associates.
16. Eppler, M. J., & Mengis, J. (2004). The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. *The Information Society, 20*(5), 325–344.
17. Estes, S. (2015). The workload curve: Subjective mental workload. *Human Factors, 57*(7), 1174–1187.
18. Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science, 3*(1), 20–29.
19. Gigerenzer, G., Hertwig, R., & Pachur, T. (Eds.). (2011). *Heuristics: The foundations of adaptive behavior*. Oxford University Press.
20. Gonzalez, C., Thomas, R. P., & Vanyukov, P. (2005). The relationships between cognitive ability and dynamic decision making. *Intelligence, 33*(2), 169–186.
21. Gutzwiller, R. S., & Clegg, B. A. (2013). The role of working memory in levels of situation awareness. *Journal of Cognitive Engineering and Decision Making, 7*(2), 141–154.
22. Gutzwiller, R. S., Clegg, B. A., & Blitch, J. G. (2013). Part-task training in the context of automation: Current and future directions. *The American Journal of Psychology, 126*(4), 417–432.
23. Hawe, G. I., Coates, G., Wilson, D. T., & Crouch, R. S. (2012). Agent-based simulation for large-scale emergency response: A survey of usage and implementation. *ACM Computin Survey, 45*(1), 1–51.
24. Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin, 121*(1), 43.
25. Hollingshead, A. B., Gupta, N., Yoon, K., & Brandon, D. P. (2011). Transactive memory theory and teams: Past, present, and future. In *Theories of team cognition: Cross-disciplinary perspectives* (pp. 421–455). Taylor & Francis.
26. Kirsh, D. (2000). A few thoughts on cognitive overload. *Intellectica, 30*(1), 19–51.
27. Kozlowski, S. W. J., & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest, 7*(3), 77–124.
28. Lankton, N. K., Speier, C., & Wilson, E. V. (2012). Internet-based knowledge acquisition: Task complexity and performance. *Decision Support Systems, 53*(1), 55–65.
29. Luna-Ramirez, W., & Fasli, M. (2018). Bridging the gap between ABM and MAS: A disaster-rescue simulation using Jason and NetLogo. *Computers, 7*(2), 24.
30. Lutz, S., & Huitt, W. (2003). Information processing and memory: Theory and applications. *Educational psychology interactive* (pp. 1–17). Valdosta State University.
31. MacLeod, C. M. (1991). Half a century of research on the stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163.
32. McLennan, J., Holgate, A. M., Omodei, M. M., & Wearing, A. J. (2006). Decision making effectiveness in wildfire incident management teams. *Journal of Contingencies and Crisis Management, 14*(1), 27–37.

33. Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81.
34. Oertel, R., & Antoni, C. H. (2015). Phase-specific relationships between team learning processes and transactive memory development. *European Journal of Work and Organizational Psychology, 24*(5), 726–741.
35. Omodei, M., Elliott, G., & Walshe, M. (2004). *Development of computer simulated wildfire scenarios for the experimental investigation of unsafe decision making* (Technical Report 2), Bushfire CRC, La Trobe University.
36. Omodei, M., Taranto, P., & Wearing, A. (2003). *Networked fire chief (version 1.33)* [computer program]. La Trobe University.
37. Omodei, M., & Wearing, A. (1995). The fire chief microworld generating program: An illustration of computer-simulated microworlds as an experimental paradigm for studying complex decision-making behavior. *Behavior Research Methods, Instruments, & Computers, 27*, 303–316.
38. Quillinan, T. B., Brazier, F., Aldewereld, H., Dignum, F., Dignum, V., Penserini, L., & Wijngaards, N. (2009). Developing Agent-based organizational models for crisis management. *8th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2009)* (pp. 45–51). IFAAMAS.
39. Rao, A. S., & Georgeff, M. P. (1991). Modeling rational agents within a BDI-architecture. *Knowledge representation and reasoning (KR & R 1991)* (pp. 473–484). Morgan Kaufmann.
40. Reuter, L., Berndt, J., & Timm, I. (2018). Simulating psychological experiments: An agent-based modeling approach. *Fourth International Conference on Human and Social Analytics (HUSO 2018)* (pp. 5–10). IARIA.
41. Reuter, L., Berndt, J. O., Ulfert, A.-S., Antoni, C. H., Ellwart, T., & Timm, I. J. (2019). Intentional forgetting in distributed artificial intelligence. *KI - Künstliche Intelligenz, 33*(1), 69–77.
42. Ruiz-Martin, C., Wainer, G., Bouanan, Y., Zacharewicz, G., & Paredes, A. L. (2016). A hybrid approach to study communication in emergency plans. In *2016 Winter Simulation Conference (WSC)* (pp. 1376–1387). IEEE.
43. Saoud, N.B.-B., Mena, T. B., Dugdale, J., Pavard, B., & Ahmed, M. B. (2006). Assessing large scale emergency rescue plans: An agent based approach. *The International Journal of Intelligent Control and Systems, 11*(4), 260–271.
44. Scholten, L., Van Knippenberg, D., Nijstad, B. A., & De Dreu, C. K. (2007). Motivated information processing and group decision-making: Effects of process accountability on information processing and decision quality. *Journal of Experimental Social Psychology, 43*(4), 539–552.
45. Singh, D., Padgham, L., & Logan, B. (2016). Integrating BDI agents with agent-based simulation platforms. *Autonomous Agents and Multi-Agent Systems, 30*(6), 1050–1071.
46. Smart, P. R., & Sycara, K. (2014). Cognitive social simulation and collective sensemaking: An approach using the ACT-R cognitive architecture. In *6th International Conference on Advanced Cognitive Technologies and Applications* (pp. 10–19). Venice, Italy.
47. Tang, J., Zhu, K., Guo, H., Gong, C., Liao, C., & Zhang, S. (2018). Using auction-based task allocation scheme for simulation optimization of search and rescue in disaster relief. *Simulation Modelling Practice and Theory, 82*, 132–146.
48. Timm, I. J., Reuter, L., & Berndt, J. O. (2020). Role assignment adaptation: An intentional forgetting approach. *Proceedings of the 53rd Hawaii International Conference on System Sciences* (pp. 4786–4795).
49. Timm, I. J., Reuter, L., Berndt, J. O., Ulfert, A., Ellwart, T., & Antoni, C. H. (2019). Analyzing the effects of role configuration in logistics processes using multiagent-based simulation: An interdisciplinary approach. In *52nd Hawaii International Conference on System Sciences, HICSS 2019, Grand Wailea, Maui, Hawaii* (pp. 1–10). ScholarSpace/AIS Electronic Library (AISeL).
50. Valette, M., Gaudou, B., Longin, D., & Taillandier, P. (2018). Modeling a real-case situation of Egress using BDI agents with emotions and social skills. In *International Conference on Principles and Practice of Multi-Agent Systems* (pp. 3–18). Springer.

51. Wickens, C., Kramer, A., Vanasse, L., & Donchin, E. (1983). Performance of concurrent tasks: A psychophysiological analysis of the reciprocity of information-processing resources. *Science, 221*(4615), 1080–1082.
52. Wooldridge, M. J. (2000). *Reasoning about rational agents*. The MIT Press.
53. Wooldridge, M. J., & Jennings, N. R. (1999). The cooperative problem-solving process. *Journal of Logic and Computation, 9*(4), 563–592.
54. Zagorecki, A., Ko, K., & Comfort, L. K. (2010). Interorganizational information exchange and efficiency: Organizational performance in emergency environments. *Journal of Artificial Societies and Social Simulation*, *13*(3).
55. Zhao, B., & Tang, T. (2020). Evaluating the learning performance of emergency procedures for operators in metro based on a cognition model. *Promet-Traffic & Transportation, 32*(3), 335–345.