Please quote as: Bittner, E. A. C., Oeste-Reiß, S., Kirmse, R., Poser, M. & Wiethof, C. (2024). Cognitive Load Theory Approach to Hybrid Intelligence: Tackling the Dual Aim of Task Performance and Learning. Forty-Fifth International Conference on Information Systems (ICIS), Bangkok, Thailand.

Cognitive Load Theory Approach to Hybrid Intelligence: Tackling the Dual Aim of Task Performance and Learning

Completed Research Paper

Eva A. C. Bittner
University of Hamburg
Hamburg, Germany
eva.bittner@uni-hamburg.de

Sarah Oeste-Reiß
University of Kassel
Kassel, Germany
oeste-reiss@uni-kassel.de

Rosemarie Kirmse, Mathis Poser, Christina Wiethof
University of Hamburg
Hamburg, Germany
firstname.lastname@uni-hamburg.de

Abstract

Knowledge workers in information-rich work environments face cognitive challenges as they must deal with multitasking, interruptions, and time pressure. In domains like customer support with high turnover rates and increasingly diverse and complex products, employees need to rapidly develop from novices to experts while showing high task performance. The objective of this paper is to develop design knowledge for hybrid intelligence systems that tackle the dual aim of task performance and learning in knowledge work. We follow a design science research approach and build on theoretical and empirical knowledge on cognitive load theory. We propose a task-user matrix that classifies expertise and task difficulty to identify cognitive challenges. We develop four intervention strategies in the form of design patterns specified through design principles to address these challenges in system design. A pattern evaluation with system developers initially supports the effectiveness, plausibility and feasibility of our patterns.

Keywords: Cognitive Load Theory, Augmented Intelligence, Hybrid Intelligence, Design Science, Pattern, Design Principle, Knowledge Work

Introduction

Knowledge work is characterized by multichannel and information-rich work environments (Dabbish et al. 2011; Wacjman & Rose, 2011). Knowledge workers are often burdened with multitasking, interruptions (Dabbish et al., 2011; Wacjman & Rose, 2011) and real-time dynamic decision making, which lead to high cognitive strains and demand for cognitive support (Lerch & Harter 2001). An illustrative example is the customer service domain. Service employees (SEs) are expected to solve customer problems about products and services that are gaining in diversity and complexity, often under time pressure. Due to high turnover rates, lack of skilled workforce and cost pressure, pre-training of new SEs is often kept to a minimum or cannot prepare these workers for the multitude of problems to be solved. Therefore, SEs are regularly asked to learn on the job continuously, while performing their problem-solving tasks. Furthermore, new SEs accrue expertise over time. Thus, their learning needs evolve, as do their performance support requirements. This concurrent dual aim of performance and learning might create even more cognitive strains that need to be dealt with. Intelligent assistance systems are a promising class of technology to augment human knowledge work (Poser et al. 2022a, 2022b). Such systems ease some of these burdens, using the potential of augmented intelligence, i.e., a person uses or collaborates with an intelligent system to improve, accelerate, and/or support their own work (Kolbjørnsrud, 2024). In the context of mutual learning, the complementary strengths of human and artificial intelligence (AI) are combined to create a hybrid intelligence (Dellermann et al., 2019). Hybrid-intelligence-systems (HIS) are characterized through joint work, with the aim of enhancing the performance of each individual agent and fostering mutual learning. HIS combine human intelligence and artificial intelligence in an everlasting learning cycle (Wiethof & Bittner 2021). However, given the relative novelty of the concept and the availability of powerful assistance tools that enable HIS in practice, much knowledge still must be accumulated on how to implement concurrent performance and learning according to the users' needs. Both problem solving performance and learning require cognitive resources of the user. The amount of cognitive resources needed to perform a given task can differ greatly depending on the user's level of expertise. While a well-designed AI based system has the potential to provide just the right amount of information at the right time in a cognitive load (CL) saving form, bad system design can even impose further CL. In the worst case, users (especially novices) might suffer from overload that prevents both task completion and learning. While recent work on HIS acknowledges the importance of CL in HIS (Hemmer et al. 2021, Yao et a. 2024) and its diversity across individuals and conditions (Ye et al. 2022), there is demand toward effective solutions for cognitive fit between user and system (Rzepka & Berger 2018). In search for guidance on how to design HIS in a CL balanced way, we address the following ***research question***: *How can research on cognitive load inform the design of HIS for users of diverse expertise levels with the dual aim of learning and task performance?*

To answer this question, we ground our work on HIS and cognitive load theory (CLT). We develop a task-user matrix in which we classify types of knowledge workers according to their expertise levels and types of tasks according to their difficulty. This enables us to identify and organize challenges and define intervention strategies formalized as design patterns. To evaluate the design patterns, we use the evaluation framework according to Petter et al. (2010). Results show that the patterns are evaluated as plausible, effective, and feasible by experts (i.e., HIS developers). Suggestions for improvement from four experts have been used to refine and contextualize the patterns. Thus, we contribute four design patterns in the form of distinct strategies for specific challenges, operationalized via design principles (DPs). These can be applied by HIS developers to guide their design for task and user specific assistance. This way, we make established principles from CLT accessible to HIS designers and researchers. In the long run, our work should contribute to inform the design of adaptive HIS that provide individual assistance to knowledge workers depending on their current expertise level, task and resulting CL.

Related Work

Hybrid Intelligence Systems and Their Dual Aim of Performance and Learning

Hybrid intelligence (HI) refers to the ability to achieve complex goals by combining human and artificial intellect and capabilities, thereby reaching superior results to those each of them could have accomplished separately, and continuously improve by learning from each other (Akata et al., 2020; Bahle et al., 2016; Dellermann et al., 2019). This definition highlights the dual aim of superior performance and mutual learning in HI through augmentation of human intelligence (Akata et al., 2020). To leverage HI, so-called Hybrid Intelligence Systems (HIS) combine "Human-in-the-loop" approaches (HITL) and "Computer-in-the-loop" learning approaches (CITL). The focus of HITL approaches is to advance computers and more precisely to foster lifelong learning of machines through human intelligence (Dellermann et al., 2019; Shneiderman 2020; Wiethof & Bittner, 2021). HIS are *learning* as they extend their knowledge base in a self-organized way; they are *collaborative* as the learning process involves reciprocal interactions between technical systems and humans working towards a common goal; they are *interactive* as there is a bidirectional information exchange. This leads to a new paradigm, a value-co-creation of humans and systems that has the potential to tackle the dual aims of performance and learning (Calma et al., 2016). In that context, literature distinguishes between augmented intelligence and augmented collective intelligence (Kolbjørnsrud, 2024). Augmented intelligence refers to a (1:1) human-AI collaboration by which the HIS supports one human agent. In the case of augmented collective intelligence there is a (n:1) human-AI collaboration by which the AI system supports the collaboration of a group of humans. The paper's focus is augmented intelligence, which is pervasive in knowledge work. In the context of our illustrative customer service example, a SE might receive customer requests via chat or e-mail. The SE gets assistance from an AI system, e.g., information or suggested answers to respond to a customer request. Depending on their current domain expertise, a novice SE will benefit more by learning, e.g. about the company's product portfolio and how to answer unfamiliar types of customer requests, whereas an experienced SE will benefit more from performance gains through using prepared answer fragments. The same customer request will be differently cognitively challenging for these SEs, resulting in different augmentation needs in terms of content and presentation and different mental capacity available for performing and learning. To inform design decisions, we turn to CLT as a theoretical lens that has been used extensively to study the effective use of limited human cognitive resources.

Cognitive Load Theory

CLT was developed as a theory of instructional design for researchers and educators. It consists of aspects of human cognitive architecture that are relevant to instruction along with the instructional consequences that flow from the architecture. "*The ultimate aim of the theory is to use our knowledge of human cognition to derive instructional design principles"* (Sweller et al. 2011, p. vii). Research on CL has built up a substantial amount of such design principles and validated them in a multitude of empirical studies. Sweller et al. (2011) provide a review of the theory, its mechanisms, CL effects and their instructional implications. A central foundation of CLT is, that *"human **working memory** is limited in capacity and duration, if dealing with novel information but unlimited in capacity and duration, if dealing with familiar information previously stored in a very large long term memory"* (Sweller et al., 2011, p. vii). The primary function of working memory is to process (a limited amount of) environmental information to store and retrieve knowledge in and from long-term memory. In working memory, one can hold seven information elements in parallel for 5-20 seconds. Thus, fundamental issue arises from this limited capacity to process information. The working memory serves as bottleneck between the environment and human's long-term memory (Sweller et al. 2011, p. 42). If all humans possess the same working memory capacity, it becomes interesting to explore how some individuals can process more complex and challenging information than others. Levels of CL are determined by **element interactivity**. Interacting elements must be processed simultaneously in working memory because they are logically related. An element is anything that needs to be or has been learned or processed. Elements are characteristically **schemas** that often consist of sub-schemas or sub-elements. Prior to a schema being acquired, those sub-elements must be processed as individual elements in working memory. When elements have been incorporated into a schema, that schema can be treated as a single element. Thus, learning reduces working memory load by converting multiple lower-level schemas into a smaller number of higher-level schemas or even a single schema that can be treated as a single entity. (Sweller et al. 2011, p.58f.). Low element interactivity refers to material that can be processed in one element at a time. This requires little working memory capacity. High element interactivity refers to material where elements interact and must be processed simultaneously. It requires more working memory capacity until interacting elements have been incorporated into a schema after learning (Sweller et al. 2011, p. 58f.). Humans are confronted with a variety of problem-solving tasks– some are overwhelming, some are underwhelming. The question thus arises as to the reasons for this phenomenon. In the notion of CLT, the **difficulty of a task** can both be determined by the number of information elements to be processed and their interactivity. Many interacting elements can be impossibly difficult to process, especially for novices who cannot build on schemas in the domain (Sweller et al., 2011, p. 61). In our work, we focus on tasks with high element interactivity. These are common and challenging in knowledge work as well as in our illustrative domain of customer support. SEs need to understand and solve customer requests that consist of several information elements. An advanced SE may store multiple information elements in long term memory as a higher-level schema. A novice SE may store the same information as individual, unconnected elements. When it comes to retrieving that information from long-term memory into the working memory, the advanced SE will have free capacity as they only hold one element in working memory. The novice SE, in contrast, will have less working memory capacity as they hold multiple elements in working memory. This, consequently, determines how working memory is utilized and whether there is free capacity, e.g., for engaging in learning, or not. Summing up, humans who can process complex information have constructed higher-level schemas that they retrieve from long-term memory. Humans who have not constructed such schemas perceive cognitive overload more easily. The question now arises, how people can be supported in building and retrieving schemas. **Intrinsic, extraneous, and germane CL:** CLT distinguishes between different types of CL that regularly occur simultaneously and must be dealt with within the cognitive capacity of a human. CLT is concerned with procedures for presenting novel information to individuals in a manner that reduces any unnecessary CL while increasing those aspects of CL that foster learning (Sweller et al., 2011), p. 56). In that respect, the "load" imposed on working memory by instructional information can be divided into two categories – i.e., intrinsic CL and extraneous CL (Paas et al., 2003, 2004; Sweller et al., 1998; van Merriënboer & Sweller, 2005). Intrinsic CL is imposed by the nature of the material, i.e., the difficulty of the basic structure of the information that the learner needs to acquire for extending their knowledge (i.e., incorporating elements into schemas) and thus, achieving learning goals (Sweller et al., 2011, p. 57). In most instructional settings and prior CL studies, the information to be learned is predetermined by the learning task. Reducing intrinsic CL by altering the nature of what is learned may be an important instructional technique, but in most cases its utility is likely to be temporary (Sweller et al., 2011, p. 64f.). Especially if learning is expected to happen on the job, manipulating the work and learning task might not be feasible. In contrast, extraneous CL results from the nature of the instructional design to present the material, i.e., the way the information is presented or the design of activities in which learners must engage, which under many circumstances can be unnecessary (Sweller et al. 2011, p. 57). Most efforts building on CLT aim at reducing extraneous CL to free working memory capacity for engaging in learning and storing knowledge in long-term memory, so-called germane CL (Sweller et al. 2011, p. 68). Imagine an incoming customer request that contains a given number of information **elements**, e.g. symptoms of a problem experienced with a specific product type and edition in a certain usage context. These elements must be considered in combination, as the problem with the complex product only occurs under specific usage conditions (high **element interactivity**). If this message is routed to an expert SE, they might only need to process a few key words from the customer's message to identify a potential solution candidate, relying on problem-solution **schemas** from previous cases in their long-term memory. The task would only induce moderate intrinsic CL on their working memory. A novice SE confronted with the very same request must process each information element separately, as they do not know yet, which elements are needed for the solution. They might even need to consult further information sources, resulting in high intrinsic CL from the task. While the intrinsic CL is predetermined by the given combination of the customer request and SE's expertise/schemas, the additional extraneous CL is subject to the HIS design. This could mean that an AI assistant presents potential solution candidates to a novice SE in the same chat window where they interact with the customer (physical integration, see Split Attention Effect), resulting in less CL for integrating information elements and searching for solutions. For more advanced SEs, this might imply reducing the level of detail of AI suggestions or hiding explanations (see Guidance Fading Effect) to avoid overwhelming them with information they already know. (See Banerjee et al. 2023 for an adaptive HIS design.)

Design Science Research Approach

We follow a Design Science Research approach (Gregor and Hevner, 2013) and structure our activities along ***Hevner's three cycle view*** (2007)**.** In the introduction, we opened the *relevance cycle*. We examined the environment and shed light on the challenge of tackling the dual aim of learning and performance for knowledge workers using the illustrative domain of customer support. In the related work section, we opened the *rigor cycle* and introduced scholarly literature on HIS and CLT that serve as knowledge base. We use CLT as our guiding kernel theory. Literature on CLT is used in three ways to guide the design of our central artefact. First, knowledge on task difficulty and expertise levels of potential HIS users is derived from CLT research. This helps us to differentiate potential task-user combinations that might come with different challenges. Second, we identify four core challenges that HIS users are faced with, depending on the combination of their level of experience and current task to be solved. Third, we derive design knowledge for addressing the four challenges in the form of design patterns grounded in DPs. In the following section, we open the *design cycle* and present our artefact, consisting of the task-user matrix for identifying prevalent challenges for specific task-user combinations and, the design patterns that provide intervention strategies to cope with these challenges. Each design pattern represents a bundle of a challenge (i.e., specification of a recurring problem) and an intervention strategy that is specified by a set of DPs. The description of the DPs adheres to the documentation conventions suggested by Gregor et al., (2020).

To evaluate our patterns, we conducted a ***pattern design and evaluation*** procedure suggested by Petter et al. (2010). A pattern describes a problem, which occurs repeatedly in our environment. Then it describes the core of the solution to that problem, in such a way that one can use this solution a million times over, without ever doing it the same way twice (Alexander et al. 1977). To develop and evaluate patterns one needs to consider that there is a natural life-cycle to patterns. The discovery, description and validation of patterns are concomitant activities (Petter et al., 2010). Thus, the evaluation of patterns is a continuous improvement activity. The pattern-life-cycle consists of a development, deployment and use phase that each offer unique opportunities for evaluation (Petter et al., 2010). This paper focuses on the *development* and *deployment phases* according to three established pattern evaluation criteria (plausibility, effectiveness, feasibility, see section "Pattern Evaluation" for detailed methods and results). The criterion "*plausible*" aims to assess whether the pattern is "sensible considering the current understanding of the domain" (Petter et al., 2010). Screening for plausibility arises though the process of building the pattern from the judgement of its creators (Petter et al., 2010) and system developers. The criterion "*effective*" aims to assess whether the "pattern is described in language that is understandable; root causes of the problem are identified and addressed by the recommended solution" (Petter et al., 2010). Effective patterns are those that are complete while being economically completed, e.g., include a meaningful name, problem statement, the context of the problem, the applicable forces and constraints, a solution, one or more examples, the context after the pattern has been applied, the rationale, a listing of related patterns and known uses of the pattern (Petter et al., 2010). The criterion "*feasible*" aims to assess whether the "pattern can be operationalized or implemented as described" (Petter et al., 2010). Feasible patterns have the quality of being implementable or operationalizable. Therefore, feasibility ensures that the pattern can be used. There are two more pattern evaluation criteria (i.e., predictive, reliable) that are subject to future work during the use phase of the pattern-life-cycle (Petter et al., 2010). We report the results from the evaluation with HIS developers in the evaluation section, with which we close the design cycle. In the discussion section, we reflect on the results, limitations, contributions and provide an outlook. The last section contains a brief conclusion.

Artefact: Design Patterns for Different Task-User Constellations

Our artefact is described in terms of (a) recurring challenges that knowledge workers of different expertise levels face in their work on tasks of different complexity, systematized in a *task-user matrix* and (b) a set of *four design patterns in the form of intervention strategies* addressing these challenges by making use of known CL effects with empirical support from CL research described in the form of DPs.

Task-User Matrix with Resulting Challenges

<table>
<tr><th colspan="2" rowspan="2"></th><th colspan="3">Prior Knowledge/ Expertise</th></tr>
<tr><th>Novice</th><th>Advanced</th><th>Expert</th></tr>
<tr><th rowspan="3">Task difficulty</th><td>Low</td><td>Onboarding Challenge: Instruction Strategy</td><td></td><td rowspan="2">No cognitive overload Intelligent Automation</td></tr>
<tr><td>Medium</td><td rowspan="2">Overburdening by Task Challenge: Simplification Strategy</td><td>Upskilling Challenge: Fading Guidance Strategy</td></tr>
<tr><td>High</td><td></td><td>Lifelong Learning Challenge: Collab. Augmentation S.</td></tr>
<tr><td colspan="5">Table 1. Task-User Matrix</td></tr>
</table>

Overall, CL results from the perceived difficulty of the task in terms of number and interactivity of information elements. This is closely connected with the expertise of the individual that is determined by their schemas of the domain. More experienced users might face different challenges in orchestrating performance and learning than novices. For example, they might need less information and different instructional strategies than novices, because information that is essential for novices may become redundant for experts (Sweller et al., 2011, p. 152). We build on this conceptualization in the development of our task-user matrix (see Table 1). In the light of the known processing limitations of humans' working memories (Cowan, 2012, 2001), we deduce a task-user matrix based on the conceptualizations in CLT literature (Sweller et al. 2011). The vertical axis refers to the user's perceived **task difficulty**.The horizontal axis refers to the **user's prior knowledge respective expertise**. *Task difficulty* refers to the element interactivity (i.e., number and interactivity of elements to be processes simultaneously in working memory) of the work task. Task difficulty can be lowered by reducing the number of elements that have to be processed at the same time, e.g., by converting several elements into one schema. *Low difficulty tasks* are those tasks that contain two to three or less (Sweller et al., 2011, p. 43) information elements that need to be processed simultaneously in working memory. The solution space contains a small and limited number of options. *Medium difficulty tasks* are such tasks, for which more than 3 information elements need to be processed. The solution space contains a medium number of options to explore. *High difficulty tasks* accordingly refer to a high number of information elements, i.e., way more than 3 that need to be processed simultaneously. The solution space contains numerous or infinite options to explore. As information elements convert into lower-level schemas, and those convert into higher-level schemas, an individual develops factual, conceptual, procedural, and metacognitive knowledge (Krathwohl, 2002) in the work domain. Concerning the cognitive user characteristics, we identified three user archetypes with different prior knowledge and expertise levels (i.e., Novice, Advanced, Expert Users) that Sweller et al. (2011) and the primary sources refer to. *Novices* have little or no relevant domain-specific knowledge. They cannot retrieve schemas from long-term memory. Most information needed to solve the task is novel to them. In line with CLT, they must use working memory capacity to solve problems. Novices tend to work backwards from the goal to solve problems using means-end strategies (Sweller et al., 2011). They must build up conceptual knowledge, i.e., they need to learn the interrelationships among the basic elements within a larger structure that enable them to function together (e.g., classifications and categories; principles and generalizations; theories, models and structures). In customer service HIS, novice users are e.g., newly hired SEs before their initial training on the company's service processes and the HIS. *Advanced users* have at least medium levels of domain-specific knowledge, i.e., initial schemas of the domain. They can use some basic conceptual and procedural knowledge (i.e., retrieve schemas) to solve problems. They apply a mixture of novice and expert problem-solving strategies in terms of CLT (Sweller et al., 2011) depending on the task. They should build up further conceptual and procedural knowledge. They need to learn the subject-specific skills and algorithms, methods and techniques and criteria to determine when to use appropriate procedures. In customer service HIS, advanced users are those who know the basics about the process and product(s) by either having completed a training or having executed customer interactions already. *Expert users* have high levels of domain-specific knowledge and high-level schemas of the domain to relate new information to. Thus, some information presented to solve a task may be redundant to them. They tend to work forward from the givens using schemas. In customer service HIS, expert users are SEs, who have worked in the organization for an extended time and have experienced and solved a variety of different customer requests, which enables them to solve most requests from experience and transfer their knowledge to unfamiliar requests. To transition from novice to expert level, knowledge workers must master three consecutive challenges. For each user archetype, we propose a core challenge. These constitute recurring problems in practice from the authors' experience in different HIS contexts, later validated by the HIS developers in the expert evaluation. *Novice* users face the *"Onboarding Challenge"*. They lack even basic knowledge of the domain and need fundamental training that is presented in forms that induce low CL to empower them for work tasks of low difficulty. *Advanced* users face the *"Upskilling Challenge"*. They need to juggle task performance and qualification for more complex tasks. *Expert* users face the *"Lifelong Learning Challenge"*, as they must sustain and optimize their ability to work in a dynamically changing work environment. By combining the user and the task archetypes, we discovered a fourth challenge – the *"Overburdening-by-Task Challenge"* that especially *novice and advanced* users face. In this challenge, the task difficulty and resulting intrinsic CL alone exceeds cognitive capacity of the user. They have no capacity to process the information and master the problem-solving task. In this case, even minimizing the extraneous CL would not result in any cognitive capacity for learning.

Design Patterns: Intervention Strategies for Challenges

To develop a dedicated solution to each challenge, we conducted a qualitative content analysis of the definitions and reviewed empirical studies on the 13 CL effects identified by Sweller et al. (2011). We mapped each effect to the user expertise and task difficulty levels at which it manifests (see Table 1) according to the original literature, allowing us to assign the effects to the four challenges. We also extracted information from the studies on promising intervention designs for the effects and documented them in the form of 13 initial DPs. The formulation of the DPs was conducted by the authors of the paper and refined iteratively through discussion considering the literature. The documentation format adheres to the 'components of a DP schema' suggested by Gregor et al. (2020). To provide concrete guidance for HIS designers, we defined a dedicated intervention strategy containing all DPs applicable to a specific challenge/task-user combination: *"instruction strategy"* to master the onboarding challenge; *"guidance strategy"* for the upskilling challenge; *"collaborative augmentation strategy"* for the lifelong-learning challenge; and *"simplification strategy"* for the overburdening-by-task challenge. Bundles of a challenge and intervention strategy specified by a set of DPs, can be considered patterns. Due to space restrictions, we present the final version of the patterns only, i.e., with implemented refinements resulting from the pattern evaluation presented in the later parts of the paper.

Instruction Strategy (Onboarding Challenge)

The Instruction strategy is directed at users with no or little domain knowledge. It should be applied to provide novices with initial instruction to build up basic knowledge and lower-level schema in a sandbox pre-training setting before being confronted with real tasks in live work settings. It aims to reduce extraneous CL imposed by provided information to free cognitive resources for learning. It is characterized by an instructional design adjusted to users' expertise and provide worked examples and tasks with low difficulty.

Effect	Design Principles
Goal-Free Effect (Sweller, 1988; Sweller et. al., 1983)	***DP1:*** *To* allow ***novice users*** to build initial schemas, *consider* giving users open ended tasks in the form of goal-free exploration to build schemas, instead of closed goal tasks, *because* in conventional problem-solving tasks, novices use means-end strategies from the goal backwards that induce high CL
Worked Example Effect (Atkinson et al., 2000; Trafton & Reiser, 1993; Oksa et al., 2010; Schwonke et al., 2009).	***DP2:*** *To* allow ***novice users*** to prepare better for subsequent problem-solving, *consider* 1. presenting worked examples in the form of a problem statement and procedure for solving the problem and then immediately follow each example by asking the user to solve a similar problem; 2. informing users before the presentation of worked examples, that they will need to solve problems on their own afterwards; 3. providing users with explanatory notes integrated into the original text, *because* it may foster comprehension, decrease CL, and enable users to study and emulate an expert's problem-solving model
Problem Completion Effect (Van Merriënboer, 1990)	***DP3:*** *To* allow ***novice users*** to engage in active learning, *consider* 1. giving users partial worked examples, where the learner must attend to and process the worked-out steps and then respond to and complete the last incomplete step(s); 2. increasing the number of steps to be solved in accordance to learning progress (from the goal backwards), *because* it can help to continuously build up problem-solving skills with manageable CL.
Split-Attention Effect (Bétrancourt & Bisseret, 1998; Florax & Ploetzner, 2010; Ginns, 2006; Kalyuga et al., 1999; Mayer, 2009)	***DP4:*** *To* allow ***novice users*** to mentally integrate several sources of information with reduced extraneous CL, *in case that* sources of information need to be processed simultaneously, *consider one or more of the following:* 1. physically integrating; 2. temporally integrating; 3. linking with color coding; 4. linking with pop-up texts; 5. linking sources of information with numbering/ labeling, *because* dispersed information may increase extraneous CL.
Modality Effect (Atkinson, 2002; Mayer, 2009; Moreno et al., 2001)	***DP5:*** *To* allow ***novice users*** to increase their effectively used working memory capacity, *a) in case that* information elements need to be processed simultaneously and that the individual information elements are limited in length and complexity, *consider* 1. using a voice channel to present some of the elements; 2. using animated agents to provide assistance in a multimodal way, *because* theories of working memory suggest, that we have two different, partially independent processors for dealing with visual and auditory information. *b) in case that* textual information elements are lengthy and complex, *consider* presenting them in permanent written form, *because* they cannot be held and processed in working memory. *c) in case that* visual information elements such as illustrations or graphics are very complex, *consider* using cueing or signaling, *because* this may help users to focus on those parts of the visual display being referred to by the auditory information.

Expertise Reversal Effect (Kalyuga et al., 2003; Kalyuga, 2007; Van Gog et al., 2008)	***DP6:*** *To* allow ***novice users*** to get the instructional guidance that matches their novice level of expertise, *consider* 1. providing detailed, direct instructional support to novice users, preferably, in integrated or dual-modality formats; 2. providing information in static diagrams showing all information in one place; 3. providing process-oriented worked examples rather than product-oriented worked examples, *because* explicit instruction may provide an effective substitute for the missing knowledge-based executive function of novices. The initial use of worked examples and isolated elements constitute a form of pre-training.
Transient Information Effect (Sweller et al., 2011, p. 219)	***DP7: To*** avoid a loss of learning by ***novice users***, *in case that* information with high element interactivity is presented in transient form, i.e., it disappears from the user's view or hearing before they have time to adequately process it or link it with new information in working memory*, consider* 1. presenting long, complex information in permanent written form rather than in spoken form to avoid transience; 2. segmenting animations into smaller fragments; 3. enabling learner control to slow down or stop an animation, *because* transience can create excess working memory load.

Table 2. Design Principles for Instruction Strategy

Guidance Strategy (Upskilling Challenge)

The Guidance strategy is directed at users with at least some basic domain knowledge. It should be applied to guide users in concurrently building up further knowledge and schemas while solving real problem tasks in their work environment. Thereby, it aims to reduce the extraneous CL imposed by the HIS and the provided information. The strategy should be characterized by an instructional design that is adjusted to the user's dynamic expertise level and that adapts to rising expertise levels by fading guidance.

Effect	Design Principles
Split-Attention Effect (Bétrancourt & Bisseret, 1998; Florax & Ploetzner, 2010; Ginns, 2006; Kalyuga et al., 1999; Mayer, 2009)	***DP4:*** *To* allow ***advanced users*** to mentally integrate several sources of information with reduced extraneous CL, *in case that* sources of information need to be processed simultaneously, *consider* 1. – 2. (see **DP4**, **Table 2. Design Principles for Instruction Strategy**)
Modality Effect (Atkinson, 2002; Mayer, 2009; Moreno et al., 2001)	***DP5:*** *To* allow ***advanced users*** to increase their effectively used working memory capacity, *consider* a) – c) (see **DP5,** Table 2)
Redundancy Effect (Sweller et al., 2011, p. 141)	***DP8:*** *To* allow ***advanced users*** to focus on processing only necessary information, *in case that* information elements are redundant or have become redundant to a specific user in reference to their currently available schemas, *consider* 1. avoiding presenting unnecessary information not required for solving the task or learning; 2. avoiding presenting the same information in different modalities, such as written and spoken form; 3. reducing presented information with rising level of expertise; 4. hiding potentially redundant information from the user and provide only upon request, *because* mentally integrating unnecessary information with essential information induces excess extraneous CL.

Expertise Reversal Effect (Kalyuga et al., 2003; Kalyuga, 2007; Van Gog et al., 2008)	***DP6:*** *To* allow ***advanced users*** to get the instructional guidance in the work environment that matches their advanced level of expertise, *consider* 1. applying a mix of direct instruction (e.g. worked examples) and problem-solving practice with reduced support; 2. applying exploratory instructional methods over worked examples for advanced users; 3. reducing process information for advanced users; 4. using animated over static information presentation for advanced users, *because* information beneficial to novices becomes redundant to advanced. At intermediate levels of expertise, an executive function based on long-term memory knowledge when dealing with familiar information can be complemented by direct instruction when dealing with unfamiliar elements of information.
Guidance Fading Effect (Atkinson et al., 2000; Kalyuga, 2006, 2008; Kalyuga & Sweller 2004, 2005; Renkl, 1997; Renkl et al., 2000, 2007)	***DP9:*** *To* allow ***advanced users*** to get the guidance that meets their expertise level and does not produce redundancy, *in case that* users have increased their expertise level (i.e. built up schemas of information elements), *consider* 1. gradually fading worked examples in a coordinated series of completion problems by replacing worked-out steps progressively with problem-solving steps to complete either in a backward-fading strategy, whereby the user must conduct the final problem-solving step, which might impose lower CL than a forward-fading strategy or fading the step(s) that the user should learn most about; 2. integrating a rapid diagnostic assessment of the user's current expertise level to provide an adaptive fading strategy that is adapted to the skill levels of the individual; 3. using intelligent assistants as user interface for providing fading guidance, *because* with increasing learner expertise, knowledge held in long-term memory can be used to decrease the demands on working memory and free working memory resources for problem solving.
Imagination Effect (Cooper et al. 2001)	***DP10:*** *To* allow ***advanced users*** to increase the flexibility in using their learned knowledge under various circumstances and prepare for continuous learning in the field, *in case that* users have freed cognitive resources from other CL reducing procedures, *consider* encouraging advanced users to imagine complex concepts and procedures, *because* the mental reproduction of activities, procedures, or concepts can reduce extraneous CL that material with redundant information would impose on advanced users.
Self-Explanation Effect (Clark et al., 2006, p. 226; Renkl, 1999)	***DP11:*** *To* allow ***advanced users*** to increase the effectiveness of prior instruction, *consider* encouraging advanced users to engage in self-explanation, defined as "a mental dialogue that learners have when studying a worked example that helps them understand the example and build a schema from it".
Element Interactivity - Effect (Sweller, 2010)	***DP12:*** *To* allow ***advanced users*** to reduce the effective intrinsic CL due to element interactivity of the task, *consider* supporting users to use pre-existing schemas to incorporate ("chunk") interacting **elements, *because*** chunking interrelated elements can decrease working memory load.
Transient Information Effect (Hasler et al., 2007; Mayer et al., 1999; Schwan & Riempp, 2004)	***DP7:*** *To* avoid a loss of learning by ***advanced users,*** *in case that* information with high element interactivity is presented in transient form, i.e., it disappears from the user's view or hearing before they have time to adequately process it or link it with new information in working memory, *consider* 1. – 3. (see **DP7**, Table 2)

Table 3. Design Principles for Guidance Strategy

Collaborative Augmentation Strategy (Lifelong Learning Challenge)

This strategy is directed at experts in the domain and work processes. It should guide them to apply their expertise for effective problem solving of complex tasks in their work environment while continuously updating and extending their knowledge and schemas. Thereby, it aims to reduce extraneous CL imposed by, e.g., redundancy of information. The strategy is characterized by the HIS augmenting users' expertise and aiming to solve familiar problems efficiently to free cognitive resources for continuous learning.

Effect	Design Principles
Modality Effect (Atkinson, 2002; Mayer, 2009; Moreno et al., 2001)	***DP5:*** *To* allow ***expert users*** to increase their effectively used working memory capacity, *consider* a) – c) (see DP5, Table 2)
Redundancy Effect (Sweller et al., 2011, p. 141)	***DP8:*** *To* allow ***expert users*** to focus on processing only necessary information, *in case that* information elements are redundant or have become redundant to a specific user in reference to their available current expertise level (i.e., schemas), *consider* 1. – 3. (**see DP8**, Table 3).
Expertise Reversal Effect (Kalyuga et al., 2003; Kalyuga, 2007; Van Gog et al., 2008)	***DP6:*** *To* allow ***expert users*** to get only the instructional guidance in the work environment that matches their level of expertise, *consider* 1. not using worked examples with experts; 2. applying problem-solving for expert users; 3. reducing process information for expert users; 4. using animated over static information presentation for expert users; 5. training new technological skills concurrently with learning a specific new subject discipline in case of technological expertise, *because* information beneficial to novices becomes redundant to experts and much of the required knowledge may be available in long-term memory of experts.
Imagination Effect (Cooper et al. 2001)	***DP10:*** *To* allow ***expert users*** to increase the flexibility in using their learned knowledge under various circumstances and prepare for continuous learning in the field, see **DP10,** Table 3.
Self-Explanation Effect (Clark et al. 2006, p. 226; Renkl 1999)	***DP11:*** *To* allow ***expert users*** to increase the effectiveness of prior instruction, *consider* encouraging expert users to engage in self-explanation, see *DP11* (see Table 3)
Element Interactivity Effect (Sweller, 2010)	***DP12:*** *To* allow ***expert users*** to reduce the effective intrinsic CL due to element interactivity of the task, *consider* supporting users to use pre-existing schemas to incorporate ("chunk") interacting elements, *because* chunking interrelated elements can decrease working memory load.
Collective Working Memory Effect (Kirschner et al. 2009)	***DP13:*** *To* allow ***expert users*** to process complex tasks effectively, *in case that* users have sufficient cognitive capacity to handle the extra CL from transactions to organize collaboration, *consider* supporting experienced users to collaboratively solve tasks with other human users or artificial agents, *because* a group of actors may have an expanded processing capacity, as the intrinsic CL caused by a task can be subdivided across cooperating working memories by dividing the interacting elements.

Table 4. Design Principles for Collaborative Augmentation Strategy

Simplification Strategy (Overburdening by Task Challenge)

The Simplification strategy is applied, if the difficulty of work tasks exceeds the expertise of the user, leading to cognitive overload in problem-solving and leaving no cognitive resources for learning. It is applied to reduce the effective intrinsic CL through reducing task complexity to 2-3 interactive information elements. This can be achieved by substituting the real task with an artificial or simplified task for training purposes, by decomposing the task into sub tasks with reduced element interactivity to be handled sequentially, or by splitting sub tasks among several working memories (either users or artificial agents).

Effect	Design Principles
Goal-Free Effect (Sweller, 1988; Sweller et. al., 1983)	***DP1:*** *To* allow ***users*** to prepare for tasks by building up initial schemas in a pre-training, in *case that* even the real work tasks exceed their cognitive capacity, *consider* 1. giving users artificial open-ended tasks in the form of goal-free exploration to build schemas, instead of closed goal tasks; 2. giving users worked examples, *because* decoupling schema-building and problem-solving can reduce CL in each phase. This can be achieved by focusing on a simplified artificial task in pre-training that helps building the schemas that can be used in later problem-solving.
Element Interactivity Effect (Sweller, 2010)	***DP12:*** *To* reduce the intrinsic CL of a task for ***users***, *in case that* it exceeds the cognitive capacity of the user, *consider* decomposing the task into subtasks that can be handled individually to reduce element interactivity, *because* non-interactive sub-tasks can be treated sequentially.
Collective Working Memory Effect (Kirschner et al. 2009)	***DP13:*** *To* simplify a task for the individual ***user*** to a manageable level, *in case that* the intrinsic CL of the task exceeds the cognitive capacity of the user, and the task can be decomposed into sub tasks, and the transaction costs for organizing the collaboration are lower than the advantages gained by off-loading some of the elements to others, *consider* organizing collaboration with other users or artificial agents, *because* a group of actors may potentially have an expanded processing capacity, as the intrinsic CL caused by a task can be subdivided across cooperating working memories by dividing the interacting elements.

Table 5. Design Principles for Simplification Strategy

Pattern Evaluation

Evaluation Procedure

The evaluation of the ***development phase*** aims to assess the criteria "*plausible*" and "*effective*". It contains (1) a continuous expert review by the pattern creators considering CLT literature, reflected in the material presented in this paper. Furthermore, (2) a peer review by two HIS designers (plus two more in the deployment phase) was executed to validate the consistency of the approach with HIS work practice. Experts were approached based on their knowledge of HIS research, an information systems research background and several years of experience in HIS design. To generate initial insights on the "*plausibility*" of the patterns, i.e., to check whether the patterns are sensible considering the current understanding of the domain, two authors conducted an ongoing review during the development phase. We individually compared all developed artefacts with the body of literature on CLT, e.g., concerning the definition of concepts and the underlying model of the human mind. To generate initial insights on "*effectiveness*", we also ensured that pattern descriptions match the documentation conventions for DPs (Gregor et al. 2020). We discussed any upcoming issues and refined the artefact descriptions iteratively until deemed sensible. To further investigate findings toward the "*plausibility*" of our patterns and to examine the "*effectiveness*", we conducted a semi-structured peer review survey with two experts. The survey started with an introduction of the task-user matrix and the challenges as stimuli and context. In the first part of the survey, each intervention strategy and its underlying DPs were described. The HIS developers were asked to develop textual and mock-up examples for each DP's instantiation to foster active engagement with the DPs and check whether the DPs can be implemented. In the second part of the survey, the developers were asked to assess the plausibility, effectiveness (and feasibility, see deployment phase) of the described DPs for each intervention strategy on a 7-point Likert-Scale using the measurement instrument proposed by Petter et al. (2010). We collected all measures on the DP level to allow for a differentiated assessment and improvement of the patterns. Developers could also add qualitative remarks in the evaluation form for every DP. The evaluation of the ***deployment phase*** aims to assess the evaluation criteria "*effective*" and "*feasible*" (Petter et al., 2010). In this phase, the validation of the feasibility of the patterns for being used by system designers has been extended beyond the formulation of textual examples and mock-ups. The two HIS developers, who had already conducted the peer review, were instructed to use the design pattern to instantiate the guidance strategy in a software prototype of a HIS. They used an existing HIS prototype as a baseline that had been developed in a prior research project for a specific organization and applied the applicable DPs from the guidance strategy to derive an adapted user interface that should optimize CL for users that have just completed onboarding and should receive guidance by the systems during the first real customer interactions they conduct. The aim of this evaluation was to explore whether the patterns can be operationalized and implemented by developers as described. To complement the survey data with more perspectives, we conducted the survey with two more developers, who are experts in different HIS, e.g., for students in self-directed learning. The developers were asked to assess the plausibility, effectiveness, and feasibility of the DPs for each intervention strategy with the same instrument as mentioned above.

Evaluation Results

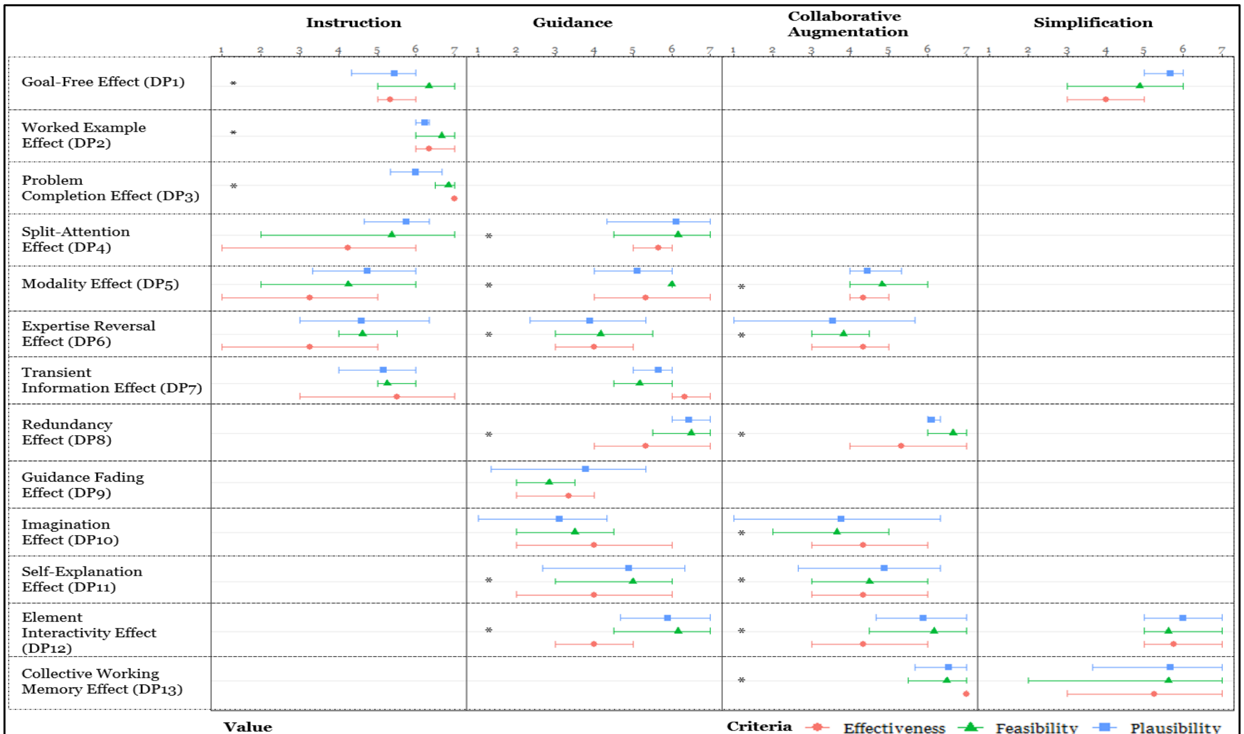

Figure 1. Means of Effectiveness, Feasibility and Plausibility across Expert Ratings (*n*=4), *indicates *n*=3, Error Bars indicate range

Figure 1 reveals that most of the DPs were on average rated above neutral in the respective criteria scales. High ratings and small ranges for DP1, DP2, and DP3 indicate a high agreement on the effectiveness, plausibility, and feasibility of those DPs within the Instruction Strategy among the experts. Similarly positive and homogenous ratings have been provided for DP7 in the Guidance Strategy, DP8 and DP13 in the Collaborative Augmentation Strategy, and DP12 in the Simplification. However, a rather heterogenous evaluation is depicted for DPs 4, 5, and 6 in the Instruction Strategy and DP10 and DP11 in the Guidance Strategy. The evaluation of DP10 in the Collaborative Augmentation Strategy and DP13 in the Simplification Strategy exhibit similar wide ranges of criteria ratings. Overall, during the development and deployment phase, the peer review revealed that the patterns are described in language that is understandable, root causes of the problem are identified and addressed by the recommended solutions. Nevertheless, we received recommendations for improvement. Required changes referred to a clarification of ambiguities and explanations of non-meaningful terms that raised clarification questions. These are already implemented in the design pattern version reported in this paper. We also removed cross-references between the DPs and strategies to allow for stand-alone usage of each pattern. As experts noted that they were not able to assess the completeness of some DPs, we added references to the primary literature. This supports users of a DP with more specific resources for design guidance. Where the description of the DP was considered as too short or unclear, we extended the description (e.g. DP10 ‘self-explanation effect’). A further refinement was to clarify the description of the concept “schema building”.

Discussion

In our study we introduce groundwork for developing HIS from a socio-technical point of view. Hybridizing human and artificial intelligence requires systems to create value for both humans and machines alike (Wiethof & Bittner 2021). Thus, HIS must intelligently augment users in their work processes to support the dual aim of performance and learning and ensure a cognitive fit with the user and their context (Rzepka & Berger 2018). However, it still underspecified in HIS literature what this means from a human cognition perspective (Ye et al. 2022). For that reason, we lay an important foundation and translate how top of previous research on CL (Sweller et al. 2011) can inform the design of HIS through the means of design patterns and DPs. System developers that use our suggested design patterns to build own HIS should consider following advice: *First*, it must be noted that the DPs are preliminary and evaluation of their actual impact in instantiated HIS is subject to future work. The patterns are in an initial state that cover those aspects that were derived from the considered CLT literature and validated by the consulted experts. *Second*, it should be considered that not all strategies address learning and performance to the same extent, but in the context of the task-user combinations' needs. This is an important insight that results from our matrix in line with the notion of cognitive fit (Rzepka & Berger 2018), as the need for learning (Wiethof & Bittner 2022) changes when employees gain expertise, as does the need for efficient performance (Rzepka & Berger 2018). *Third*, when system developers implement the patterns for their own purposes, we suggest using them more like a toolbox than a cooking recipe. Not all DPs might be applicable to all types of HIS and specifics may arise from the (real-world) application context, e.g., in terms of available user interface options. Therefore, not all DPs from an intervention strategy might be applicable or useful to all context as one expert noted *"(…) there might be restriction given the specific situations and contexts"*. From a user-centered perspective, HIS have the potential to enable learning on the job for humans with varying levels of expertise, where task performance and continuous learning are dual goals for the humans involved. Thus, the paper illustratively shows that it becomes a matter of design, how to allocate and sparsely use cognitive resources. Informed by extensive work on CLT from other domains (Sweller et al. 2011), we propose initial design knowledge in the form of design patterns for HIS design. The design patterns provide solution ideas for reducing extraneous CL to free capacity for learning and performance. What is new is that augmentation in a HIS might also challenge the common assumption that intrinsic CL by that respective task cannot be manipulated. Should the AI subsystem be able to assume certain tasks on behalf of the human, thereby facilitating a collaborative effort, it is possible that the overall intrinsic CL for the human may be altered. Intrinsic CL can also be reduced by the act of learning itself. This transformation marks a shift from novices to experts in a domain with impacts on both instructional needs and task performance capacities. It allows individuals to perceive tasks as trivially simple that previously were impossible or even inconceivable (Sweller et al., 2011, p.vii). Thus, the expertise dimension of our task-user matrix might also (and maybe more precisely) be interpreted as a continuum and trajectory of evolvement rather than distinct groups. In line with that notion, there are several DPs that are applicable to more than one design pattern. While their expertise increases, the users' objective is still mastering the dual aim of performance and learning, but they need different support in a HIS. Thus, a technical implication for future HIS research is providing real-time measurement of expertise or CL to enable personalized user augmentation. Especially modern (conversational) AI systems show high potentials to adapt their assistance to the current state of the specific user (see e.g. Wambsganss et al. 2021, Banerjee et al. 2023). In the long run, an adaptive HIS, where the AI system complements the current level of expertise and CL of the user would support learning on the job without interfering with task performance. With our artefact, we contribute toward making the cognitive foundations from CLT accessible to this timely research stream on adaptive systems (Ritz, 2024).

Limitations and Outlook for Future Research

Our work is not without limitations. First, the scope of our paper refers to the development and deployment phases of the pattern-life-cycle (Petter et al., 2010). Our solution is mainly developed in a theory-based and conceptual way. The evaluation ended at the first part of the deployment phase. Therefore, the actual effect of applying the DPs in HIS design on the levels of CL, learning and task performance need to be tested in future experimental and field studies. To investigate the evaluation criteria of "predictiveness" (i.e., pattern produces expected result) and "reliability" (i.e., pattern produces similar results regardless of the implementer and technique) an evaluation during the use phase of the pattern-life-cycle is necessary. Second, in the paper the description of the DPs adheres to the scholarly documentation conventions for DPs suggested by Gregor et al. (2020). For use by practitioners (i.e., system developers) it is requested to extend the documentation with visual examples. We found that some DPs are still underspecified in comparison to others, resulting from varying coverage in CLT literature and preliminary expert evaluation with mixed assessments, calling for more in-depth work on the individual DPs. Moreover, our expert evaluation supports the need for developing HIS that are adaptive. Experts perceived DP12 on the element interactivity effect (supporting users to use pre-existing schemas to chunk interacting elements) as abstract. This can be attributed to the difficult challenge of capturing human cognition in real-time. In this light, experts questioned *“How can we measure “live” in active HIS, if and when a task exceeds the human’s cognitive capacity?”* or *“How does one know unfamiliar elements of information to integrate them into worked examples?*”. Novel AI-based techniques are required to measure human’s expertise level in real-time. Some experts’ comments referred to the assessment that some DPs are difficult to implement in HIS with current technology (e.g., DP9 Realtime Guidance Fading). Although we refer to some literature that provides starting points for such a design, future research should assess the feasibility of the proposed solutions in different technological surroundings. Moreover, this indicates a need for developing new AI-based solutions. Next to this, there are hints that DPs might be combined for practical purposes (e.g., DP6 and DP2 on how to guide the contextualized use of worked examples and when to adapt or reduce worked examples with users of increasing expertise levels) in future iterations. For now, we kept DPs separate according to the CLT based effects they are grounded in to keep their origin transparent. However, this might be consolidated in future versions to ease application by practitioners.

Contributions and Conclusion

For developers of HIS, we propose a toolbox of four design patterns to design HIS that direct users’ cognitive resources in a more economical way. Through the accompanying task-user matrix, practitioners can be guided toward identifying potential challenges that may apply to their focal user group. A challenge is determined by the expertise level of a user and difficulty of the tasks. For each challenge, we present a set of DPs bundled in design patterns that can be considered for providing CL sensitive assistance that balances the dual aim of task performance and learning in HIS (Jones & Gregor 2007; Gregor & Hevner 2013). We make contributions toward a design theory for a CLT approach for HIS and outline its purpose and scope, kernel theory as well as principles of form and function represented by our design patterns (Jones & Gregor, 2007). Prior HIS studies have acknowledged CL as important factor (Ye et al. 2022, Poser et al. 2022a), but applied CLT only to a limited extent – to the best of our knowledge, e.g., the individual CL effects discussed in Sweller et al.(2011) have not been assessed in their entirety for applicability to IS design. Thus, we translated a vast amount of empirical evidence and theoretical knowledge on the functioning of the human mind and learning from CLT to the timely domain of HIS to make it accessible to IS researchers in that domain. With this, we contribute to the timely research stream on tracking, understanding and actively managing human factors such as cognitive load and fatigue (Yao et al. 2024) toward HIS that better fit the cognitive augmentation needs of knowledge workers (Rzepka & Berger 2018).

Acknowledgements

This research was funded by the German Federal Ministry of Education and Research(BMBF) in the context of the project HyMeKI (reference number: 01IS20057).

References

Akata, Z., Balliet, D., Rijke, M. de, Dignum, F., Dignum, V., . . . Welling, M. (2020). A research agenda for hybrid intelligence: Augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. *Computer*, *53*(8), 18–28.

Alexander, C., Ishikawa, S., Silverstein, M., Jacobson, M., King, I. F., & Shlomo, A. (1977). A pattern language: Towns, buildings, construction (48th printing). *Center for environmental structure: vol. 2*. Oxford University Press.

Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology*, *94*(2), 416–427.

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, *70*(2), 181–214.

Bahle, G., Calma, A., Leimeister, J. M., Lukowicz, P., Oeste-Reiss, S., Reitmaier, T., Schmidt, A., Sick, B., Stumme, G., & Zweig, K. A. (2016). Lifelong learning and collaboration of smart technical systems in open-ended environments -- Opportunistic collaborative interactive learning. *Proceedings of the 2016 IEEE International Conference on Autonomic Computing (ICAC)* (pp. 315–324).

Banerjee, D., Poser, M., Wiethof, C., Subramanian, V. S., Paucar, R., Bittner, E. A., & Biemann, C. (2023). A system for human-AI collaboration for online customer support. *arXiv preprint arXiv:2301.12158*.

Bétrancourt, M., & Bisseret, A. (1998). Integrating textual and pictorial information via pop-up windows: An experimental study. *Behaviour & Information Technology*, *17*(5), 263–273.

Calma, A., Leimeister, J. M., Lukowicz, P., Oeste-Reiss, S., Reitmaier, T., Schmidt, A., Sick, B., Stumme, G., & Zweig, K. A. (2016). From active learning to dedicated collaborative interactive learning. *Proceedings of the 29th International Conference on Architecture of Computing Systems* (pp. 1-8).

Clark, R. C., Nguyen, F., Sweller, J., & Baddeley, M. (2006). Efficiency in learning: Evidence-based guidelines to manage cognitive load. *Performance Improvement*, *45*(9), 46–47.

Clarke, T., Ayres, P., & Sweller, J. (2005). The impact of sequencing and prior knowledge on learning mathematics through spreadsheet applications. *Educational Technology Research and Development*, *53*(3), 15–24.

Cooper, G., Tindall-Ford, S., Chandler, P., & Sweller, J. (2001). Learning by imagining. *Journal of Experimental Psychology. Applied*, *7*(1), 68–82.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, *24*(1), 87-114; discussion 114-85.

Cowan, N. (2012). Working memory capacity. *Psychology Press*.

Dabbish, L., Mark, G., & González, V. M. (2011). Why do I keep interrupting myself? In D. Tan, G. Fitzpatrick, C. Gutwin, B. Begole, & W. A. Kellogg (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3127–3130).

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

Florax, M., & Ploetzner, R. (2010). What contributes to the split-attention effect? The role of text segmentation, picture labelling, and spatial proximity. *Learning and Instruction*, *20*(3), 216–224.

Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction*, *16*(6), 511–525.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, *37*(2), 337–355.

Gregor, S., Kruse, L., & Seidel, S. (2020). Research perspectives: The anatomy of a design principle. *Journal of the Association for Information Systems*, *21*, 1622–1652.

Hasler, B. S., Kersten, B., & Sweller, J. (2007). Learner control, cognitive load and instructional animation. *Applied Cognitive Psychology*, *21*(6), 713–729.

Hemmer, P., Schemmer, M., Vössing, M., & Kühl, N. (2021). Human-AI Complementarity in Hybrid Intelligence Systems: A Structured Literature Review. *Proceedings of the Twenty-fifth Pacific Asia Conference on Information Systems*.

Hevner, A. R. (2007). A three cycle view of design science reserach. *Scandinavian Journal of Information Systems*, *19*(2), 87–92.

Jones, D., & Gregor, S. (2007). The anatomy of a design theory. *Journal of the Association for Information Systems*, *8*(5), 312–335.

Kalyuga, S. (2006). Rapid cognitive assessment of learners' knowledge structures. *Learning and Instruction*, *16*(1), 1–11.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*(4), 509–539.

Kalyuga, S. (2008). When less is more in cognitive diagnosis: A rapid online method for diagnosing learner task-specific expertise. *Journal of Educational Psychology*, *100*(3), 603–612.

Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology*, *96*(3), 558–568.

Kalyuga, S., & Sweller, J. (2005). Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. *Educational Technology Research and Development*, *53*(3), 83–93.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*(1), 23–31.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, *13*(4), 351–371.

Kirschner, F., Paas, F., & Kirschner, P. A. (2009). A cognitive load approach to collaborative learning: United brains for complex tasks. *Educational Psychology Review*, *21*(1), 31–42.

Kolbjørnsrud, V. (2024). Designing the Intelligent Organization: Six Principles for Human-AI Collaboration. *California Management Review*, *66*(2), 44–64.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, *41*(4), 212–218.

Lerch, F. J., & Harter, D. E. (2001). Cognitive support for real-time dynamic decision making. *Information systems research*, *12*(1), 63-82.

Mayer, R. E. (2012). Multimedia learning. *Cambridge University Press*.

Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology*, *91*(4), 638–643.

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, *19*(2), 177–213.

Oksa, A., Kalyuga, S., & Chandler, P. (2010). Expertise reversal effect in using explanatory notes for readers of Shakespearean text. *Instructional Science*, *38*(3), 217–236.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and Instructional Design: Recent Developments. *Educational Psychologist*, *38*(1), 1–4.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive Load Theory: Instructional Implications of the Interaction between Information Structures and Cognitive Architecture. Instructional Science, 32(1/2), 1–8.

Petter, S., Khazanchi, D., & Murphy, J. D. (2010). A design science based evaluation framework for patterns. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, *41*(3), 9–26.

Poser, M., Wiethof, C., Banerjee, D., Shankar Subramanian, V., Paucar, R., & Bittner, E. A. (2022a). Let's team up with AI! toward a hybrid intelligence system for online customer service. *Proceedings of the International Conference on Design Science Research in Information Systems and Technology* (pp. 142-153).

Poser, M., Wiethof, C., Bittner, E. (2022b). Integration of AI into customer service: a taxonomy to inform design decisions. *Proceedings of the European Conference on Information Systems (ECIS)*.

Ritz, E. (2024). Balancing bytes with brains: Exploring the role of learners' control in personalized learning tasks. *Proceedings of the European Conference on Information Systems*.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science*, *21*(1), 1–29.

Renkl, A. (1999). Learning mathematics from worked-out examples: Analyzing and fostering self-explanations. *European Journal of Psychology of Education*, *14*(4), 477–488.

Renkl, A., Atkinson, R. K., & Maier, U. H. (2000) From studying examples to solving problems: Fading worked-out solution steps helps learning. *Proceedings of the 22nd Annual Cognitive Science Society 2000*, pp. 393–398.

Renkl, A., Schwonke, R., Wittwer, J., Krieg, C., Aleven, V., & Salden, R. (2007). Worked-out examples in an intelligent tutoring system: Do they further improve learning? *Proceedings of the International Cognitive Load Theory Conference*.

Rzepka, C. and B. Berger (2018). User Interaction with AI-enabled Systems: A Systematic Review of IS Research". *Proceedings of the 39th International Conference on Information Systems*.

Schwan, S., & Riempp, R. (2004). The cognitive benefits of interactive videos: learning to tie nautical knots. *Learning and Instruction*, *14*(3), 293–305.

Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Aleven, V., & Salden, R. (2009). The worked-example effect: Not an artefact of lousy control conditions. *Computers in Human Behavior*, *25*(2), 258–266.

Shneiderman, B. (2020). Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human–Computer Interaction*, *36*(6), 495–504.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, 22, 123–138.

Sweller, J., Mawer, R. F., & Ward, M. R. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology: General*, *112*(4), 639–661.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Cognitive Load Theory. *Springer New York*.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, *10*(3), 251–296.

Trafton, J. G., & Reiser, R. J. The contribution of studying examples and solving problems to skill acquisition. *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp. 1017–1022).

van Gog, T., Paas, F., & van Merriënboer, J. J. (2008). Effects of studying sequences of process-oriented and product-oriented worked examples on troubleshooting transfer efficiency. *Learning and Instruction*, *18*(3), 211–222.

Van Merriënboer, J. J. G. (1990). Strategies for programming instruction in high school: Program completion vs. program generation. *Journal of Educational Computing Research*, *6*, 65–285.

Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. *Educational Psychology Review*, *17*(2), 147–177.

Wajcman, J., & Rose, E. (2011). Constant connectivity: Rethinking interruptions at work. *Organization Studies*, *32*(7), 941–961.

Yao, B., Li, X., Ji, Z., Xiao, K., & Xu, W. (2024). Task reallocation of human-robot collaborative production workshop based on a dynamic human fatigue model. *Computers & Industrial Engineering*, *189*, 109855.

Ye, P., Wang, X., Zheng, W., Wei, Q., & Wang, F. Y. (2022). Parallel cognition: Hybrid intelligence for human-machine interaction and management. *Frontiers of Information Technology & Electronic Engineering*, *23*(12), 1765-1779.

Wambsganss, T., Kueng, T., Soellner, M., & Leimeister, J. M. (2021). ArgueTutor: An adaptive dialog-based learning system for argumentation skills. *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1-13).

Wiethof, C., & Bittner, E. (2021). Hybrid Intelligence - Combining the Human in the Loop with the Computer in the Loop: A Systematic Literature Review. *Proceedings of the International Conference on Information Systems*.