Unlearning before creating new knowledge: A cognitive process.

Thomas Grisold  
Vienna University of Economics and Business  
Austria  
thomas.grisold@wu.ac.at

Alexander Kaiser  
Vienna University of Economics and Business  
Austria  
alexander.kaiser@wu.ac.at

Julee Hafner  
The Chicago School of Professional Psychology  
Chicago, Illinois, USA  
wehaf2talk@gmail.com

Abstract
Recent research expresses serious doubts on the concept of unlearning. It is argued that knowledge cannot be discarded or eliminated in order to make space for the creation of new knowledge. Taking into account the recent scepticism, we focus on the cognitive dimension of unlearning and propose an alternative conceptualization. Considering how far unlearning can go from a psychological/cognitive scientific perspective, we propose that unlearning is about reducing the influence of old knowledge on our cognitive capacity. This study: (a) investigates the unlearning process within the cognitive domain and on an individual level and (b) proposes unlearning process triggers that detract or facilitate the knowledge change process, which could subsequently contribute to unlearning on an organizational level.

1. Introduction
Undoubtedly, unlearning has been attracting increasing interest in the fields of organizational learning, innovation, change and crisis management, and other fields [1], [2]. The general idea is that organizations must discard knowledge in order to keep pace with environmental changes and remain innovative [3].

However, the idea has not been without controversy. Researchers doubt that there is such a thing as unlearning. For example, they claim that knowledge cannot be discarded or eliminated; the concept is built on wrong premises and its implications are misleading. They suggest to forget unlearning [4].

In this paper, we want to make a contribution to the debate circling around various concerns and explore how we could provide clarification. Our motivation is to see how unlearning could be used in order to create new knowledge. We will review recent critique and see how an alternative view on unlearning could look like focusing on processes, which deal with unlearning of old knowledge (and previous experiences) but at the same time, are not about discarding or eliminating them. We focus on unlearning and its role to support knowledge creation.

The remainder of this paper is organized as follows. First, we will provide an overview of the research on unlearning and summarize major points of critique. We will suggest an alternative view on unlearning, taking into account how far unlearning can possibly go. In the third part, we will present a method that entails a phase of unlearning as defined in section 2. We will describe the results of two experimental settings and highlight how this new understanding of unlearning holds in practice. In the discussion section, we will point to emerging opportunities and discuss how the research on unlearning could evolve in the future.

2. Theoretical Background
2.1 Past and current research on unlearning
Interest in unlearning has been increasing since the 1980s. Drawing on experiments from psychology, researchers suggested that hindering or obsolete knowledge should be discarded or eliminated to make space for the creation of new one [5]. Learning of a new knowledge base to successfully perform tasks without errors has become an important focus [6]. As knowledge changes, the ability to maintain competitive advantage becomes difficult for both organizations and employees.

Unlearning has been approached in a variety of theoretical frameworks. While there is agreement that knowledge and/or behaviour should require unlearning, there is continued disagreement about what it this process actually is. The confusion about the characteristics of unlearning involves anecdotal evidence and lacks empirical agreement about the specifics of the process. Although the term unlearning is present within many disciplines, disagreement rests on a lack of a consensus regarding a clear definition, process understanding and usage of this term.

In its original sense, as proposed by Hedberg [7], and Nystrom and Starbuck [5], unlearning refers to the intentional elimination of knowledge, which is obsolete and may detract from new knowledge acquisition [8]–[10]. This approach may be related with the idea that organizations possess memory and how this memory can be cleared [11]; an idea, which
remains undeveloped and is still under debate [12]. Furthermore, it has been suggested that organizational knowledge change processes are related to learning and unlearning within the organization [13]. Knowledge change and acquisition is speculated to involve a “replacement” of prior knowledge [14], [15]. The idea that an individual should “eliminate pre-existing knowledge or habits that would otherwise represent formidable barriers to new learning” [5, p. 36] has not been established. Also, it has been proposed that unlearning occurs when previously held views and attitudes are being recognized and rethought [16]. In a similar vein, some authors focus on organizational forgetting [17]–[19]. As opposed to unlearning, which is an intentional process, forgetting is a “loss of knowledge that is not necessarily planned or intended” [17, p. 311].

Some researchers focus on organizational unlearning involving a number of individuals who aim at getting rid of distributed knowledge [20], [21], while others concentrate on the individual context [22], [23] and partially try to see if they can draw conclusions or suggestions for the organizational level [24], [25].

2.2 Critique and Issues

In the following, we highlight three aspects, which show serious critique on the concept.

2.2.1 Issue 1: Unlearning as discarding or eliminating knowledge?

In a very recent article, Howells and Scholderer [4] emphasize that knowledge cannot be unlearnt as it has been originally proposed in [5], [7]. They argue that the concept rests on an erroneous interpretation of psychological experiments, and the term is only occasionally used for related processes such as extinction [26]. The authors reason all subsequent research was built on wrong premises. The term is not even part of the PsycInfo-database, hence, the concept does not provide the scientific ground to which it explicitly refers. They conclude that researchers should forget unlearning [4].

Can we select specific “pieces” of knowledge to delete them? In support of the recent critique, research in psychology, cognitive science and neuroscience suggests a connectionist perspective on knowledge and cognition [27], [28]. Thereby, knowledge is distributed across neural networks where views, beliefs and behaviours are coherent and closely entwined [29], [30]. This implies that most knowledge is implicit and interconnected and it cannot be simply removed. Discarding or eliminating knowledge seems only possible if our brain is seriously damaged by a tumour or an accident and parts of the neural networks are destroyed [31]. Furthermore, defining unlearning as a process of discarding or eliminating knowledge evokes the impression that once, the process is finished, knowledge would be gone. However, it (implicitly and/or partially) remains in the distributed network and may even be activated after it has been “silent” for a long time, as it can be seen in people suffering from traumata or former drug addicts [32].

2.2.2 Issue 2: What is the difference between learning and unlearning?

As a further point of critique, it is under debate what the difference between learning and unlearning is [33]. In its broadest sense, learning is seen to be an acquisition of new knowledge while unlearning is thought to be the reduction of old knowledge [34]–[36]. Individuals or organizations face conflicts between their knowledge and the environment and to catch up with external changes, they must get rid of their old knowledge [36]–[38]. How would this process be any different than learning? After all, learning involves periods of reflection where subjects use meta-cognitive perspectives to see if and to what extent their knowledge is suitable to perform a task [39]. For example, Argyris and Schoen [40] suggest that there are different levels on which learning can take place: as opposed to single-loop learning, where subjects slightly adjust and improve their behaviour, they can also engage in double-loop learning where they reflect on mismatching experiences between them and the environment and assumptions, premises or paradigms are being changed [41], [42]; this refers to a change (e.g. in theory) and it would resemble to what many researchers refer to as unlearning [4]. Using the term unlearning to describe phases of reflection seems redundant, as it would highlight what learning theories already acknowledge. Learning and unlearning would be two sides of the same coin [33].

2.2.3 Issue 3: What should be unlearnt?

Knowledge is a broad term including explicit and implicit knowledge [43] or declarative and non-declarative knowledge [44]. However, the term unlearning is being used for both knowledge types simultaneously. For example, unlearning is used with regards to changes in routines and beliefs [21], routines, habits and cognitive frameworks or understanding and behaviour [35], [45]. Both knowledge types are closely connected as our beliefs and assumptions navigate our behaviour and become implicit over time [46]. However, they are not the same. Imagine you are used to take specific route from the subway station to your office. Eventually, a colleague tells you that there is a faster route you
could use. You realize that it is much faster (i.e. you unlearn your previous assumptions) and you decide to use the new route. At the same time, you might be tempted to use the old route for some time as your behavioural routines have not been affected (e.g. taking the left instead of the right exit at the station, etc.). The question is if we should use unlearning for the two notions of knowledge interchangeably given there is a lot of confusion regarding the terms in both theory and practice [47]. Rather, it may be better to first see what we would need unlearning for in order to then clarify which level we should consider.

To sum up, unlearning has been attracting interest in the field (organizational) learning. The concept goes through serious criticism because (1) it seems questionable if knowledge can be discarded or eliminated, (2) there is no clear distinction between learning and unlearning, and (3) it often remains imprecise on what level one unlearns.

2.2. Research question and research methodology

Based on the analysis of the unlearning concept presented in the last section, we can articulate the following research question:

How can the term unlearning be redefined with respect to the creation of new knowledge, taking into account recent critique in literature and thereby, clarifying the ground for future research?

Research methodology

To answer this research question, we use a comprehensive literature review to build the groundwork for an explorative analysis and a theoretical foundation.

3. Towards a new definition: cognitive unlearning to create new knowledge

As discussed in section 2, some authors argue that the term unlearning may be useless and irrelevant. Existing definitions seem misleading and/or redundant as they use unlearning and learning interchangeably.

It could be helpful to reframe the concept and find a definition, which considers recent critique and at the same time, implies that we get rid of previous knowledge in order to improve the capacity to create new knowledge.

In line with other researcher is in the field of organizational learning and unlearning, we refer to knowledge as a capacity that makes (collective) action possible, i.e. knowledge as a capacity to act that can be manifested on a cognitive as well as on behavioural level [18], [48], [49]. Since we investigate the concept of unlearning in the context of knowledge creation, we focus on the cognitive level exclusively and thus, we refer to knowledge as the explicit and implicit assumptions, beliefs and hypotheses that allow us to interpret the world and form the basis for the creation of subsequent knowledge [50].

We are concerned with the question of how we can overcome old thinking patterns, overcome past experiences and get rid of past-driven thinking in order to create new knowledge [30], [51]. In line with Wittgenstein who noted that we need words to think in terms of their concepts [52], we will call for ‘saving’ the term unlearning and reinforce it to the current research on knowledge creation as this could guide the focus of current and future research to enable unlearning of past-driven thinking.

3.1. Past driven learning and thinking

As proposed by research in psychology, neuroscience and cognitive science, our cognitive performance is “driven by the past” [51]. Thereby, our thinking and behaving is always affected by what we have already learnt [51]. Research in cognitive science and neuroscience suggests that we interact with the world by applying cognitive schemata that have been successful in the past to predict incoming sensory signals [29]. Over time, our thinking becomes entrenched by a set of causal beliefs, which navigate thinking, perception and behaviour; they underlie most of our assumptions, opinions and premises. What we see, feel and hear is driven by top-down processes in the brain, which are dependent on past experiences; they navigate our cognitive processes and are being projected on future events [50], [51], [53].

Learning from the past is well developed and underlies all major learning methodologies, best practices and approaches to organizational learning. Influencing learning theories refer to learning as experiential learning which is “the process whereby knowledge is created through the transformation of experience. Knowledge results from the combination of grasping and transforming experience” [54, p. 41]. An overview of some past-driven learning theories can be found in [55].

3.2. Unlearning as a process to reduce the influence of old knowledge

We might not be able to discard or eliminate previous knowledge and we cannot step out of our knowledge structures and start from a blank slate [56]. It is argued that our selves are modelled by the knowledge which we have constructed; this means that our most inner ideas, assumptions and perceptions of the world will always guide what we think and do [57].
regularity and stability [58]. When we speak about cognitive unlearning and stress its role for knowledge creation, the question should be how we can free ourselves from our past. Therefore, unlearning old knowledge to create new forms of knowledge does not mean that we would have to eliminate or discard knowledge but to reduce the past-driven nature of our cognition to detach from knowledge stemming from previous experiences, assumptions and beliefs. In contrast to previous research on unlearning, we do not suggest that unlearning helps to increase the memory space for new knowledge. What appears crucial for the cognitive dimension of unlearning in order to create new knowledge is that subjects are free to reduce the influence of previous knowledge while they are in a process of creating new knowledge so that they can interpret and interact with the world that is less determined by their past experiences and their previous assumptions, beliefs and proven ways of thinking. Thus, in the context of knowledge creation, unlearning could be defined as a process where subjects can overcome their entrenched ways of thinking and reasoning to improve their capacity to create new knowledge.

Therefore, we define cognitive unlearning as a process where subjects reduce the influence of old knowledge for the sake of creating new knowledge and/or patterns of thinking.

This definition seems to resolve the major points of critique as present in section 2 of this paper. First, this definition does not imply that old knowledge is discarded or eliminated but its influence is reduced for a specific duration. Second, unlearning is a phase of reducing the influence in order to subsequently learn new knowledge; thus, the processes differ from each other. And third, we avoid confusion regarding the level of unlearning and refer to the cognitive domain.

Our proposed definition overlaps with previous approaches claiming that unlearning serves to make space for the acquisition of new knowledge [36]. However, knowledge cannot be eliminated in binary way of thinking (i.e. either the knowledge is there or it is gone) but unlearning can be seen as a reduction of existing knowledge while creating new knowledge.

4. Unlearning in practice: presenting a method and discussing empirical results

Unlearning as detaching from experiences from the past and existing knowledge could enhance the capacity of organizations and individuals to create new knowledge. How can we achieve this in practice? In the following, we present learning from an envisioned future as a method to illustrate how subjects can unlearn the boundaries of their current knowledge in order to create new forms of knowledge.

4.1 Unlearning to learn from an envisioned future

Learning from an envisioned future is a method which we have been applying to (organizational) learning processes [59]. Thereby, subjects are guided into an ideal future scenario and learn from what they experience there. By projecting themselves in situations where everything is just perfect and fulfills their dreams and most inner wishes, they formulate answers to questions such as “What has happened in this future scenario that makes it perfect?”, and “What has ended in this scenario that makes it perfect?”. They are asked to describe this ideal situation, what they see and how it feels to be there. This method is useful for a variety of organizational learning processes, such as vision development processes, where members of an organization should formulate where they want to develop.

Learning from an envisioned future consists of two phases.

1. In the first phase, participants are encouraged to overcome what they have experienced in the past. This is done with a mental time-travelling guiding them out of their current world-view. This a phase where subjects unlearn as they project themselves in a future point of time and thereby, reduce the influence of old knowledge.

2. In the second phase, they learn after they have arrived in this ideal future where everything is possible and just perfect. They interact with this future, experience how it looks/feels there and at the same time, they create new knowledge, which is less affected by previous experiences.

The two phases are depicted in Figure 1.

![Figure 1: Unlearning to learn from an envisioned future](image)
participants feel uneasy and even anxious when they are asked to unlearn their common word-view [60]. Therefore, it is essential to provide enabling spaces [61] allowing participants to feel safe and encouraged to test and follow paths for the creation of new knowledge. In a similar vein, Nonaka et al. suggest the concept of ba (for example [62]) arguing that a shared space is the foundation for the creation of (individual and/or collective) knowledge. Providing such enabling spaces and ba appears particularly important for the phase of unlearning as it seems that participants are reluctant and must overcome an inner burden until they can let their thoughts flow and themselves go.

Furthermore, a facilitator engages them in a narrative where they are encouraged to detach from their implicit knowledge and are gradually guided into an ideal future scenario [63].

In the following, we will present the results of two experiments that were done within two organizational learning processes. In comparing the outcomes of two conditions (i.e. participants who created knowledge after a phase of unlearning versus participants did not unlearn), we show that the capacity to create new knowledge increases after reducing the influence of old knowledge.

4.2 Case Study High-School in Austria

The first experiment was part of a larger project to assess the needs of pupils from a high school in Lower Austria [64]. Within this project we organized two workshops with two classes to suggest how their ideal school would look like in a future point of time. All pupils were about the same age (17-18 years). In total, a number of 31 pupils and teachers participated in the study; 12 pupils and 2 teachers were learning from an envisioned future after they went through an unlearning phase (workshop 1), while 17 pupils were exposed to a learning where they reflected on past and current experiences to decide what should be changed today to have an ideal future (workshop 2).

In workshop 1, the class that learnt from an envisioned future was exposed to a setting to facilitate the unlearning of their present situation, previous experiences and current expectations. A facilitator guided them into a scenario taking place in the year 2020; the narrative time journey implied that they were leaving the year 2014 and all doubts and concerns would become obsolete. This unlearning phase took up to several minutes and the imagined time leap was illustrated with Richard Strauss’ Zarathustra; this piece of music has been reportedly useful for subjects to feel excited and enabled to imagine that a time travelling takes place. After this phase of unlearning, the subjects were welcomed in the year 2020 where they experienced their ideal school to learn from their imagination. They were asked to write down what has emerged in this ideal school, and what has come to an end.

In workshop 2, the pupils were exposed to a learning setting where they were asked to reflect on their previous experiences in their school and to subsequently think of what they would like to change today in order to have an ideal school in 2020. They did not undergo a phase of unlearning to subsequently learn from an envisioned future; they did not perform a mental time travelling to unlearn. Similarly, they wrote down what will have emerged, and what will have come to an end.

Analysis and results

In workshop 1 participants generated a total number of 369 satisfiers, whereas in workshop 2 the respective participants generated a total number of 520 satisfiers.

In order to see the differences between the two conditions, we used the Paradigm Relatedness Framework to evaluate the novelty of an idea with regards to the status-quo of a particular system [65], [66]. Thereby, we can see the extent to which subjects could unlearn the influence of their previous knowledge as we can assess to what degree an idea is in line with the current system. An idea is (1) paradigm-preserving if it refines the current situation but the situation itself remains the same; it is a minor incremental improvement. In contrast, paradigm-modifying ideas change a current situation by (2) adding a new element to the context, (3) redesigning the situation changing the relationship between the elements, or (4) by transforming the system by both adding new elements and changing the relationship. Category 1-ideas are the least innovative ideas while category 4-ideas are fundamental breaches and radically innovative. This is depicted in figure 2.

![Figure 2: Paradigm-relatedness framework](image)

We randomized the collected data, removed any bias to see whether they were produced in workshop 1 or 2. In the following, we present examples from the data set to show how ideas for one paradigm (teaching) look like for each category.
- **Category 1**: Better explanations by teachers (refining current situation)
- **Category 2**: New teaching methods (adding new element to system)
- **Category 3**: Curriculum is organized as a flexible module system (changing the relationship between existing elements)
- **Category 4**: No attendance at all, pupils can attend school via Skype (changing relationship of the elements and adding a new element)

Overall, workshop 1 (WS1) produced more paradigm modifying ideas compared to workshop 2 (WS2). WS1 generated 90% of all satisfiers that are ascribed to category 4 (i.e. containing the suggestions that are most radical new for the system). Also, the production of category 3-satisfiers was facilitated as it is evident with about 65%. Accordingly, unlearning to learn from an envisioned future produced output that is more challenging to the status-quo of a social system and yields a higher degree of novelty. At the same time, WS2 produced a considerably higher number of satisfiers that are paradigm preserving, i.e. that refine the current state of the system, with a percentage of 65%. Therefore, there is an overall tendency for providing more moderate and less novel ideas in a learning taking into account past experiences. Figure 3 shows the distribution of the two learning modes for each category.

![Figure 3: Distribution of ideas](image)

**Figure 3: Distribution of ideas**

Figure 4 shows a comparison of both conditions with respect to the categories 1 to 4.

![Figure 4: Distribution of ideas for each condition](image)

**Figure 4: Distribution of ideas for each condition**

The results indicate that subjects produce output which seems less influenced by current knowledge and have a higher degree of novelty after they went through an unlearning phase and subsequently learnt from their future.

### 4.3 Case Study Austrian Economic Chamber

The Austrian Chamber of Economics represents tradesmen and tradeswomen in Austria covering annual revenue of approximately 80 billion euros. In 2015, we conducted a project to develop a strategy for the industry section of crafts and trades, which is one of the seven sectors represented by the Austrian Chamber of Economics. Within this project, we held two workshops where we presented a number of representatives with pre-specified strategic goals and invited them to develop a set of concrete actions, which could potentially reach these goals. Similar to case study 1, the workshop differed in terms of the utilized learning approaches; in workshop 1, we asked participants to reflect on their past experiences and in workshop 2, we asked them to learn from an envisioned future after going through a phase of unlearning.

A number of 35 representatives participated in workshop 1. The main objective was to develop concrete ideas and actions in order to achieve pre-specified goals. Participants were exposed to a setting to facilitate the unlearning of their present situation, previous experiences and current expectations. A facilitator guided them into a scenario taking place in the year 2020. This unlearning phase again took up to several minutes and the imagined time leap was illustrated with Richard Strauss’ Zarathustra. After this phase of unlearning, they were welcomed in the year 2020 where they experienced and learnt how their imagined ideas achieved the pre-specified goals.

Workshop 2 had a total number of 18 representatives. The goal was to develop concrete actions to realize the five predefined goals, by inviting participants reflect on the current situation as well as past experiences. A facilitator asked them to reflect on what has worked and what has not worked in the past, they were encouraged to formulate ideas, which they thought had the potential to reach these goals in the future.

**Analysis and results**

In workshop 1, participants came up with 62 actions and workshop 2 resulted in 41 actions. We randomized the collected data of both workshops. Subsequently, we used the paradigm relatedness framework.

The analysis reveals that a significant majority of the actions that were found to belong to category 4, i.e. the most paradigm-challenging, originate from
the unlearning and learning from an envisioned future-condition, as it is evident with 86%. On the contrary, actions that were suggested in WS2 make up almost two thirds of category 1, i.e. the least radical and possibly least innovative category. It has to be stated that WS1 produced more actions than workshop 1, which may be due to the higher number of participants in WS2. However, the distributions in category 2 and category 3 confirm the trend that unlearning to learn from an envisioned future leads to more status-quo challenging actions.

![Figure 5: Distribution of the four categories](image)

Figure 6 depicts a comparison of the approach with unlearning and the approach without unlearning with respect to the categories 1 to 4.

![Figure 6: Distribution of ideas for each condition](image)

Results indicate that going through an unlearning phase enables people to create new knowledge which seems to be less influenced by current knowledge.

5. Discussion and conclusion

5.1. Implications for theory and practice

We acknowledge that our definition of unlearning departs from current definitions. However, current allegations (as brought forward by Howells and Scholderer and colleagues) are serious and we argue that the unlearning concept is in need of an interdisciplinary approach to clarify how/if it can be realized. We argue that our definition of unlearning as reducing the influence of old knowledge for the sake of creating new one is in scope of what is possible from a psychological/cognitive scientific point of view. This is why this definition could contribute the research on knowledge creation.

At the same time, our definition provides a new perspective on other research strands in the field of (organizational) learning where such an unlearning is an inherent but implicit part of. For example, Scharmer suggests to learn from the future as it emerges [67]. He suggests that we should overcome our current ways of thinking to see potentials which the environment yields but which we cannot recognize. A closer look at this idea reveals that it implicitly entails unlearning as he suggests to suspend our current thinking patterns and dwell in a state where we attune to unknown features in the environment [41].

Similarly, the concept of mindfulness also entails aspects of unlearning. By taking a non-judging stance towards our environment, we prevent previous experiences to be projected on current situations. It provides a deep, non-conceptual seeing into the world [68]. This has measurable effects for cognitive performance; for example, it has been found that mindfulness improves creative thinking [69] and that it reduces intentional blindness and improves the perception of unexpected stimuli in goal-directed tasks [70].

We reason that our concept of unlearning could guide researchers to focus on the period where we overrule our current knowledge and previous expectations in order to create new knowledge.

Our results reflect that we can reduce the influence of knowledge on an individual level. How does unlearning on an individual level interrelate with organizational unlearning [21], [24]? In the case of our proposed definition of unlearning, we argue that the integration of the individual and the organizational level depends on the purpose of the overall unlearning/learning process; for example, when the goal is to develop a new vision for an organisation, the outcomes of the individual unlearning/learning process are being merged on the collective level and thus, the overall output (i.e. the vision) will have contributed from the reduction of old knowledge. Unlearning as proposed in this paper should facilitate the innovation process in general as it enables organizations to transcend the boundaries of their current thinking. We propose that this could be particularly useful for the design of new services, products and processes in various fields, when the aim is to create something literally new.

5.3 Limitations and future research

This research provided a better understanding of the complex process of unlearning and how it may
References


[47] P. M. de Holan and N. Phillips,


