Empirical Test of a Formal Framework of Forgetting

Sara Todorovikj (sara.todorovikj@hsw.tu-chemnitz.de)
Daniel Brand (daniel.brand@hsw.tu-chemnitz.de)
Marco Ragni (marco.ragni@hsw.tu-chemnitz.de)
Predictive Analytics, Chemnitz University of Technology
Straße der Nationen 62, 09111 Chemnitz, Germany

Abstract

Forgetting is a fundamental cognitive process that allows to efficiently discard outdated and irrelevant knowledge and manage information for humans and artificial systems alike. Originating from the latter, Beierle et al. (2019) propose a formal framework for modeling forgetting operators from a commonsense point of view. In this article, we aim for bridging the gap by investigating the formal forgetting operators with respect to their effects on human cognition utilizing an abstract experimental paradigm – the Counting Game. Using participants' accuracy and response time in rule-based tasks, we examine if and how the forgetting operators affect human performance in response to rule manipulations. We found that the addition, contraction and revision of a rule have the most remarkable effect on performance. Therefore, we modeled them using features describing the rule system and current scenario to predict the change in accuracy they prompt. By analyzing the effect that the operators have on humans when used to manipulate rule systems, we make a step towards further modeling the formal aspect of changes to rule systems in applications in order to understand the potential effect they have on the users.

Keywords: Forgetting Operations; Empirical Validation; Rule Updating

Introduction

Humans are constantly exposed to new situations that require them to use various cognitive mechanisms to categorize and interpret information. Understanding how we acquire and update such categorization rules is central in cognitive science. Rule changes, in general, play a crucial role in problem solving, decision making and learning. However, sometimes knowledge is irrelevant or outdated and requires to be discarded. As an everyday phenomenon, forgetting is a fundamental cognitive process that allows humans to efficiently manage and restructure information. Beyond psychology, forgetting has been studied in different contexts, like logical systems, e.g., propositional or first order logic (Delgrande, 2014), or belief revision (Alchourrón, Gärdenfors, & Makinson, 1985). However, a generally accepted theory or formalization that unifies all such approaches does not yet exist. Towards that goal, Beierle, Kern-Isberner, Sauerwald, Bock, and Ragni (2019) propose a framework that conceptualizes different forms of forgetting, offering a structured approach to modeling belief change and formalizing forgetting from a common-sense point of view.

Experimental approaches in the field of rule changes often use rule-based systems to analyze how individuals update concepts. For example, Brand, Dames, Puricelli, and Ragni (2022) investigated the cognitive costs of adding new categorization rules or altering previously learned ones, finding that modifying an existing rule was less cognitively demanding than adding a new one, indicating the presence of processes depending on the number of rules. Additionally, they found that performance improves when new rules align with old ones, but in case of conflicting situations, this was no longer apparent. At their core, rules function as if-then conditions: if *a precondition is satisfied* then *a corresponding action follows*. This structure underlies processes such as learning to categorize (Maddox, Ashby, Ing, & Pickering, 2004; Maddox, Ashby, & Bohil, 2003), implementing intentions (Oettingen & Gollwitzer, 2010; Gollwitzer & Brandstätter, 1997) and is central to cognitive architectures like ACT-R (Anderson, 2007), among others. Studying how rules are acquired, updated and modified is essential for understanding how individuals adapt to changing environments.

To bridge the gap between formal models and human cognition, we utilize a new experimental paradigm designed to empirically test the predictions made by the formal framework. Using controlled rule-based tasks, we examine how and to which extent participants adapt and apply their knowledge in response to rule changes. This paradigm allows us to investigate different forgetting operators through rule manipulations and assess associated cognitive costs. Tying formally defined forgetting operators that describe changes in knowledge and rule systems can also allow their application in human-centered fields. If the effects (e.g., their impact on cognitive load) that the operators have when used by humans can be determined and understood, those operators can be used as a means to model not only the formal aspect of changes to rule systems in applications (e.g., software systems and work processes), but also provide estimates on the effects such changes will have on the users.

In this article, we aim to provide an empirical validation of formal definitions of forgetting operators. More specifically, we examine whether human behavior aligns with forgetting patterns predicted by the formal framework, by looking into how these operators influence accuracy and response time. Building up on our analysis results, we present linear regression models for three rule operators that aim to represent the effect they have on accuracy by considering features that describe the participants' knowledge base and the impact a rule has on the specific scenario.

Theoretical Background

Within the framework for modeling forgetting operators presented by Beierle et al. (2019), given a belief state that can have any type of representation, such as logical formulas or rankings over possible worlds, an agent's inference relation determines what follows from the beliefs. Using ordinal conditional functions (OCFs) they can assign degrees of plausibility, allowing for more nuanced reasoning and belief modification and removal, i.e. forgetting. We focus on the following five forgetting operators.

Contraction is an operator describing explicit intention to remove information, e.g. a navigation system being informed about a permanent closure of a street. Formally, contraction of a belief results in not believing the proposition afterwards, leading to a consistent belief state.

Revision involves presentation of new information that contradicts already established knowledge, prompting the forgetting of the old information. E.g., receiving new information that a person who was previously a student is now a staff member. In the case of revision, forgetting is implicit - given new information, any beliefs conflicting with it have to be forgotten in order to adopt the new, prioritized belief. It can be made explicit formally, by first contracting the old information and then adding the new one.

Abstraction is a common change operation that involves generalizing specific beliefs or rules in order to create a more broadly applicable one. For example, a rule can be that an object with two wheels and a bike bell is a bicycle, whereas another rule can state that an object with two wheels and no bike bell is also a bicycle. From these rules, a simplified rule can be abstracted, namely that an object with two wheels is a bicycle. Both, deductive and inductive reasoning may be involved in deriving a more general principle, which in turn allows to potentially forget the more complicated initial rules.

Conditionalization is an operator that restricts a belief set by omitting ones that do not adhere to a specific case, context or precondition. Inspired by probabilistic conditionalization where posterior beliefs are determined by conditional probabilities, the operator centers the beliefs around the condition leading to forgetting when it is not satisfied.

Fading out is a gradual forgetting process where information becomes increasingly difficult to retrieve over time, especially in the case of infrequent use. For example, a PIN code for a credit card that is rarely used is more likely to be forgotten eventually. Formally, it describes the increase in required effort to make an inference based on a set of beliefs in a later point in time.

Commonly, such belief systems are illustrated with declarative instead of procedural knowledge (e.g., like in the PIN code example above) but are in fact applicable on the latter as well. This allows to utilize the presented operators in the context of rule changes and rule learning, providing a formal framework for describing changes to a rule system. Thereby, the question arises if and how those formal forgetting operators are performed in human memory, and what the cognitive costs of each operation are. To this end, first investigations of *revision* have been performed, indicating that updating and altering an existing rule is less cognitively demanding than learning a new rule (Brand et al., 2022). However, in order to assess the applicability of the forgetting operators also as a cognitive modeling framework, all operators need to be investigated under controlled and comparable conditions. To this achieve this, we utilize a new abstract experimental paradigm based on shapes and colors, called the *Counting Game* (anonymized). Its goal is to provide a mean to empirically investigate and compare cognitive costs of forgetting operators, among many other concept changes.

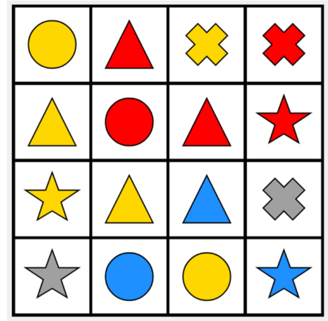

Figure 1: Representation of the Counting Game showing a trial whose winner is yellow. A rule "Red triangles count twice" changes the winner to red.

Method

The Counting Game consists of *trials*, which are 4×4 grids of 16 objects with various *shape* (circle, triangle, cross, star) and *color* (red, yellow, blue, gray), as shown in Fig. 1. Each trial has a winner – the color that has the largest amount of objects, among the "competing" colors – red, yellow and blue. A valid trial has only one winning color (which cannot be gray) and at most half of the total number of elements can be of the same color. A study implemented within the Counting Game paradigm presents multiple trials to participants and prompts them to determine which color is the winner as fast as possible. To motivate fast, but also accurate answers, we introduce a scoring system that awards 10 points for each correct answer, with an additional bonus timer counting down from 20 seconds and awarding 1 point per second remaining. So, a correct answer with 8 seconds left on the timer awards $10 + 8$ points). Incorrect answers do not award points, regardless of how fast they were given.

At the beginning, participants simply count objects – each one, regardless of color and shape, counts as one. Throughout the game, we introduce two types of rules that change how specific objects are counted:

1. *Count-twice* rules state that objects of a certain color and shape count twice from now on (e.g., "Red stars count twice.")

2. *Count-as* rules state that objects of a certain color and shape count as a different color from now on (e.g., "Blue

Table 1: Rules implementing the tests of the forgetting operators.

Operator	Rules
Contraction	Blue stars count twice. Blue stars count as one again.
Revision	Red circles count as yellow. Red circles count as blue.
Abstraction	Red circles count as yellow. Red stars count as yellow. Red crosses count as yellow. Red triangles count as yellow.
Conditionalization	Gray crosses count as blue.
Fading Out	Blue crosses count twice.

crosses count as yellow.")

Using these rules, we test the previously introduced forgetting operators, as shown in Table 1. We manipulate the individual's rule system within the game by changing how specific objects count, which then in turn allows to quantify the impact of the forgetting operations on performance.

In this work, a *Phase* refers to a stage within a round when a specific rule configuration is active. All rounds start with no rules in *Phase 0* and progress through multiple phases. A new phase begins each time a rule is introduced, updated or removed. So, *Phase 1* begins after the first rule, *Phase 2* after the second, and so on. Each phase consists of the same trials in randomized order. That allows us to investigate how participants adapt to new rules and how/if prior rules affect the application of recent ones. All rules have corresponding *critical trials*. They are trials whose winner changes after a rule is applied. For example, the trial depicted in Fig. 1 is critical for the rule "Red triangles count twice" as its application changes the winner from yellow to red.

Forgetting Operators

An overview of rules used to implement the forgetting operators is provided in Table 1. Contraction is implemented by providing one *count-twice* rule and explicitly withdrawing it later. In the case of revision, we give two *count-as* rules, such that the second one overrides the first one, implicitly making it irrelevant. For abstraction, the participants got four *count-as* rules that cover all shapes of a given color, prompting them to ultimately replace the four rules with one, e.g. "Red shapes count as yellow". The presence of gray elements in a trial is optional, therefore for conditionalization we use specific *count-as* rules where the initial color is gray, introducing the precondition that *if* there are gray shapes, they will count as a different color. In the case of fading out, we only use one *count-twice* rule, however, here the temporal component of the operator is important. We had multiple fading out rounds manipulating the amount of time (trials) between the rule presentation and the corresponding critical trials. We added filler rules, so that the rounds are comparable with four rules each.

Participants

Each participant was randomly and evenly assigned to one of the five operator conditions. To prevent unwanted effects or bias, each participant had a randomly assigned set of colors and shapes. So, a rule "$color_1$ $shape_1$ count twice" could be about red crosses for one person and yellow stars for another. The general structure of the rounds, rules and trials is otherwise preserved.

We obtained data from 117 participants recruited on Prolific[1]. The experiment was performed online as a web-experiment. Participants needed on average 25 min. to complete the experiment and received monetary compensation at the end. All participants were native English speakers. The participants first had to provide consent to have their experimental data stored. Afterwards, they were introduced to the game and its rules, the possible shapes and colors, the goal to find the winner as fast as possible and the scoring system. Then, they did a practice round with a few tasks in order to get accustomed to the format. Once finished with the instructions, the participants started with the study. The data, analysis and modeling scripts developed for this article are available on GitHub[2].

Analysis

We identified as outliers four participants whose mean accuracy score diverged by at least 2 standard deviations away from the mean and excluded them from the analysis. The final number of participants is 113. The overall accuracy across all participants and conditions was 83.62% (Median $= 100\%, SD = 37.01\%$). The mean accuracy of trials critical for at least one rule was 74.62% and 95.68% for non-critical trials. The median response time (RT) across all trials was 2.06s. Critical trials had a higher median RT (2.22s) in contrast to non-critical trials (1.90s).

Contraction We examine the average response accuracy of critical and non-critical trials when no rules were present, after a rule addition and after that rule's contraction, as presented in Table 2. The accuracy decreases once a rule is added (Wilcoxon: $z = 68.50$, $p = .010$) and we identified that all errors are due to participants not applying the rule. Once the rule is contracted, the accuracy improves once again ($z = 4.00$, $p = .014$). Analyzing the errors, we discovered out that 93.75% of wrong responses were because the rule was still applied, i.e. it was not contracted. Furthermore, we confirmed that correct responses after contraction were not a consequence of not applying the rule in the first place (Spearman $\rho = 0.13$, $p = .227$). In general, contraction should fa-

[1] https://www.prolific.com

[2] https://github.com/saratdr/ICCM2025-CountingGame

Table 2: Average response accuracy on critical and non-critical trials in the contraction and revision round. Crit and NonCrit indicate critical and non-critical, respectively.

		No rule	Rule	After Op.
Contraction	Crit	88.89%	63.33%	82.22%
	NonCrit	91.11%	94.44%	91.67%
Revision	Crit	85.71%	75.19%	80.45%
	NonCrit	94.30%	92.98%	94.74%

cilitate rule application, reducing the cognitive load by explicitly removing a rule. To investigate this effect, we compared two scenarios where three rules were added. We focus on response accuracy of critical trials for the last rule if the second rule was previously contracted and not (Mean: $contr = 76.79\%$, $notcontr = 57.81\%$). Since these scenarios used different trials and rules, we normalized the values by the participants' general performance on them in a phase with no rules. We observed a marginal effect suggesting that contraction may indeed ease the process of remembering and applying rules (Mean, normalized: $contr = 87.61\%$, $notcontr = 72.87\%$; Mann-Whitney: $U = 149.50$, $p = .062$).

Revision For revision, the accuracies are depicted in Table 2. Like in the case of contraction, non-critical trials remained unaffected and stayed at a consistently high accuracy of over 90%. For critical trials, a similar pattern to the contraction emerged, with a dip in performance when the rule was added, with a slight improvement after revision. In this case, however, participants performed better after the addition of the rule, which was likely due to an easier structure of the critical trials compared to those of the contraction condition. However, since revision does only change the rule to a new rule with an arguably comparable difficulty, it seems unexpected that performance would increase compared to the simple addition of a rule. A possible explanation could be that participants, who forgot or did not learn the first rule properly, got essentially a "second chance", since the revision removed the rule in question. 75.76% of the errors after a rule was added (with a total of 24.81% of the critical trials being answered incorrectly) were in fact due to the rule not being applied, indicating that the rule was not perfectly learned. This at least partially supports the possible explanation, that participants indeed simply did not learn the first rule properly, but were better prepared later. After the rule was revised, 7.69% of the remaining errors (19.55%) was due to the previously learned rule still being present. This indicates that the revision itself was performed successfully, leaving little artifacts of the old rule.

Abstraction In the case of abstraction, we analyzed how the response accuracy on critical trial changes after the addition of each rule. Thereby, the mean accuracy decreased after each rule until the final rule necessary for performing abstraction was introduced (70.83%, 56.25%, 43.06% to 47.22%, for the respective phases). We can immediately notice a tendency of accuracy to decline with each rule addition, except for a marginal increase after the last rule. Given that the addition of the last rule enabled participants to abstract over all four rules and apply only one, a better performance is indeed expected in the last phase. Such increase was found for the majority of the participants (58.33%).

Table 3: Response accuracy ratio between Phase 1 (rule applied) and Phase 0 (no rules) on irrelevant trials that do not contain the object affected by the rule for conditionalization and the rest, and trials with no gray elements for conditionalization.

Rule	Trials	Accuracy Ratio
Conditionalization	Irrelevant	99.78
	No Gray	104.58
Rest	Irrelevant	101.58

Conditionalization For conditionalization, an important distinction is between the effect of the knowledge level (i.e., are rules not present/active at the moment, so that they don't affect cognitive load) and the effect of rule application itself (i.e., the counting itself is easier when no additional rules need to be considered). Thereby, the important part for our evaluation is the former, since it is the one that the logical framework is concerned with. However, in practice, both effects occur together, making it hard to distinguish between them. To this end, we first investigate the effects on *irrelevant* trials, i.e., trials that do not contain elements affected by the rule. That way, we can see if the rule is generally less present when it is not applicable compared to other rules. Additionally, since gray is a color only relevant due to the conditionalization rule, we also consider trials with no gray elements at all. Table 3 shows the ratio between the accuracy of *Phase 1* (where the first rule is present) and *Phase 0* (where no rule is present) for the conditionalization on irrelevant trials and trials without gray shapes, as well as for the other conditions as a comparison. It becomes apparent that the accuracy does not substantially change for either condition, and is also only slightly improved when gray shapes are not present at all, indicating that conditionalization does seem to affect the cognitive load mostly on the applicational level.

To estimate the effect on the applicational level, we compare the response times between relevant and irrelevant trials, as well as trials with and without any gray elements. Thereby, no significant effect was shown between relevant and irrelevant trials (Median: $rel = 2.65s$, $irrel = 2.66s$, Wilcoxon: $z = 368.5$, $p = .061$). For trials with and without gray elements, however, the effect on the response time was significant Median: $rel = 2.76s$, $irrel = 2.51s$, Wilcoxon: $z = 389.5$, $p = .026$). Put together, this shows that the conditionalization of the prominent surface feature (i.e., gray color)

Table 4: Average response accuracy on critical trials immediately after rule addition and at the end of the round. Rounds 1, 2 and 3 have an increasing number of trials. Round 4 introduces additional rules, irrelevant to the fading out rule.

Round	After rule	End of round	Mean Difference
1	69.44%	51.39%	20.32%
2	66.67%	48.61%	20.53%
3	76.39%	45.83%	32.50%
4	65.28%	20.84%	44.65%

did help to accelerate the process of solving the tasks, however, as soon as the more detailed condition (i.e., color and shape) comes into play, that advantage vanished.

Fading Out In the case of fading out, we tested participants' performance in four different rounds. In the first three rounds, participants were presented with only one rule in *Phase 1* followed only by trials and no additional rules. The length of the rounds based on the amount of presented trials corresponds to 3, 4 and 5 phases, respectively. The critical trials were presented once among the first 12 trials after the rule and once among the last 12 trials of the round. That way, we ensured an infrequent application of the rule and we can examine the temporal effect of the operator. The average response accuracy and their difference between the first and last trial presentation are presented in Table 4. While a decrease in accuracy over time can be noted, it is not a drastic one, with a 32.50% difference in the longest round.

In the fourth round, we used the same principle of only presenting the critical trials of the relevant rule at the beginning and end, but we also added rules at the beginning of each phase. This led to a much substantial negative effect on performance, suggesting that information is rather increasingly difficult to retrieve over manipulations than time. In order to further confirm the effect of additional rules within this round, we looked into their corresponding critical trials. In Table 5 we can see that a rule addition leads to a lower accuracy for its critical trials. Moreover, introducing yet another operation causes a further substantial decline for trials critical for previous rules, indicating that a recency effect potentially comes into play as well.

Modeling

In our analyses, we found strong evidence for the presence and effect of three operators involved in rule manipulation: *addition* (implicitly used to build up to rule system), *contraction* and *revision*. In order to further evaluate their impact and estimate the associated cognitive costs we use linear regression models to predict the difference in accuracy that an applied operator induces on a trial per trial basis. To account for the influence of the cognitive demand of the participants' current rule system, a feature representing the amount of rules that are currently active in a given trial was included (*CurrActive*). In order to contrast the abstract nature of the operators with their dependency on the context in which they are applied, we add *RelevantBool* and *RelevantNum* as model features. They incorporate the current scenario by indicating if and how many relevant shapes for the currently added, contracted or revised rule are present in the trial, regardless if it's a critical trial one or not.

Table 5: Additional rules in Round 4 of fading out. The second rule (R2) is introduced in Phase 2, the third (R3) in Phase 3. R2 is contracted in Phase 4. R2 Crit and R3 Crit are critical trials for the second and third rule, respectively.

		Phase 2	Phase 3	Phase 4
Accuracy	R2 Crit	58.34%	25.00%	80.56%
	R3 Crit	80.56%	72.22%	36.11%

In Table 6 we present the means and standard deviations of the accuracy differences that the models aim to fit, along with the models' mean absolute errors (MAE) and parameter values. For the critical trials the *RelevantBool* parameter is irrelevant, since all of them contain relevant shapes by default. Additionally, the parameter *CurrActive* is irrelevant for revision, since in our experiment, it always took place in the same phase to facilitate the statistical analyses. We first examine our target variables and model fits. Addition and revision of a rule harm performance in critical trials, whereas contraction improves it. This is in line with the intuition behind the operators. In contrast, the operators seem to have minimal effect on average for non-critical trials. In terms of the impact on performance on critical tries, revision is substantially less impactful compared to addition, which is in line with the results of Brand et al. (2022). Thereby, revision is showing a performance similar to a combination of contraction and addition.

Considering the model fits, it becomes apparent that revision achieved a substantially better fit, while both addition and contraction have R^2 values close to 0, thereby not being able to utilize the given features. This, however, is likely due to the fact that revision is only present in isolated cases (i.e., conditions that explicitly test revision), while especially addition is present across all conditions, making the data more noisy and less dependent on the limited features. Nevertheless, the coefficients provide some insight about the used features. Compared to the number of relevant shapes, the number of active rules has a substantially lower importance. This highlights the strong relationship between the context (i.e., the actual composition of shapes in a trial) and the rules, while the effects of the manipulations to the rule system itself are rather indirect.

Discussion

In this article, we investigated a formal framework of forgetting operators introduced by Beierle et al. (2019) on a cognitive level, in particular for describing changes of a rule system. Therefore, we utilized a new experimental paradigm to

Table 6: Mean and standard deviation (STD) of differences in accuracy after adding, contracting or revising a rule for critical and non-critical trials. Mean absolute error (MAE), coefficient of determination (R^2) and parameter values of our models fit. *CurrActive* is the amount of currently applicable rules, *RelevantBool* indicates the presence of shapes affected by the rule in the trial and *RelevantNum* is the amount of those relevant shapes. The parameter *RelevantBool* is irrelevant for critical trials, since all of them contain relevant shapes. *CurrActive* is irrelevant for revision, since it always took place during the same phase.

Trials	Operator	Mean	STD	MAE	R^2	Model Parameters			
						CurrActive	*RelevantBool*	*RelevantNum*	Intercept
Critical	Addition	−19.04	17.49	13.00	0.11	−0.003	–	0.313	−0.356
	Contraction	13.87	20.64	15.04	0.09	0.089	–	0.114	−0.039
	Revision	−10.00	12.99	6.14	0.62	–	–	0.477	−0.383
Non-critical	Addition	−2.40	16.71	10.85	0.07	−0.019	−0.002	0.302	−0.033
	Contraction	−4.42	17.20	11.33	0.10	−0.024	0.007	−0.336	0.038
	Revision	5.00	9.52	4.04	0.46	–	−0.185	0.496	0.046

conduct a study in which participants had to apply various rules in a simple counting task: Participants had to decide which color was the most prominent in a set of shapes. Within this paradigm, distinct manipulations of the counting rules (e.g., some shapes are ought to be counted twice) allowed to assess the impact the forgetting operators have on performance. We conducted a study within the paradigm in which we used rule manipulations in order to assess the effects of forgetting operators and evaluate the predictions made by the formal framework. This allowed us to gain insights about how and if those formal operators are affecting human rule manipulation in procedural memory.

Thereby, we investigated five operations: contraction, revision, abstraction, conditionalization, and the fading out of a rule if it is not utilized over time. Implicitly, we also tested the addition of a new rule, as the only not forgetting-related operator. We found that adding a rule significantly decreased the accuracy of solving the task, and even to a bigger extend as the revision of a rule – which is in line with the findings of Brand et al. (2022). For contraction, i.e., the removal of a rule, we found that performance was restored to almost it's original level. Furthermore, errors after contraction mostly related to the rule still being applied, which indicates that artifacts of the rule still remain. Additionally, we found typical effects of directed forgetting (e.g., Dames & Oberauer, 2022; Dames, Brand, & Ragni, 2022): being able to forget a rule, allowed participants to free resources for learning other rules. This also can explain the lower impact on performance of a revision compared to a simple addition, since revision can be interpreted as the combination of forgetting of an old rule and learning of a new one.

Furthermore, we found evidence that abstraction, i.e., the simplification of several specific rules into a more general one, seemed to take place. While the addition of rules typically decreases performance, the majority of participants showed improvements after the last missing rule was added to allow for abstraction. For Conditionalization, we found that the effects are mostly related to the application of rules itself, and do not influence the rule system itself. While participants improved when a rule was not relevant for a trial, the type of the rule itself did not affect performance on the trials where it was not applicable, thereby giving no indication for a difference in general cognitive load.

Finally, for the last operator, fading out, we found a subtle trend that longer phases where a rule is not utilized indeed reduce performance when it is relevant again. However, the effect is only within the magnitude of error, thereby not warranting a conclusion. Overall, this can also be an effect of the generally short time-spans within the experiment, which may have been too short for time-based forgetting to occur. We additionally found a more substantial decrease in performance when additional rules are presented, implying that manipulations have a larger negative impact on information retrieval than time. This, in fact, aligns with the concept of *retroactive interference*, as new rules make previous ones vulnerable to decay (Ricker, Nieuwenstein, Bayliss, & Barrouillet, 2018; Dames & Oberauer, 2022).

Since the three basic operators, addition, contraction and revision, appeared to be the most distinctive building blocks, we modeled them using linear regression based on features describing the trial with respect to a rule (i.e., how many shapes are relevant for the rule, if there are any) and with respect to the rule system itself (i.e., how many rules are active at the time). For the models, we found a strong dependency on the trial only, across all operators. The strong dependency of the model to the structure of the trial itself (rather than to the rule system and memory operations) highlights the problem to discern the effects of manipulations of a rule system from the application of the rule itself. Whereas investigations of forgetting in declarative memory can directly rely on recognition of items, investigations of procedural memory often needs application of rules as an intermediate step, thereby assessing forgetting indirectly. While being a drawback for investigating cognitive operations, it can be an advantage for applications, since real-world rule systems are not occurring free of the specific instances and contexts.

Acknowledgments

This project has been funded by a grant to MR in the DFG-projects 529624975, 427257555 and 318378366.

References

Alchourrón, C., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: partial meet contraction and revision functions. *Journal of Symbolic Logic*, *50*(2), 510–530.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York: Oxford University Press.

Beierle, C., Kern-Isberner, G., Sauerwald, K., Bock, T., & Ragni, M. (2019). Towards a general framework for kinds of forgetting in common-sense belief management. *Künstliche Intelligenz*, *33*, 57-68. doi: https://doi.org/10.1007/s13218-018-0567-3

Brand, D., Dames, H., Puricelli, L., & Ragni, M. (2022). Rule-based categorization: Measuring the cognitive costs of intentional rule updating. In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenzoni (Eds.), *Proceedings of the 44th annual meeting of the cognitive science society* (pp. 2810–2817).

Dames, H., Brand, D., & Ragni, M. (2022). Evidence for multiple mechanisms underlying list-method directed forgetting. In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenzoni (Eds.), *Proceedings of the 44th annual conference of the cognitive science society.* Retrieved from `https://escholarship.org/uc/item/46921378`

Dames, H., & Oberauer, K. (2022). Directed forgetting in working memory. *Journal of Experimental Psychology*, *151*(12).

Delgrande, J. (2014). Towards a knowledge level analysis of forgetting..

Gollwitzer, P. M., & Brandstätter, V. (1997). Implementation intentions and effective goal pursuit. *Journal of Personality and Social Psychology*, *73*(1), 186–199. doi: 10.1037/0022-3514.73.1.186

Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 650.

Maddox, W. T., Ashby, F. G., Ing, A. D., & Pickering, A. D. (2004). Disrupting feedback processing interferes with rule-based but not information-integration category learning. *Memory & Cognition*, *32*(4), 582–591.

Oettingen, G., & Gollwitzer, P. M. (2010). Strategies of setting and implementing goals: Mental contrasting and implementation intentions. In J. E. Maddux & J. P. Tangney (Eds.), *Social psychological foundations of clinical psychology* (pp. 114–135). New York: Guilford Press.

Ricker, T., Nieuwenstein, M. R., Bayliss, D. M., & Barrouillet, P. (2018). Working memory consolidation: Insights from studies on attention and working memory: An overview of working memory consolidation. *Annals of the New York Academy of Sciences*, *1424*(1), 8–18. doi: 10.1111/nyas.13633