Rule-Based Categorization: Measuring the Cognitive Costs of Intentional Rule Updating

Daniel Brand[1] **(daniel.brand@hsw.tu-chemnitz.de)**
Hannah Dames[2] **(damesh@cs.uni-freiburg.de)**
Leonardo Puricelli[1] **(leonardo.puricelli@hsw.tu-chemnitz.de)**
Marco Ragni[1,2] **(marco.ragni@hsw.tu-chemnitz.de)**
[1] Predictive Analytics, Chemnitz University of Technology, Germany
[2] Department of Computer Science, University of Freiburg, Germany

Abstract

The ability to categorize visual information is essential for human cognition. Often, this categorization is achieved via internalized rules. In rule-based categorization tasks, participants categorize stimuli according to given decision rules. In this study, we created a framework aimed at measuring the respective impact of single memory operations on task performance. We present a study investigating two central mental operations – the addition of a new and the update of an existing rule – by confronting participants with Alien images they needed to assign to planets. Both conditions showed interference effects for task performance with previously learned ones. We found improved categorization task performance when old and new rules were in accordance, but no significant effect for conflicting situations. Our experimental setting promises to be well-suited to investigate the impact of memory operations on participants' behavior in a controlled environment.

Keywords: Category learning; rule-based learning; memory operations; updating rules

Introduction

Facing unknown objects and situations on a daily basis, we are challenged to conduct an ensemble of cognitive manoeuvres in order to make sense of incoming visual information. For instance, when we need to judge whether a mushroom is edible or poisonous, we rely on our ability to combine visual input with prior knowledge to derive a decision (e.g., red + white spots = poisonous). A widespread paradigm to explore the cognitive mechanisms behind our ability to categorize information are category learning tasks.

In typical category learning experiments, exemplars of several categories are presented, showing relevant and irrelevant features of their respective group. Participants then have to assign the exemplars to a category, typically followed by feedback on whether their response was correct or not (e.g., Ashby & Waldron, 2000; Erickson & Kruschke, 2002; Hughes & Thomas, 2021; Zeithamova & Maddox, 2006).

There is still an ongoing debate about whether only a single system (Nosofsky & Johansen, 2000; Nosofsky & Kruschke, 2002; Nosofsky & Zaki, 1998), revolving around processes of storing instances or exemplars (e.g., Medin & Schaffer, 1978; Nosofsky, Clark, & Shin, 1989; Nosofsky, Kruschke, & McKinley, 1992) or multiple cognitive systems (e.g., Ashby, Alfonso-Reese, Waldron, et al., 1998; Ashby & O'Brien, 2005; Erickson & Kruschke, 1998; E. E. Smith & Grossman, 2008) are involved in solving such categorization tasks. The latter propose at least one separate system based on implicit learning during observation of exemplars as well as an explicit reasoning system for processing of rules.

Looking further into explicit rule-based categorization, simple rules (e.g., "All blue exemplars belong to group A. All red exemplars belong to group B."), and conjunctions or disjunctions of logical statements will lead to categories that are usually easy to describe verbally in a subsequent recall test. In contrast, categories generated using more complex, integrated information or implicit learning are harder or even impossible to verbalize (Ashby et al., 1998). Even though there is a lot of evidence regarding the role of working memory and executive attention in rule-based categorization tasks (Ashby et al., 1998; Maddox, Ashby, Ing, & Pickering, 2004; Waldron & Ashby, 2001; Zeithamova & Maddox, 2006), neither *single operations* of working memory like keeping information active, prioritizing, modification or protection from interference (Bledowski, Kaiser, & Rahm, 2010), nor their impact on performance have been engaged in previous studies.

The aim of the present study was to investigate the cognitive costs of altering previously-learned rules in a rule-based categorization task. As a starting point, we conducted a study assessing the influence of adding a new rule as well as a change to an existing rule on participants' performance. We estimated cognitive costs by the differences in error rates and reaction times between different memory operations. With this, we are laying the foundation for subsequent experimental testing aimed at investigating the effects of rule-updating and corresponding cognitive costs in a controlled framework. We conducted a study comparing two operations in a rule-based categorization scenario: Performing a change to an existing rule and adding an additional rule. Our analyses provided insights into the differences between the operations as well as the effects that occur for them. However, the experimental setting will also allow us to systematically assess and compare a variety of other operations.

The remainder of the paper is structured as follows: First, we will present an overview over moderators that have been found to affect performance in rule-based categorization tasks that have to be considered our the study design.

Second, we will describe our study design in detail. Third, we will present the analyses and results, leading to a discussion of the limits and potentials of this new experimental setting.

Moderators of category learning

Several previous studies investigating category learning have been designed to target suspected moderators affecting task performance. One method is to manipulate *similarity* between exemplars with shared features, in order to raise or lower the likelihood of them being associated (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976), distinguishing between emphasizing commonalities or differences *within* and *between* categories (see Brunel, Carvalho, & Goldstone, 2015; Carvalho & Goldstone, 2014; Kornell & Bjork, 2008; Zulkiply & Burt, 2013).

Another method is to vary *complexity* of typizations, meaning that participants have to either apply simple, category-defining rules or integrate more complex information implicitly (e.g. Ashby & Valentin, 2017; Hughes & Thomas, 2021; Maddox, Ashby, & Bohil, 2003; E. E. Smith, Patalano, & Jonides, 1998). Distinctions in category types are well established by findings that, for example, only implicit learning processes profit from an increased amount of trials (Hélie, Waldschmidt, & Ashby, 2010). This suggests that when focusing on explicit rule learning, the influence of implicit learning can be reduced by keeping the amount of trials low.

Moreover, even though participants receive feedback on their responses in standard designs, it has been shown that this typically enhances only learning of information-integration-based but not of rule-based categories (Filoteo, Maddox, Ing, Zizak, & Song, 2005; Maddox et al., 2004). Studies varying the delay or type of feedback further support this differentiation (Dunn, Newell, & Kalish, 2012; Maddox et al., 2003; Maddox, Love, Glass, & Filoteo, 2008; J. D. Smith et al., 2014).

Beside of that, it has been demonstrated that even minimal prior knowledge can influence category learning (Kaplan & Murphy, 2000). In order to reduce such influences, Carvalho and Goldstone (2017) recommend to use a scenario with no available background knowledge (e.g., a fantasy scenario) to investigate category learning.

Method

The goal of this study was to investigate how changing an established rule affected participants' performance in terms of their accuracy and response times compared to the addition of a new categorization rule.

Furthermore, for the addition of a new rule, we aimed at comparing two different variants: By designing the rules in a hierarchical manner, they can be understood as a decision tree. Rules that are located in the same sub-tree are therefore closer to each other and form a group, which we will call a *division* for the rest of this article. A new rule could either be consistent with the existing divisions (*within division*) or conflict with the existing hierarchy, linking features from different sub-trees (*across divisions*). Based on these operations, we used three conditions in our study: rule change, additional rules within division and additional rules across divisions.

Generally, across all conditions, situations can occur in which the rules interfere/interact with each other. If at least a part of the rule's precondition is fulfilled, it can be considered to be applicable or mix-ups between rules can become apparent. This can happen in two ways: First, the rules can contradict each other, leading to a *conflict*. Second, they could both lead to the same result, which we call a *consensus* situation. As the conflict will be most affected by mix-ups, we expect these situations to be substantially more challenging, while the consensus situations could actually improve the performance due to multiple rules being available to classify the item correctly. Furthermore, we expect a change of an existing rule to be easier than an additional rule, as it holds the total number of rules that have to be memorized and considered for classification constant, by allowing to *forget* the old rule. The reduced number of rules should not only improve the performance in terms of correctness, but also in terms of the time needed to classify an item. While the rule change might be more prone to mix-ups, e.g., artefacts of the old rule incorporated in the updated rule, even a slightly wrong memorization might still be sufficient to solve most tasks. Therefore, the performance for classifying items affected by a rule change is better compared to rule additions. Finally, for additions across divisions or within a division, there are two possibilities. On the one hand, it could be easier to stay within the same division, as it only requires to extend an already existing hierarchy. On the other hand, it might lead to more confusion and mix-ups, which could be reduced when breaking the hierarchies.

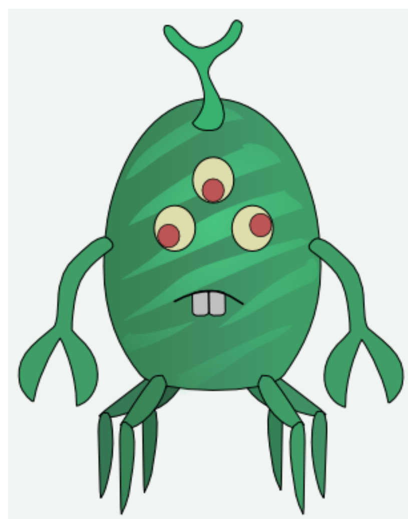

Figure 1: Example for the aliens used as stimuli. The depicted alien would have the following features: 1 antenna, 3 eyes, 2 arms, 6 legs, stripes and teeth.

Study Design

Participants needed to have an age between 18 and 35 years and be native English speakers in order to participate. After obtaining consent, participants were asked about their demographics and had to pass an attention-check before proceeding with the main study. The main study consisted of a rule-based classification task, in which participants were asked to categorize items based on several features using pre-

Table 1: Rules for the different phases and conditions as logical terms. Note that implicitly assumed negations originating from mutual exclusion of the rules are omitted.

Feature	Possible values
Antennae	None, One, Two
Arms	None, Two, Four
Eyes	One, Two, Three
Legs	None, Two, Four, Six
Teeth	Yes, No
Pattern	Plain, Striped, Dotted

viously memorized rules. In order to reduce influences of background knowledge (for more information, see Kaplan & Murphy, 2000), we used a fantasy scenario based on the stimuli used by Carvalho and Goldstone (2017). In our scenario, different aliens had to be assigned to one out of four planets based on their appearance (see Figure 1 for an example of the presented aliens).

The participants were presented with instructions and the rules needed to classify the aliens. Each planet had an unique combination of two features that allowed to categorize the aliens unambiguously. The features and the possible values are shown in Table 1. However, the rules were created in a hierarchical manner, so that one of the two features was shared with another planet and were presented. The initial set of rules was presented as follows:

If an alien has 2 antennae, it comes from either *planet 1* or *planet 2*.

If an alien has 2 arms, it comes from *planet 1*.
If an alien has 4 arms, it comes from *planet 2*.

If an alien has stripes, it comes from either *planet 3* or *planet 4*.

If an alien has 3 eyes, it comes from *planet 3*.
If an alien has 2 eyes, it comes from *planet 4*.

The hierarchy divides the planets into two divisions (Planet 1 & 2 and Planet 3 & 4). Planets were named *Sala*, *Diro*, *Kemi* and *Laru* with a randomized order of the names. Participants were instructed to give their response using the keyboard only, by pressing the letter S, D, K or L (in accordance to the first letter of the name of the respective planet). After the rules were presented, the participants were given one example task for each planet, which showed the respective rule again after the response was given. After that, a training phase with 20 tasks (5 of each planet) started. Throughout the whole study, after each given response, it was indicated for 0.5s if the response was correct, and if not, what the correct response was.

When the training phase was completed, participants were assigned to one of three conditions:

1. Rule Change: The rules for planet 1 and planet 2 are changed, now relying on teeth instead of arms.
2. Additional Rule (within division): An additional rule for planet 3 and planet 4, relying on a dotted pattern and the number of legs.
3. Additional Rule (across divisions): An additional rule for planet 1 and planet 4, relying on a dotted pattern and the number of legs.

Table 2 provides an overview of the rules for all conditions and phases. The aliens in the training phase were randomized, but constructed in a way that they avoided situations in which they would be categorized differently according to the changed rules later on (in order to avoid memory effects accidentally carrying over from the training phase). Following the instructions containing the new rules, another set of 28 tasks had to be completed by the participants (test phase). Depending on the condition, the set contained a special selection of tasks to investigate the effects of the rule change or the addition of a new rule. These test tasks will be described in detail in the following sections.

Rule Change

In the rule change condition, the old rules for distinguishing between *planet 1* and *planet 2* had to be replaced, while the high-level rule remained unchanged. In doing so, three different situations could occur: First, only the updated rule could be applicable to an alien. Second, the old rules and the updated rule could both be applicable, with both rules yielding the same categorization (*Consensus*). Third, the old and the updated rules could both be applicable, but would result in a different categorization (*Conflict*). A total of 6 tasks for each of the three cases was included, each with 3 tasks for *planet 1* and *planet 2*, respectively. The remaining 10 tasks consisted of aliens from *planet 3* and *planet 4* and thus were not affected by the rule changes.

Additional Rule within division

The addition of a new rule yielded more options than the rule changes. As both the new and the old rule are still valid, both rules had to be tested without interference of the other. For each, the new and the old rules, 4 tasks were included. Although the rules are mutual exclusive when not crossing the divisions (as the old and the new rule rely on the same feature), the tasks for testing single rules were constructed in a way that if one rule matched, the other rule was not matching with any feature (i.e., not only relying on the striped pattern, but also restricting the possible number of legs to 2 and 4 when testing the old rule in order to avoid matching parts of the new rule). Due to the rules being mutual exclusive, the conflicts and the cases with consensus also differ from the rule change condition. Therefore, interference could only occur partially on one feature. When referring to these situations, the conflict and consensus is annotated by the applicable rule. For consensus, a total of 4 tasks were included (with

Table 2: Rules for the different phases and conditions as logical terms. Note that implicitly assumed negations originating from mutual exclusion of the rules are omitted.

Phase	Planet 1	Planet 2	Planet 3	Planet 4
Training	2 Antenna $\wedge$ 2 Arms	2 Antenna $\wedge$ 4 Arms	Stripes $\wedge$ 3 Eyes	Stripes $\wedge$ 2 Eyes
Rule Change	2 Antenna $\wedge$ Teeth	2 Antenna $\wedge$ No Theeth	Stripes $\wedge$ 3 Eyes	Stripes $\wedge$ 2 Eyes
New Rule (within)	2 Antenna $\wedge$ 2 Arms	2 Antenna $\wedge$ 4 Arms	(Stripes $\wedge$ 3 Eyes) $\vee$ (Dots $\wedge$ 6 Legs)	(Stripes $\wedge$ 3 Eyes) $\vee$ (Dots $\wedge$ No legs)
New Rule (across)	(2 Antenna $\wedge$ 2 Arms) $\vee$ (Dots $\wedge$ 6 Legs)	2 Antenna $\wedge$ 4 Arms	Stripes $\wedge$ 3 Eyes	(Stripes $\wedge$ 3 Eyes) $\vee$ (Dots $\wedge$ No legs)

2 tasks for the old and the new rule, respectively), while a total of 8 tasks was used for the conflicts, as we expected those to have a higher variance. The remaining 8 tasks were, again, used as baseline, consisting of *planet 1* and *planet 2*.

Additional Rule across divisions

When testing the old and the new rules on their own, we relied on the same number of tasks as in the previous condition, using 4 tasks for each, new and old rules. When crossing divisions, a major difference occurs compared to conditions staying withing a division: As the rules are no longer mutually exclusive for all planets (for *planet 1* both rules are applicable at the same time), the consensus can now also be tested accordingly. This leads to a total of 3 tasks for testing the consensus, as new- and old-rule cases collapse into a single task for *planet 1*. However, for the conflicting cases, it must still be warranted that tasks can be solved unambiguously, which requires to avoid fulfilling the criteria for one of the rules. As in the previous condition, 8 tasks were used to test the conflict situations. The remaining 9 tasks provided the baseline, consisting of *planet 2* and *planet 3*.

Dataset

The data was acquired via online study on the platform Prolific[1]. Two studies were conducted, obtaining data from 85 participants in the first and 113 participants in the second study. Due to an error, the data for the third condition (additional rule across divisions) could not be used, leading to a total of 55 participants remaining from the first run. To rebalance conditions, assignment to the conditions was adjusted accordingly in the second run. Besides correction of the error, the only difference between our studies was an additional survey at the end of the second study, where participants had to write down how they remember the rules for each planet in free text. After combining both datasets and excluding participants with a sub-random performance ($n = 3$) in training phase, we obtained data from 165 participants (102 female, 62 male, 1 diverse). Over all three conditions, there are 53, 57 and 55 participants for rule change, rule addition within division and rule addition across divisions, respectively. Datasets and materials are publicly available on GitHub[2].

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

[2]https://github.com/Shadownox/cogsci-2022-rulebasedcat

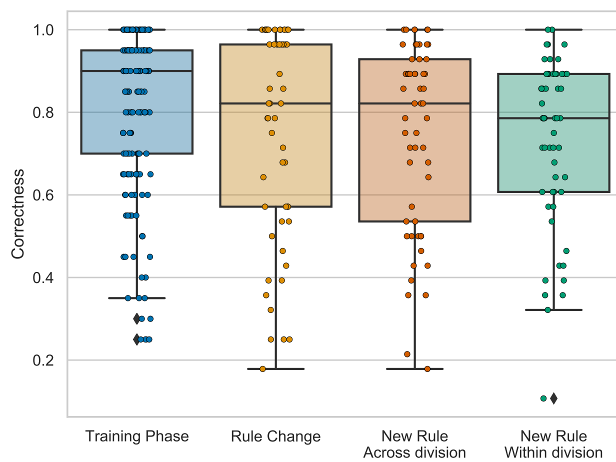

Figure 2: Proportion of correct responses for the different phases and conditions. Dots indicate individual participants.

Results

First, we analyze the general performance for the conditions. Figure 2 shows performance for the training phase and the test phases for all three conditions. Note that only training performance was used to exclude participants, which is why sub-random performances can occur in the test phase (e.g., by participants mixing up different rules, leading to systematic errors). The training performance is the best ($M = 0.9$, $MAD = 0.1$) by a substantial margin, followed by the rule change condition ($M = 0.82$, $MAD = 0.18$), the new rule conditions across divisions ($M = 0.82$, $MAD = 0.14$) and finally the addition of a new rule within the same division ($M = 0.79$, $MAD = 0.11$). This is not surprising, as all conditions put additional cognitive load on the participants and the rules learned in the training phase will slowly fade out over time.

Between the two conditions for rule additions, the performance was better across divisions than within a division, but showing a higher variance. This indicates that the hierarchy has a small effect on the performance, but it is not clear if the effect is stable or comes down to inter-individual differences in representing the rules.

The performance in the rule change condition was slightly

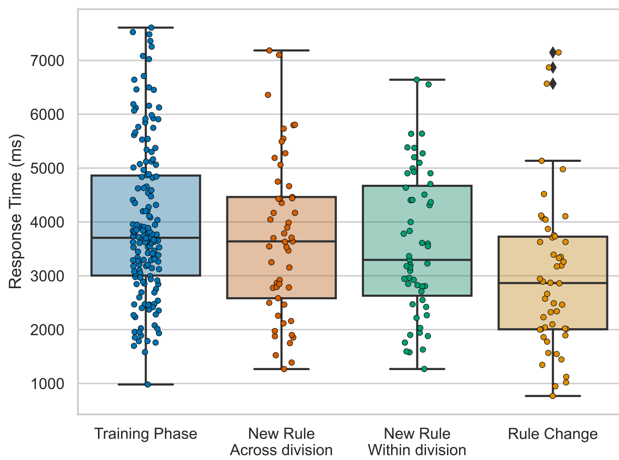

Figure 3: Mean reaction times for the different phases and conditions. Dots indicate individual participants. Participants with mean reaction times that deviate by more than two standard deviations were excluded.

better compared to the addition of a new rule, however the effect was much weaker than expected (mean performance: $change = 0.78$, $addition = 0.74$) and did not reach significance (Mann-Whitney U test: $U = 2524.5$, $p = 0.24$, all p-values are Bonferroni corrected to correct for multiple comparisons). However, when taking the time needed for classifying an item into account, another picture emerges. Figure 3 shows the mean reaction times for the different conditions. While there are only slight differences between training and the rule addition conditions, indicating that possible practice effects get cancelled out by the additional rule, the rule change condition had significantly faster response times compared to the additional rule conditions ($mean = 3508ms$ compared to $mean = 3838ms$, Mann-Whitney U test: $U = 2295$, $p = 0.04$). We assume that this is due to the lower number of rules that have to be considered and thus allowing participants to become faster with more practice.

Figure 4 shows the performances for all conditions broken down by the task category. For comparison, the training performance of participants in the respective condition group (training) and the test performance for tasks not affected by the condition (baseline) are included in blue. The first thing that becomes apparent is that there seems to be a recency effect. The baseline performance is substantially worse compared to the training performance, indicating that the memory strength for the rules faded out over time. This is also supported by the fact that the performance for new rules is always superior compared to the old rule.

In the following, we investigate the effects of interfering rules, i.e., *conflicts* and *consensus*. Generally, we can see in Figure 4 that the consensus always has the best performance (even compared to the training performance in some cases), while the conflicting situations are on the opposing side of the spectrum. However, for conflicting situations, an interaction with the previously described recency effect seems to occur: For the third condition (rule addition across divisions), the conflicting situation based on the new rule shows a higher performance compared to the the old rules. Overall, the consensus situations has a significantly better performance compared to the non-interfering situations ($mean = 0.816$ compared to $mean = 0.75$, Mann-Whitney U test: $U = 10908.5$, $p = 0.002$). For the conflicting situations, the opposite effect seems to be weaker due to the interaction with the recency effect and thus not reaching significance level ($mean = 0.71$ compared to $mean = 0.75$, Mann-Whitney U test: $U = 12729.5$, $p = 0.6$). This shows that the positive effects of multiple rules in this scenario seem to have a greater impact than potential conflicts emerging from them. A possible explanation would be a race-condition between rules. For the consensus situations, it would be beneficial if either rule wins. For conflicting situations, there is still the chance that a wrong rule is rejected, even if it won the race condition, limiting the negative impact.

At last, we analyze the results from the end-survey, where participants had to recall the rules (see Table 3). Note that these results required some interpretation, as the participants were allowed to provide free-text responses. For example, some participants reformulated the rules based on other rules utilizing the exclusion principle (e.g., instead of *has two antennae*, it would also be valid to use *has not stripes or dots*) and therefore also be counted as correct (about 7% of the participants provided such rules). Responses that provided a conjunction of parts from the old and the new rule were counted as a mix-up, while mentions of the old rule without parts of the new rule are counted as old (same holds for new rules, respectively). Optionally mentioned parts of rules (e.g., *might also have dots*) were omitted. Due to the survey being added in the second iteration of the study, the responses are not available for all participants. As the across-divisions condition was only conducted in the second iteration of the study, the condition is over-represented in this analysis. However, while the results should not be used for strong quantitative statements, some tendencies can be seen. Across all conditions, the recall rate of the rules are similar for the manipulated rules: the rules are only recalled correctly in about 50% of the cases. The baseline rules were recalled slightly better with about 60%. This contradicts the assumed recency effect. When looking at the type of errors it also becomes apparent that most errors originate from the participants only being able to recall mixed rules, or the old rule. At first, these findings seem to contradict the performance, as the classification performance for new rules is generally better. However, when analyzing if the participants do indeed follow their rules, a different picture emerges. To do this, we derived logical rules comparable to the ones shown in Table 2 as literally as possible from the written responses and applied the rules in order to obtain the responses from the rules. Overall, participants only selected an option in line with their recalled rules in about 61% of the tasks on average. Only 15% of the errors

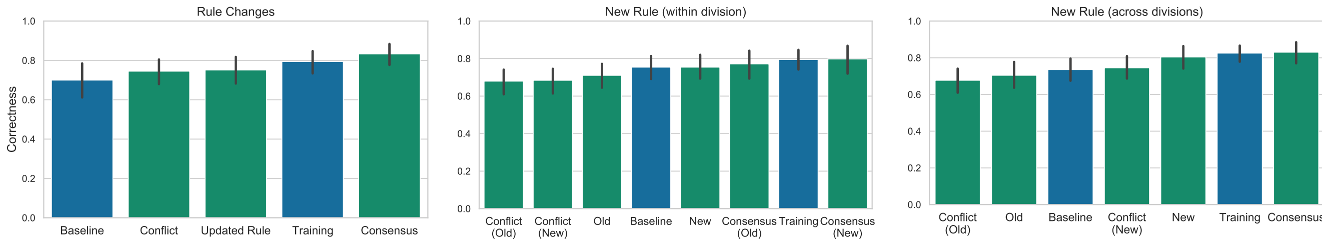

Figure 4: Mean correctness in the test phase for the three conditions. *Baseline* refers to the classification performance for items that were not affected by the changed rules, *Training* denotes the performance of the participants in the training phase.

Table 3: Number of correctly and incorrectly recalled rules for rules affected by the condition (left) and the baseline (right). Additionally, the errors for the condition-affected rules are broken down by potential pitfalls (mix-ups where both rules formed a wrong rule, only recalled the old/new rule correctly).

	Correct		Wrong		Mix	Old	New
Change	24	37	26	13	12	2	
Within	28	38	28	18	8	7	3
Across	53	83	53	23	15	18	4

where due to the rule. When having recalled a wrong rule, for about 64% of the cases the correct answer was selected nevertheless. This indicates that participants were not able to verbalize learned rules in a logically correct form.

Discussion

The ability to manipulate prior category knowledge (e.g., learning new and update existing categories in the light of new information) is an essential cognitive ability. In this work, we developed an experimental setting that allows to investigate the impact of different rule-updating operations on recalling and applying memorized rules. As a starting point, we assessed the influence of an addition of a new rule as well as a change to an existing rule. Although we expected the rule change to have less impact on categorization performance compared to the rule addition (due to the reduced complexity of rules that have to be remembered), our results did not supported this assumption. However, participants were significantly faster to apply a changed as compared to a new rule, which still indicates that processes depending on the number or complexity of the rules were present. There were no significant differences between the different addition conditions that stayed within or broke the initial hierarchy.

When investigating the interactions between old and new rules after a change, we found that accuracy was enhanced in situations where both rules resulted in the same classification. This was in line with our expectation, that both rules can contribute to the categorization (e.g., due to race conditions between rules). For the case of a rule change, this means that the old rule left some cognitive artefacts that affected participants' decisions. However, for a conflicting situation, the effect was no longer apparent. We assume that this is due to the interaction with another effect: Assuming a recency effect, new rules might be more active compared to old ones, potentially counteracting the influence of cognitive artefacts from the old rule.

When looking at participants' ability to verbally recall the applied rules, we found that participants were only able to recall about 50% of the rules for categories that required to add or update rules and 60% of the rules that remained unchanged. Furthermore, when the recalled rules would lead to an incorrect categorization, participants applied the correct rule instead of their recalled rule in about 64% of the cases. This is surprising, as the rules did not have to be (implicitly) learned over time, but where explicitly given to the participants. As Ashby et al. (1998) showed, participants are usually able to verbalize their decisions, they might use a similar strategy when recalling the rules and verbalize their decisions for typical instances. While the self-reports should be interpreted with caution, the general trend suggests that participants implicitly encoded rules that were not necessarily in line with the instructed ones but allowed them to categorize the images, more often than not, accurately. We speculate that those implicitly learned rules interfered (or replaced) the explicitly-instructed rules. Changing the free-text to a more constraint way to query for the rules should be considered for further investigations in this direction. Participants need to be queried for the rules at several stages of the study, allowing to gain insight into the processes behind the effect.

Put together, the main contribution of this work lies in the presented experimental setting, which allows a systematical investigation of rule-manipulations. Our findings hint at several possible processes that we plan to further investigate by using the presented experimental scenario. By proceeding with our investigations, we aim to establish an estimation of the cognitive costs of mental operations involved in rule-updating.

Acknowledgements

This research was supported by the German Research Foundation, DFG (Grant RA1934/5-1 and RA1934/8-1) within the SPP 1921 "Intentional Forgetting" and by the Saxony State Ministry of Science and Art (SMWK3-7304/35/3-2021/4819) research initiative "Instant Teaming between Humans and Production Systems" on the basis of the budget passed by the deputies of the Saxony state parliament.

References

Ashby, F. G., Alfonso-Reese, L. A., Waldron, E. M., et al. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological review*, *105*(3), 442.

Ashby, F. G., & O'Brien, J. B. (2005). Category learning and multiple memory systems. *Trends in cognitive sciences*, *9*(2), 83–89.

Ashby, F. G., & Valentin, V. V. (2017). Multiple systems of perceptual category learning: Theory and cognitive tests. In *Handbook of categorization in cognitive science* (pp. 157–188). Elsevier.

Ashby, F. G., & Waldron, E. M. (2000). The neuropsychological bases of category learning. *Current Directions in Psychological Science*, *9*(1), 10–14.

Bledowski, C., Kaiser, J., & Rahm, B. (2010). Basic operations in working memory: contributions from functional imaging studies. *Behavioural brain research*, *214*(2), 172–179.

Brunel, L., Carvalho, P. F., & Goldstone, R. L. (2015). It does belong together: Cross-modal correspondences influence cross-modal integration during perceptual learning. *Frontiers in psychology*, *6*, 358.

Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & cognition*, *42*(3), 481–495.

Carvalho, P. F., & Goldstone, R. L. (2017, 03). The sequence of study changes what information is attended to, encoded, and remembered during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*. doi: 10.1037/xlm0000406

Dunn, J. C., Newell, B. R., & Kalish, M. L. (2012). The effect of feedback delay and feedback type on perceptual category learning: the limits of multiple systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 840.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, *127*(2), 107.

Erickson, M. A., & Kruschke, J. K. (2002). Rule-based extrapolation in perceptual categorization. *Psychonomic Bulletin & Review*, *9*(1), 160–168.

Filoteo, J. V., Maddox, W. T., Ing, A. D., Zizak, V., & Song, D. D. (2005). The impact of irrelevant dimensional variation on rule-based category learning in patients with parkinson's disease. *Journal of the International Neuropsychological Society*, *11*(5), 503–513.

Hélie, S., Waldschmidt, J. G., & Ashby, F. G. (2010). Automaticity in rule-based and information-integration categorization. *Attention, Perception, & Psychophysics*, *72*(4), 1013–1031.

Hughes, G. I., & Thomas, A. K. (2021). Visual category learning: Navigating the intersection of rules and similarity. *Psychonomic Bulletin & Review*, 1–21.

Kaplan, A. S., & Murphy, G. L. (2000). Category learning with minimal prior knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(4), 829.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological science*, *19*(6), 585–592.

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.

Maddox, W. T., Love, B. C., Glass, B. D., & Filoteo, J. V. (2008). When more is less: Feedback effects in perceptual category learning. *Cognition*, *108*(2), 578–589.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological review*, *85*(3), 207.

Nosofsky, R. M., Clark, S. E., & Shin, H. J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, memory, and cognition*, *15*(2), 282.

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review*, *7*(3), 375–402.

Nosofsky, R. M., & Kruschke, J. K. (2002). Single-system models and interference in category learning: Commentary on waldron and ashby (2001). *Psychonomic Bulletin & Review*, *9*(1), 169–174.

Nosofsky, R. M., Kruschke, J. K., & McKinley, S. C. (1992). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 211.

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science*, *9*(4), 247–255.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive psychology*, *8*(3), 382–439.

Smith, E. E., & Grossman, M. (2008). Multiple systems of category learning. *Neuroscience & Biobehavioral Reviews*, *32*(2), 249–264.

Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition*, *65*(2-3), 167–196.

Smith, J. D., Boomer, J., Zakrzewski, A. C., Roeder, J. L., Church, B. A., & Ashby, F. G. (2014). Deferred feedback sharply dissociates implicit and explicit category learning. *Psychological science*, *25*(2), 447–457.

Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic bulletin & review*, *8*(1), 168–176.

Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & cognition*, *34*(2), 387–398.

Zulkiply, N., & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition*, *41*(1), 16–27.