Van 't Wout, F. M., & Jarrold, C. (2022). How does language support the acquisition of novel cognitive tasks? Investigating the role of task complexity and task instructions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(3), 416-431. https://doi.org/10.1037/xlm0001131

How does language support the acquisition of novel cognitive tasks? Investigating the role of task complexity and task instructions

Félice van 't Wout & Christopher Jarrold

Abstract: 249 words; Text: 11578 words; 36 references; 2 tables; 6 figures.

This manuscript was submitted to Journal of Experimental Psychology: Learning, Memory, and Cognition on the 7th of January 2022.

Abstract

Recent findings have shown that language plays an important role in the acquisition of novel cognitive tasks (Van ‘t Wout & Jarrold, 2020). The current study sought to elucidate the factors that influence the contribution of language to novel task learning, focussing specifically on the role of task complexity (defined by the number of stimulus-response rules per task) and the role of task instructions (by comparing trial-and-error learning to instruction-based learning). In each experiment participants were required to learn the correct response to novel sets of picture stimuli. When analysed as a function of stimulus occurrence within a task, both experiments found that initial performance was worse under articulatory suppression (AS; verbal distractor task) than under foot tapping (FT; non-verbal distractor task), but only if the task was more complex (consisting of 6 S-R rules) and not if it was less complex (consisting of 3 or 4 S-R rules), suggesting that the acquisition of a simpler task by trial-and-error might not require verbal mediation. Experiment 2 furthermore found that the role of language was modulated by the manner of acquisition: For trial-and-error learning, the detrimental effect of AS increased and then decreased again as a function of stimulus occurrence. Conversely, for instruction-based learning, AS exclusively affected the first few stimulus occurrences, suggesting that participants are able to create a verbal representation of the task during the instruction phase. Together, these experiments demonstrate that the role of language in novel task learning is modulated by task complexity and task instructions.

Keywords: Skill acquisition, instruction following, language, learning.

1. Introduction

We are all inhabitants of an ever-changing, dynamic world, and it is essential to our survival and success that we are able to flexibly adapt to our environment when required. In order to do so we must rely on the many skills we already possess. However, to succeed in a changing environment we must also frequently learn new skills. Humans are very adept at skill acquisition, and are able to acquire novel cognitive skills with relative ease and very little practice, especially compared to other species.[1]

Theories of skill acquisition (e.g., Anderson, 1982) and instruction following (e.g., Brass, Wenke, Spengler & Waszak, 2009; Brass, Liefooghe, Braem, & De Houwer, 2017) have argued that the first phase of learning a novel skill is aided by language. According to Anderson (1982), language supports the maintenance of task rules in working memory during an initial “declarative phase”. Similarly, Brass et al.’s (2017) model of instruction following includes an “instruction phase”, during which linguistic information is translated into a task model. Both of these theories suggest that the role of language is especially important during the early stages of skill acquisition, and then declines with practice. Surprisingly, until recently there was no evidence to support the claim that the role of language is most crucial during the early stages of learning. Van ‘t Wout and Jarrold (2020) provided the first evidence consistent with this claim: In two experiments, participants were required to learn a series of novel tasks (consisting of five arbitrary stimulus-response (S-R) mappings) by trial-and-error. On each trial, participants responded to a centrally presented line drawing of an object with one of five keyboard keys. On incorrect trials, feedback was provided to aid learning. To investigate the role of language, participants performed either articulatory suppression (AS; a verbal distractor task which blocks the use of inner speech, e.g., Baddeley, Chincotta & Adlam, 2001), foot tapping

---

[1] For example, Nakahara, Hayashi, Konishi and Miyashita (2002) managed to teach macaque monkeys a simplified cognitive set shifting task, but only after several months of training (also see Stoet & Snyder, 2003). Humans, on the other hand, were able to learn the same task with very little practice.

(FT; a non-verbal distractor task well-matched to AS in terms of difficulty, e.g., Miyake, Emerson, Padilla & Ahn 2004) or no distractor task during the first or second half of each task. Results showed that accuracy was decreased under AS compared to FT, but only during the first half of each task, and not when the task was well-practised (during the second half). The precise nature of the role of language was further elucidated by a more detailed analysis, in which error data were plotted as a function of practice (indexed by stimulus occurrence within each task). That analysis revealed that the detrimental role of AS on performance followed a bow-shaped curve: For the first few stimulus occurrences, performance was no worse under AS compared to FT. The detrimental effect of AS then emerged, and subsequently disappeared again. These results suggest that the role of language emerges as participants gradually construct a linguistic representation of the task; and then disappears with practice as performance ceases to rely on language.

The results of Van ‘t Wout and Jarrold (2020) are consistent with the abovementioned theories of skill acquisition and instruction following, suggesting that language might be used to establish and maintain the S-R rules in verbal working memory during the initial stages of learning. They are also in line with some previous findings, but not others. Specifically, one previous task switching study manipulated the phonological similarity of the stimulus names in a task (Van ‘t Wout, Lavric & Monsell, 2013). Performance was worse for tasks containing phonologically similar items compared to phonologically dissimilar items, but only right at the beginning of the experiment, when participants were practicing the tasks in single task blocks. Conversely, phonological similarity did not affect task switching performance in mixed blocks once the tasks were well-practiced, consistent with a diminishing role of language with practice. Additionally, Monsell and Graham (2021) have recently manipulated the phonological similarity of the object names in a choice RT task, and found that performance was worse for phonologically similar items only on the first few encounters with each stimulus, and not thereafter.

However, the results of other studies have been inconsistent with such a transient role of language in novel task learning. For example, Kray, Eber and Karbach (2008) investigated the effect of AS on task switching performance, and found that the detrimental effect of AS persisted even after more than a thousand trials. Additionally, other studies investigating the effect of verbal labelling on performance have shown that the beneficial effects of verbal labelling can sometimes *increase* during practice (Lupyan & Swingley, 2012; Ferdinand & Kray, 2017). One possible explanation for these seemingly inconsistent results is that the precise contribution of language to novel task learning is determined by the nature of the task, and the manner in which it is acquired. For example, if participants are using language to create a verbal representation of the S-R rules in working memory, then the role of language would be expected to be restricted to the early stages of learning, as predicted by theories of skill acquisition (e.g., Anderson, 1982). However, in task switching paradigms, language might be used to retrieve the relevant task goal into working memory on each trial (Miyake et al., 2004), which could explain why Kray et al. (2008) found that the effect of a verbal distractor task did not disappear with practice. Together, these results suggest that the role of language is multi-faceted, and likely depends on the task requirements. The aim of the experiments reported here was to investigate the role of language in learning novel tasks. Consistent with theories of skill acquisition (Anderson, 1982) and instruction following (Brass et al., 2009), the findings of Van ‘t Wout and Jarrold (2020) and Monsell and Graham (2021) suggest that language is used to establish a verbal representation of the S-R rules during the early stages of learning. However, the precise nature of the contribution of language to novel task learning remains poorly specified within these theories. For this reason, the aim of the two experiments reported here was to examine the factors that influence the role of language in novel task learning, focussing in particular on the complexity of the task (see below), and the manner of acquisition (trial-and-error learning versus instruction-based learning).

Specifically, Experiment 1 investigated how the acquisition of a new task is influenced by the complexity of the task (determined by the number of S-R rules per task; cf. Van ‘t Wout, 2018). With regards to the role of language, the most likely outcome would be that the role of language in task learning is diminished when the task is less complex, on the assumption that a less complex task is proceduralised more rapidly. Accordingly, the effect of articulatory suppression on performance would be expected to be more short-lived for less complex tasks than for more complex tasks. Experiment 1 explored this possibility, by manipulating the number of S-R mappings per task so that it was either more complex (six S-R mappings) or less complex (three S-R mappings; for a similar manipulation see Van ‘t Wout, 2018).[2] To investigate the role of language in learning, each task was performed either under articulatory suppression (verbal distractor task), foot tapping (a non-verbal distractor task), or in the absence of a distractor task.

Experiment 2 was designed to investigate whether language differentially affects trial-and-error learning and instruction-based learning. Like Experiment 1, Experiment 2 varied the distractor task to be performed (articulatory suppression or foot tapping), and the complexity of the tasks. With regards to the latter manipulation, the more complex condition again consisted of six S-R rules per task, but (prompted by the outcome of Experiment 1) in Experiment 2 the less complex condition comprised four (rather than three) S-R rules per task. Crucially, Experiment 2 also manipulated the manner of acquisition (trial-and-error learning or instruction-based learning). Unlike the study by Van ‘t Wout and Jarrold (2020) and Experiment 1, in which participants were required to learn each task by trial-and-error, many cognitive psychology experiments as well as many real-life situations (such as learning to drive a car or compiling flat pack furniture) involve the acquisition of a novel task via instructions. Whether or not the role of language differs in instruction-based and trial-and-error learning, depends in part on whether participants are able to “proceduralise” the task during the

---

[2] For the more complex condition, six S-R rules were chosen as this was similar to a previous study (Van ‘t Wout & Jarrold, 2020), which used five S-R rules per task. For the less complex condition, three S-R rules were chosen (two S-R rules per task was not a viable option as the experiments did not include stimulus repetitions).

instruction phase, prior to task performance. Studies with adults (Cohen-Kdoshay & Meiran, 2009) and children (Verbruggen, McLaren, Pereg & Meiran, 2018) using instruction-based learning paradigms have found that in some cases, instructions can influence behaviour in a way that suggests instant "proceduralisation:" without any practice. Brass, Wenke, Spengler and Waszak (2009) claim that this powerful effect of instructions on performance (or “prepared reflex”) is mediated by language, though there is no direct evidence to support this assumption. If participants are able to proceduralise instructions prior to task performance, then the effect of articulatory suppression may be more short-lived (or even absent) in instruction-based learning, compared to trial-and error learning. Experiment 2 explored this possibility.

Experiment 2 furthermore asked whether task complexity affects instruction-based learning and trial-and-error learning differently. A previous study by Ruge, Karcz, Mark, Martin, Zwosta and Wolfensteller (2018) found that task complexity affected only trial-and-error learning, which they deemed more demanding on working memory than instruction-based learning. However, their task (specified by four S-R rules at most) was arguably not sufficiently taxing on adult’s working memory to capture such an effect. Therefore, one of the aims of Experiment 2 was to further investigate trial-and-error learning and instruction-based learning across two levels of complexity.

Finally, we also conducted two additional analyses, which aimed to further elucidate the process by which a novel task is acquired. Firstly, a serial order analysis examined performance as a function of response finger in order to determine whether participants acquire novel tasks in a serial manner (from left to right), or not. Secondly, a post-error analysis explored the possibility that “post error slowing” (Rabbitt, 1966), a robust phenomenon often observed in choice reaction time experiments, might not occur in the early stages of trial-and-error learning, as participants take time to update their mental representation of the task-set following a correct response.

In summary, the experiments reported here aimed to further explore how people learn new tasks, by investigating whether the previously found effect of AS on task acquisition is modulated by the complexity of the task and the manner of acquisition (via instructions or via trial-and-error learning). In Experiment 1, participants were required to learn novel tasks of greater (six S-R rules per task) or lesser (three S-R rules per task) complexity via trial-and-error. Experiment 2 also manipulated the complexity of the task (four versus six S-R rules); but the main aim of Experiment 2 was to compare the effect of articulatory suppression on trial-and-error learning versus instruction-based learning.

2. Experiment 1 Method

2.1 Participants

Forty-eight participants (aged between 18 and 27 [mean age= 20]; 40 female) took part in Experiment 1. This number of participants was determined in advance (and data were not analysed prior to completion of data collection), as it was constrained by the between-subject counterbalancing of certain factors (see below for more detail). Data from one participant who did not complete the experiment were replaced. All participants provided informed consent prior to taking part and received either course credit or £12 in return for their participation. Both experiments were approved by the University of Bristol's School of Psychological Science Human Research Ethics Committee (ID 79562 and ID 86903, respectively).

2.2 Procedure

The experimental task was a modified version of the task employed by Van 't Wout and Jarrold (2020), programmed in PsychoPy (Peirce et al., 2019) and run on a Dell Laptop. Specifically, participants had to learn six novel tasks by trial-and-error. Each task consisted of a set of six or three picture stimuli, requiring a unique keyboard response (x, c, v, b, n or m on a standard QWERTY keyboard). All pictures were black and white line drawings, selected from the International Picture Naming Project (IPNP; Bates et al., 2003).

Nine sets of three images were created (see Table 1); and combinations of two sets were used to create sets of six pictures for the six stimulus-response (6 S-R) condition. The combination of sets was balanced between participants, so that across participants, each image featured in the three S-R and six S-R condition. Items within each set were selected as to avoid phonological, visual or semantic similarity. All items were selected to have a naming agreement of at least 98% (proportion of all trials on which participants produced the dominant target name). Additionally, mean naming latency (an indirect measure of word frequency, Oldfield & Wingfield, 1965) was matched across sets. Naming latencies and naming agreement data were obtained from the IPNP (Bates et al., 2003).

At the beginning of each task, participants were informed of the number of S-R mappings in the task (three or six); and of the distractor task condition (articulatory suppression, foot tapping or no distractor task; see below for more detail). Importantly, participants were not shown the stimuli prior to the start of each task. Instead, participants were told that they must learn the task by trial-and-error, and to try their best to respond as quickly and accurately as possible.

Participants performed 300 trials with each of the six novel tasks. As a result, for each task each stimulus was responded to 100 times in the three S-R condition, and 50 times in the six S-R condition. For each task, participants were required to respond to a centrally presented stimulus by pressing a key on the computer keyboard. In the six S-R condition, participants were instructed to place the ring, middle and index finger of each hand on the x, c, v, b, n and m keys of the keyboard. In the three S-R conditions, half of the participants used only the left hand (x, c, and v keys), and half used only the right hand (b, n and m keys). Trials within a block were pseudorandomized so that there were no immediate stimulus repetitions.

Each trial began with a 250ms fixation cross, followed by the stimulus, which remained on screen until a response was made. If the response was incorrect, the word “Error!” appeared on screen for 1000ms (see Figure 1 for an example of the trial sequence). Each novel task was performed under one of three distractor task type conditions: articulatory suppression (AS; saying “tick, tick, tick”), foot tapping (FT; tapping one foot) or no distractor task. AS and FT were performed to the beat of a metronome set to 100 beats per minute. Participants were instructed to ignore the metronome during the no distractor task condition. The experimenter remained present throughout to experiment to ensure that participants were indeed performing the distractor tasks as required. Each distractor task condition was performed twice: once in the three S-R condition and once in the six S-R condition. The order of distractor task type and S-R condition was balanced between participants; as was the assignment of stimuli to responses and stimuli to distractor task conditions.

Prior to completing the experimental task, all participants completed a baseline “colour matching” task. This task required participants to respond to a set of three pictures (also selected from the IPNP) which could be presented in either blue or green. On each trial, half of the participants had to press the “a” key if the picture was blue; and the “l” key if the picture was green (the response assignment was reversed for the other half of the participants). A blue and green circle remained on the screen throughout the experiment at the respective response locations (bottom left and right corner of the screen), to minimise the memory load of the task. The trial sequence was otherwise identical to the experimental task. Participants performed four blocks of 30 trials in the baseline condition: one practice block, followed by a further three blocks (one per distractor task condition; the order of distractor task conditions was balanced between participants). The purpose of the baseline task was twofold: 1) to familiarise participants with the distractor tasks; and 2) to ensure that AS was no more difficult than FT under conditions that do not require the acquisition of a novel task. Participants were tested one at a time, in a quiet room. In total, the experiment lasted approximately one hour, after which participants were thanked and debriefed.

#	Set 1	RT	%	#	Set 4	RT	%	#	Set 7	RT	%
1	book	656	100	10	car	751	100	19	train	838	100
2	pig	855	100	11	tree	796	100	20	house	745	98
3	watch	780	100	12	hand	723	98	21	bell	703	100
	Mean	**764**	**100**		**Mean**	**757**	**99**		**Mean**	**762**	**99**
#	**Set 2**	**RT**	**%**	**#**	**Set 5**	**RT**	**%**	**#**	**Set 8**	**RT**	**%**
4	leaf	848	100	13	sun	762	100	22	dog	702	100
5	hat	684	98	14	bread	773	98	23	key	738	100
6	tent	744	100	15	pen	753	100	24	moon	804	100
	Mean	**758**	**99**		**Mean**	**763**	**99**		**Mean**	**748**	**100**
#	**Set 3**	**RT**	**%**	**#**	**Set 6**	**RT**	**%**	**#**	**Set 9**	**RT**	**%**
7	fish	777	100	16	eye	700	98	25	sock	712	100
8	chair	732	100	17	horse	809	100	26	cheese	843	100
9	box	753	100	18	comb	717	100	27	heart	720	100
	Mean	**754**	**100**		**Mean**	**742**	**99**		**Mean**	**758**	**100**

Table 1. Picture names for the stimulus sets used in Experiment 1. Stimuli from two sets were combined for the 6 S-R condition. Stimuli were matched for percent name agreement (%) and RT target mean (in ms; mean latency for dominant responses only).

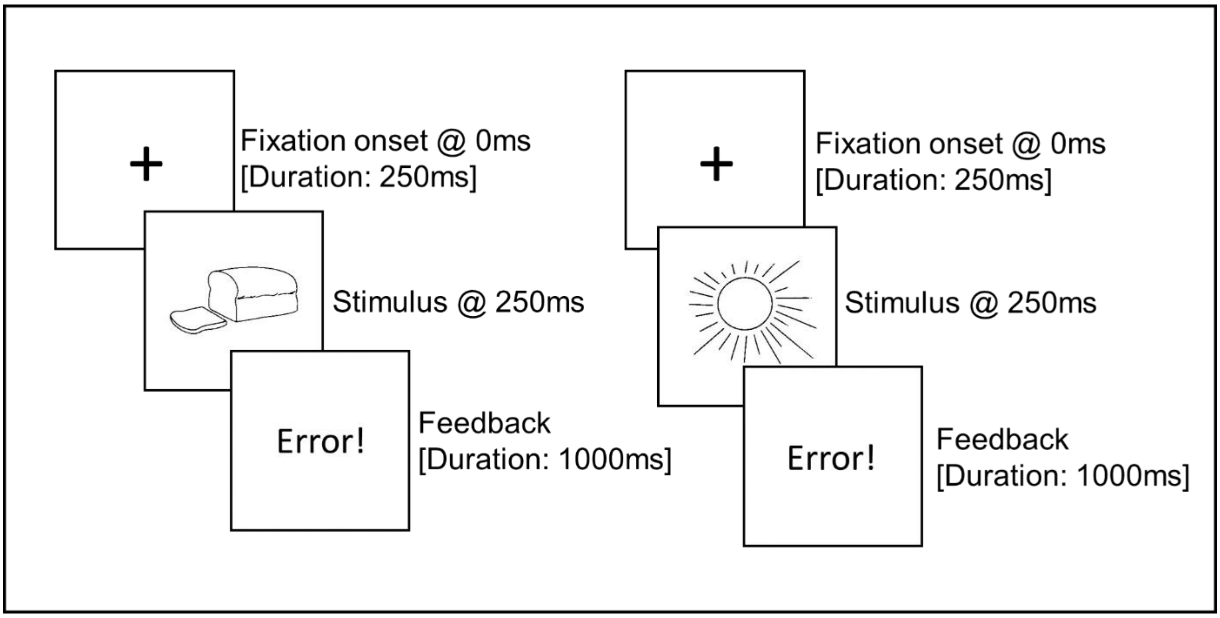

Figure 1. Example of a sequence of two trials in Experiments 1 and 2.

3. Experiment 1 Results

Reaction times greater than 5000ms (0.2% of correct responses) were removed from the data set prior to conducting the analyses described below. Throughout, the number following the ± symbol indicates the standard error of the mean (SEM).

3.1 Baseline task

A one-way repeated measures ANOVA with the within-subjects factor distractor task type (AS, FT or none) was run on the mean correct RT and % error data from the baseline task. With regards to the % error data, a significant main effect of distractor task type revealed that error rates were increased under AS (7.0±0.6%) and FT (5.6±0.7%) compared to when participants did not have to perform a distractor task (3.9±0.5%), F(2,94)=8.61, p=.001, η_p^2=.155 (Huynh-Feldt; H-F[3]). Further one-way ANOVAs found significant differences between AS and the control condition, F(1,47)=24.58, p<.001, η_p^2=.343, and between FT and the control condition, F(1,47)=5.00, p=.030, η_p^2=.096; but not between AS and FT, F(1,47)=2.73, p=.105, η_p^2=.055.

With regards to the mean correct RT analysis, RTs were significantly greater under FT (590±22ms) compared to AS (523±16ms) and the no distractor task condition (505±14ms), F(2,94)=18.37, p<.001, η_p^2=.281 (H-F). Further ANOVAs revealed significant differences between FT and the no distractor task condition, F(1,47)=28.56, p<.001, η_p^2=.378, and between FT and AS, F(1,47)=21.72, p<.001, η_p^2=.316; but not between AS and the no distractor task condition, F(1,47)=1.61, p=.210, η_p^2=.033.

---

[3] For main effects of, and interactions with, within-subject variables with three or more levels, Huynh-Feldt corrected F and p-values were reported (with uncorrected degrees of freedom).

Importantly, analysis of the baseline data showed no evidence that AS was a more difficult distractor task than FT: Accuracy did not differ significantly between AS and FT[4] and RTs were significantly *faster* under AS, so there was no evidence to suggest that AS is a more difficult distractor task per se.

3.2 Experimental task

To examine whether the contribution of language to novel task learning is modulated by the complexity of the task, RTs and error rates were analysed as a function of distractor task and S-R condition. Additionally, as it has previously been shown that the detrimental effect of AS is modulated by practice (Van 't Wout & Jarrold, 2020), the data were also analysed as a function of stimulus occurrence. Note that only the first 50 occurrences of each stimulus (grouped into pairs) were included in the analysis. This means all the data were included for the 6SR condition (300 trials per task), but only the first half of the data (150 trials per task) were included for the 3SR condition. Importantly, RT and accuracy reached asymptote well before then in the 3SR condition (see Figure 2), and so this filtering of the data was deemed appropriate and necessary in order to make the analysis comparable between the 3SR and 6SR conditions. Accordingly, a 3 (distractor task type: AS, FT or none) x 2 (task complexity: 3SR or 6SR) x 25 (stimulus occurrence pair: 1 to 25) repeated measures ANOVA was run on the % error data and the mean correct RT data.

For the % error data (see Figure 2), this analysis revealed significant main effects of distractor task type, F(2,94)=24.25, p<.001 (H-F), η_p^2=.340, complexity, F(1,47)=116.82, p<.001, η_p^2=.713, and stimulus occurrence, F(24,1128)=213.00, p<.001, η_p^2=.819 (H-F). All three two-way interactions and

---

[4] Any concerns over the (numerically) larger error rates under AS compared to FT should be alleviated by Experiment 2, in which that numerical difference was reversed (though still not significant); and by the baseline data of Van 't Wout and Jarrold (2020), which similarly showed no significant difference in accuracy under AS and FT.

the three-way interaction were also significant: Distractor task type x complexity, F(2,94)=8.00, p=.001, η_p^2=.145 (H-F); distractor task type x stimulus occurrence, F(48,2256)=2.54, p<.001, η_p^2=.051 (H-F); complexity x stimulus occurrence, F(24,1128)=55.95 (H-F), p<.001, η_p^2=.543 (H-F); distractor task type x complexity x stimulus occurrence, F(48,2256)=2.42, p<.001, η_p^2=.049 (H-F).

To further examine the effects of AS and FT on performance in the light of these highly significant interactions, separate two-way repeated measures ANOVAs (distractor task type (AS or FT) x stimulus occurrence (1-25)) were run on the error data from the 3SR and 6SR conditions. In the 6SR condition, there were significant main effects of distractor task type, F(1,47)=11.84, p=.001, η_p^2=.201, and stimulus occurrence, F(24,1128)=135.34, p<.001, η_p^2=.742 (H-F), and a significant distractor task type x stimulus occurrence interaction, F(24,1128)=2.56, p=.010, η_p^2=.052 (H-F). In the 3SR condition, only the main effect of stimulus occurrence was significant, F(24,1128)=32.52, p<.001, η_p^2=.409 (H-F). The main effect of distractor task type and the interaction between distractor task type and stimulus occurrence were not significant, F(1,47)=2.13, p=.151, η_p^2=.043, and F(24,1128)=0.71, p=.789, η_p^2=.015 (H-F), respectively. As can be seen from Figure 2 (top right panel), the difference in accuracy between articulatory suppression and foot tapping increased and then decreased again in the 6SR condition, but not in the 3SR condition.

Given the absence of a significant effect of distractor task on error rates in the 3SR condition, Bayesian analyses were also conducted to assess the evidence for H1 (a difference between AS and FT) and H0 (no difference between AS and FT). This was done in order to follow up the significant interactions with distractor task, and to explore the evidence for or against the effect of distractor task in each condition separately. Two separate 2 (AS or FT) x 25 (stimulus occurrence) Bayesian ANOVAs were conducted (in JASP, using default priors) on the error rates for the 3SR and 6SR conditions. As we were predominantly interested in the evidence for or against the main effect of distractor task in each condition, we have reported only the Bayes Inclusion Factors (BF_{incl}=evidence in favour of H1 over H0 for models that contain the effect compared to models stripped of the effect) for the main effect of distractor task produced by each analysis. With regards to the main effect of distractor task, this analysis found extreme evidence for H1 in the 6SR condition ($BF_{incl}=6.357*10^{30}$), and moderate evidence for H0 in the 3SR condition ($BF_{incl}=0.271$).

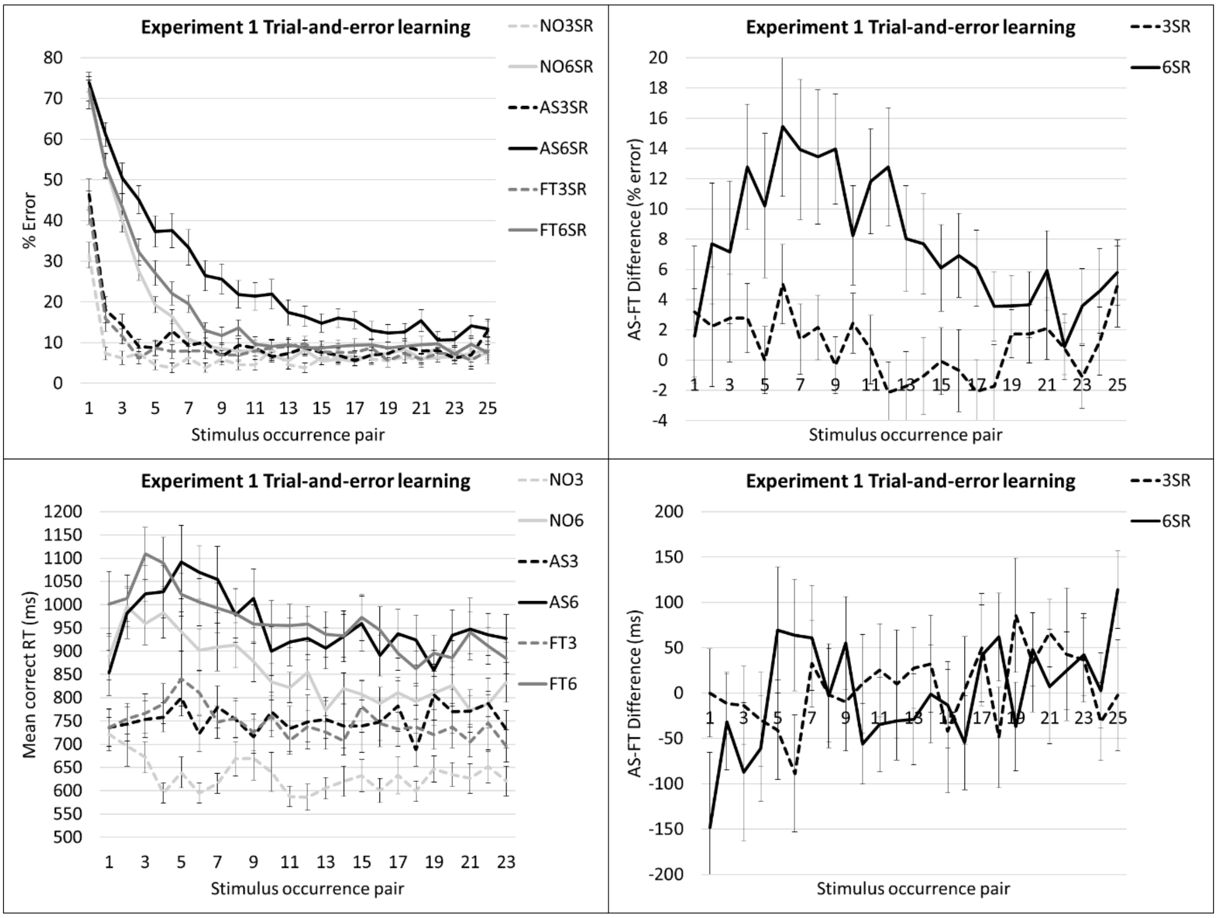

Figure 2. Mean % error (top) and mean correct RT (bottom) data (±SEM) for Experiment 1. On the left, data are plotted separately for the distractor task (AS, FT or None) and SR (3SR or 6SR) conditions, as a function of stimulus occurrence pair. On the right, the same data (AS and FT conditions only) are plotted as a difference score (AS-FT).

For the mean correct RTs, a 2 (distractor task type: AS or FT) x 2 (complexity: 3SR or 6SR) x 25 (stimulus occurrence pair: 1-25) repeated measures ANOVA was conducted. Note that 13 participants were excluded from this analysis due to empty cells[5]. This analysis revealed significant main effects for complexity and stimulus occurrence, F(1,34)=82.98, p<.001, η_p^2=.709 (H-F) and F(24,816)=3.76, p<.001, η_p^2=.100 (H-F), respectively. It also produced a significant stimulus occurrence by complexity interaction, F(24,816)=1.84, p=.023, η_p^2=.051 (H-F), reflecting that in the 6SR condition, RTs increased and then decreased as a function of stimulus occurrence; whereas in the 3SR condition this pattern was not observed to the same extent (see bottom left panel of Figure 2). Importantly, the main effect of distractor task type was not significant, F(1,34)=0.02, p=.889, η_p^2=.001, nor were the interactions with distractor task type (all F's below 1.21).

4. Experiment 1 summary

The results of Experiment 1 showed that the contribution of language to novel task learning is modulated by the complexity of the task. AS resulted in significantly more errors during the acquisition of a novel task than FT, but only when the task had six S-R rules, and not when it had three S-R rules. For the 6SR condition, the results looked very similar to those of Van 't Wout and Jarrold (2020) – error rates steadily increased with stimulus occurrence under AS compared to FT, and then decreased again. The absence of this pattern in the 3SR condition suggests that either participants do not use language to acquire a simpler task; or that they do, but that the effect is too small or short-lived to be detected. Experiment 2 directly examined this latter possibility by slightly increasing the number of S-R mappings from three to four in the less complex condition. However,

---

[5] The same analysis was also run including all participants. For that analysis, missing values (20 values across 13 participants) were interpolated (condition mean RT – (participant mean RT - group mean RT)). This analysis produced very similar results to the one excluding participants with missing values: Again the main effects of complexity and stimulus occurrence were significant, F(1,47)=117.38, p<.001, η_p^2=.714 and F(24,1128)=2.91, p<.001, η_p^2=.058 (H-F), respectively. The only (qualitative) difference was that the complexity by occurrence interaction was no longer significant, F(24,1128)=1.26, p=.214, η_p^2=.026 (H-F). Most importantly, as for the analysis excluding participants with missing cells, neither the main effect of distractor task nor any interactions with this factor were significant (all F's below 1.02).

the main aim of Experiment 2 was to compare the role of language in instruction-based learning versus trial-and-error learning, as described in the Introduction.

5. Experiment 2 Method

5.1 Participants

Forty-eight participants (aged between 18 and 44 [mean age= 23]; 34 female) took part in Experiment 2. As for Experiment 1, this number of participants was determined in advance, and data were not analysed prior to the completion of data analysis. Data from two participants (who did not perform the distractor tasks adequately when required) were removed and replaced. Participants provided informed consent prior to taking part, and all participants received £12 in return for their participation.

5.2 Procedure

Experiment 2 was very similar to Experiment 1, with the following modifications: In Experiment 2 participants learnt eight novel tasks, each consisting of either four or six S-R rules. Experiment 2 used four rather than three (as in Experiment 1) S-R rules per task in the less complex condition, to further investigate the possibility that the acquisition of a simpler task is not supported by language. Eight sets of six stimuli were selected from the IPNP in the same way as in Experiment 1 (see Table 2). The assignment of complete sets of stimuli (for the 6SR condition) or subsets of stimuli (for the 4SR condition) to distractor task condition was balanced between participants. Each participant completed half of the eight tasks under articulatory suppression, and half of the eight tasks under foot tapping[6]. For each distractor task condition, four tasks consistent of four S-R rules, and four tasks consisted of six S-R rules.

The main difference between Experiment 1 and Experiment 2 was that whereas in Experiment 1, all tasks were acquired by trial-and-error learning; in Experiment 2 half of the tasks were acquired by trial-and-error learning, and for the other half of the tasks participants received instructions on the correct response for each stimulus prior to the start of the task. Specifically, prior to each task participants were informed of the number of S-R rules in that task. In the instruction-based learning condition, participants subsequently received an instruction screen displaying the (four or six) stimuli in the task serially and simultaneously on screen, from left to right (mapping onto the x, c, v, b, n and m keys of a QWERTY keyboard). Participants were able to view these instructions for as long as they wanted, and they were not required to perform any distractor task for the duration of the instruction phase. The information provided to participants for the trial-and-error condition was identical to that provided in Experiment 1.

Each stimulus occurred 48 times in each novel task, resulting in 288 trials per task in the 6SR condition, and 192 trials per task in the 4SR condition. The order of SR condition, distractor task type (foot tapping or articulatory suppression) and manner of acquisition (instruction-based learning or trial-and-error learning) was balanced between participants, as was the assignment of stimuli to responses and conditions. Additionally, in the 4SR condition, half of the participants responded using the x, c, v and b keys, and half of the participants responded using the v, b, n and m keys (all using the index and middle finger of both hands).

---

[6] As the total number of tasks had to be increased in Experiment 2 to accommodate the new instruction method variable, the “no distractor task” control condition was omitted from the main experimental task in this experiment in order to simplify the design and analysis of this experiment.

As in Experiment 1, prior to the performing the main experimental task, all participants performed the same baseline colour matching task (24 trials each for the foot tapping, articulatory suppression and no distractor task condition). The experiment was conducted in a quiet space, in the presence of the experimenter, and lasted approximately 1 hour, after which participants were thanked and debriefed.

#	Set 1	RT	%	#	Set 4	RT	%	#	Set 7	RT	%
1	egg	874	98	13	heart	720	100	31	door	719	100
2	car	751	100	14	Owl	837	98	32	broom	821	100
3	tree	796	100	15	foot	758	98	33	saw	863	100
4	fan	865	98	16	moon	804	100	34	nose	721	100
5	sock	712	100	17	key	738	100	35	bell	703	100
6	hat	684	98	18	bread	773	98	36	flag	847	100
	Mean	**807**	**99**		**Mean**	**772**	**99**		**Mean**	**801**	**100**

#	Set 2	RT	%	#	Set 5	RT	%	#	Set 8	RT	%
7	spoon	777	100	19	frog	751	100	37	horse	809	100
8	tent	744	100	20	chair	732	100	38	comb	717	100
9	box	753	100	21	hand	723	98	39	sun	762	100
10	pig	855	100	22	train	838	100	40	ring	785	100
11	ear	681	100	23	snake	775	100	41	book	656	100
12	watch	780	100	24	kite	796	100	42	glove	848	100
	Mean	**758**	**100**		**Mean**	**735**	**99**		**Mean**	**763**	**100**

#	Set 3	RT	%	#	Set 6	RT	%
13	bus	777	100	25	bed	706	100
14	leaf	744	100	26	fish	777	100
15	pen	753	100	27	cheese	843	100
16	house	855	100	28	clock	772	98
17	dog	681	100	29	dress	840	100
18	cake	780	100	30	eye	700	98
	Mean	**758**	**100**		**Mean**	**775**	**100**

Table 2. Picture names for the stimulus sets used in Experiment 2. For the 4 S-R condition, subsets of four stimuli were selected from each set. Stimuli within a set or subset were matched for percent name agreement (%) and RT target mean (in ms; mean latency for dominant responses only).

6. Experiment 2 Results

As for Experiment 1, reaction times greater than 5000ms (0.3% of correct responses) were removed from the data set prior to conducting the analyses described below. Throughout, the number following the ± symbol indicates the standard error of the mean (SEM).

6.1 Baseline task

As for Experiment 1, a one-way repeated measures ANOVA (distractor task type: AS, FT or none) was run on the mean correct RT and % error data from the baseline task. The results of Experiment 2 replicated those of Experiment 1. Specifically, for the % error analysis, a significant main effect of distractor task type, F(2,94)=6.78, p=.002, η_p^2=.126 (H-F), indicated that error rates were greater under AS (4.2±0.7%) than in the no distractor task condition (2.0±0.5%), F(1,47)=8.34, p=.006, η_p^2=.151, and under FT (4.6±0.8%) than in the no distractor task condition, F(1,47)=14.50, p<.001, η_p^2=.236; but the difference between AS and FT was not significant, F(1,47)=0.27, p=.604, η_p^2=.006.

With regards to the mean correct RT analysis, a significant main effect of distractor task type, F(2,94)=10.07, p<.001, η_p^2=.176 (H-F), reflected significantly greater RTs under FT (607±22ms) compared to both AS (560±20ms), F(1,47)=10.38, p=.002, η_p^2=.181, and the no distractor task condition (543±16ms), F(1,47)=17.14, p<.001, η_p^2=.267. As for Experiment 1, the difference between AS and the no distractor task condition was not significant, F(1,47)=1.46, p=.233, η_p^2=.030.

6.2 Experimental task

A 2 (manner of acquisition: trial-and-error or instruction-based learning) x 2 (distractor task type: AS or FT) x 2 (task complexity: 4SR or 6SR) x 24 (stimulus occurrence pair: 1 to 24) repeated measures ANOVA was run on the mean correct RT and the % error data.

For the error data, this analysis yielded significant main effects of manner of acquisition, F(1,47)=115.51, p<.001, η_p^2=.711, distractor task type, F(1,47)=7.19, p=.010, η_p^2=.133, complexity, F(1,47)=33.91, p<.001, η_p^2=.419, and stimulus occurrence, F(23,1081)=125.02, p<.001, η_p^2=.727 (H-F). It also produced significant two-way interactions between manner of acquisition and distractor task type, F(1,47)=5.23, p=.027, η_p^2=.100; manner of acquisition and task complexity, F(1,47)=30.61, p<.001, η_p^2=.394; distractor task type and task complexity, F(1,47)=6.42, p=.015, η_p^2=.120, manner of acquisition and stimulus occurrence, F(23,1081)=127.62, p<.001, η_p^2=.731 (H-F), distractor task and stimulus occurrence, F(23,1081)=1.92, p=.026, η_p^2=.039 (H-F); and task complexity and stimulus occurrence, F(23,1081)=15.35, p<.001, η_p^2=.246 (H-F). The analysis also produced two significant three-way interactions between manner of acquisition, distractor task and stimulus occurrence, F(23,1081)=3.05, p<.001, η_p^2=.061 (H-F); and between manner of acquisition, task complexity and stimulus occurrence F(23,1081)=11.79, p<.001, η_p^2=.201 (H-F).

To further explore how the role of language in novel task learning is modulated by task complexity and manner of acquisition, and given the significant three-way interactions reported above, four separate 2 (distractor task type: AS or FT) x 24 (stimulus occurrence pair: 1 to 24) repeated measures ANOVA were run (4SR trial-and-error learning, 6SR trial-and-error learning, 4SR instruction-based learning and 6SR instruction-based learning). These ANOVAs revealed a significant interaction between distractor task type and stimulus occurrence only when participants learnt the more complex 6SR task by trial-and-error, F(23,1081)=2.71, p=.005, η_p^2=.055 (H-F). The same interaction was not significant for the 4SR trial-and-error condition, F(23,1081)=1.55, p=.105, η_p^2=.032 (H-F), the 4SR instruction-based learning condition, F(23,1081)=1.06, p=.386, η_p^2=.022 (H-F), or the 6SR instruction-based learning condition, F(23,1081)=1.667, p=.059, η_p^2=.034 (H-F). As can be seen from Figure 3, data from the more complex 6SR trial-and-error condition mirrored the data from Experiment 1: There was a significant detrimental effect of AS, which was modulated by stimulus occurrence. This pattern of results was not significant in the other three conditions.

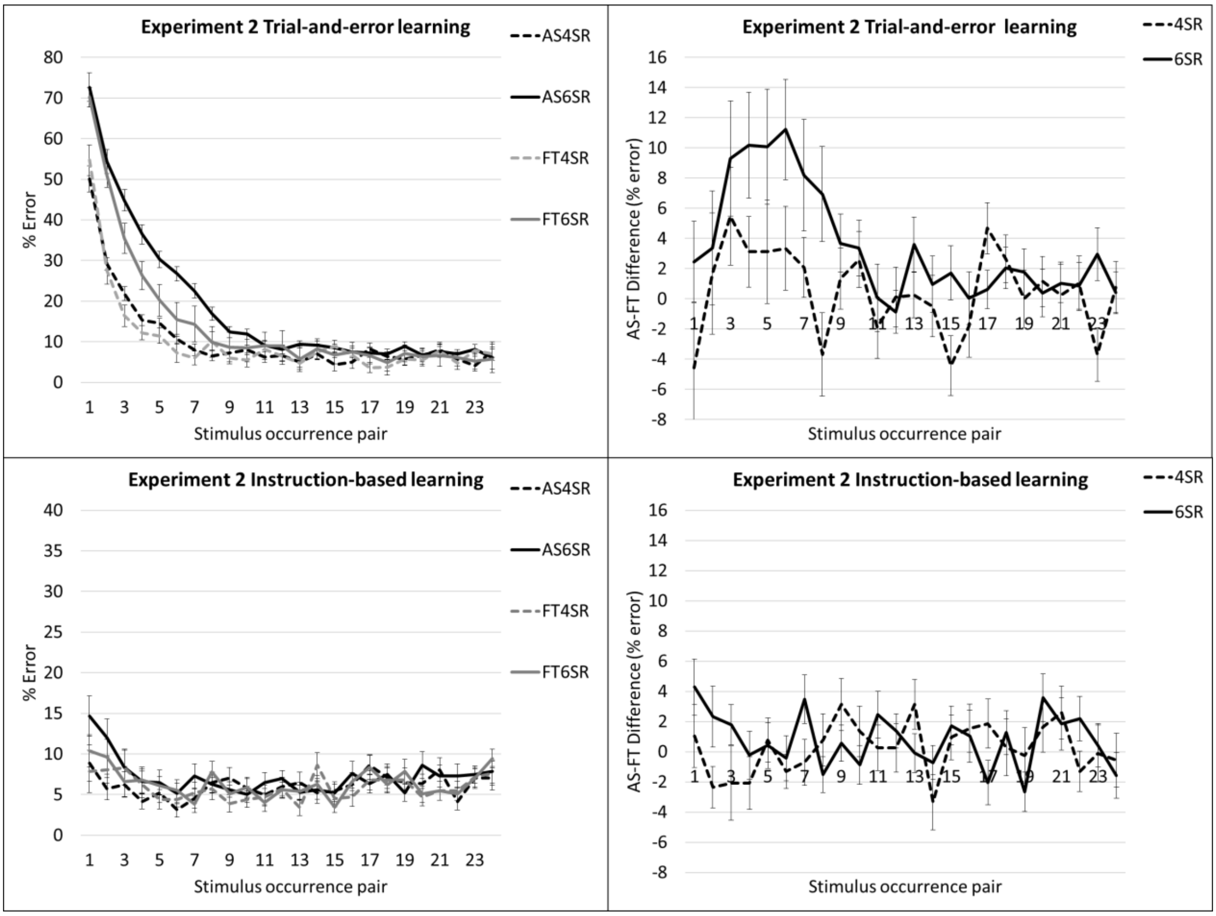

Figure 3. Mean % error data (±SEM) for Experiment 2 in the trial-and-error condition (top) and the instruction-based learning condition (bottom). On the left, data are plotted separately for the distractor task (AS or FT) and SR (4SR or 6SR) conditions, as a function of stimulus occurrence pair. On the right, the same data are plotted as a difference score (AS-FT).

Hence, there appears to be no detectable effect of AS on performance in instruction-based learning. However, given that instruction-based learning is likely to be fast, examining the effect of AS across a large number of trials may not be informative in that condition[7]. Indeed, it is possible that the role of language in instruction-based learning is so short-lived, that the effect of AS is restricted to the first (few) occurrence(s) of each stimulus. To examine this possibility, one-way repeated measures ANOVAs compared error rates under AS and FT for the first stimulus occurrence pair in each of the four conditions separately. Consistent with the prediction that the detrimental effect of AS is extremely short-lived in the instruction-based learning condition, error rates were significantly higher under AS compared to FT for the first stimulus occurrence pair only in the 6SR instruction-based learning condition, F(1,47)=5.32, p=.026, η_p^2=.102 (difference of 4.3±1.9%). The same difference was not significant in the 4SR instruction-based learning condition, F(1,47)=0.25, p=.619, η_p^2=.005 (difference of 1.0±2.1%), the 4SR trial-and-error condition, F(1,47)=1.13, p=.294, η_p^2=.023 (difference of -4.6±4.3%), or the 6SR trial-and-error condition, F(1,47)=0.84, p=.365, η_p^2=.018 (difference of 2.5±2.7%). This confirms that the effect of AS is extremely short-lived (with 6 S-R rules) or absent (with 4 S-R rules) in instruction-based learning. These findings were in contrast to the 6SR trial-and-error condition, in which the effect of AS is absent for the first stimulus occurrence pair, after which it gradually increases, and then decreases again.

As for Experiment 1, separate 2 (AS or FT) x 24 (stimulus occurrence) Bayesian ANOVAs were conducted on the accuracy data in each condition (4SR trial-and-error learning, 6SR trial-and-error learning, 4SR instruction-based learning and 6SR instruction-based learning), in order to follow up the significant interactions with distractor task. As for Experiment 1, below we have reported the Bayes Inclusion Factors (BF_{incl}=evidence in favour of H1 over H0 for models that contain the effect compared to models stripped of the effect) for the main effect of distractor task for each of the four conditions. With regards to the main effect of distractor task, these analyses yielded extreme evidence for H1 in the 6SR trial-and-error condition (BF_{incl}=1.564*10^8), and very strong evidence for

---

[7] The rationale for including a large number of trials in the instruction-based condition was to facilitate the (statistical) comparison between the trial-and-error and instruction-based learning conditions.

H0 in the 4SR trial-and-error condition (BF_{incl}=0.087). For the instruction-based condition, this analysis found strong evidence for H0 in the 4SR condition (BF_{incl}=0.056), and weak evidence for H0 in the 6SR condition (BF_{incl}=0.642). This latter finding is not surprising, given that there was indeed a significant but short-lived effect of distractor task in this condition (for the first two stimulus occurrences, see above).

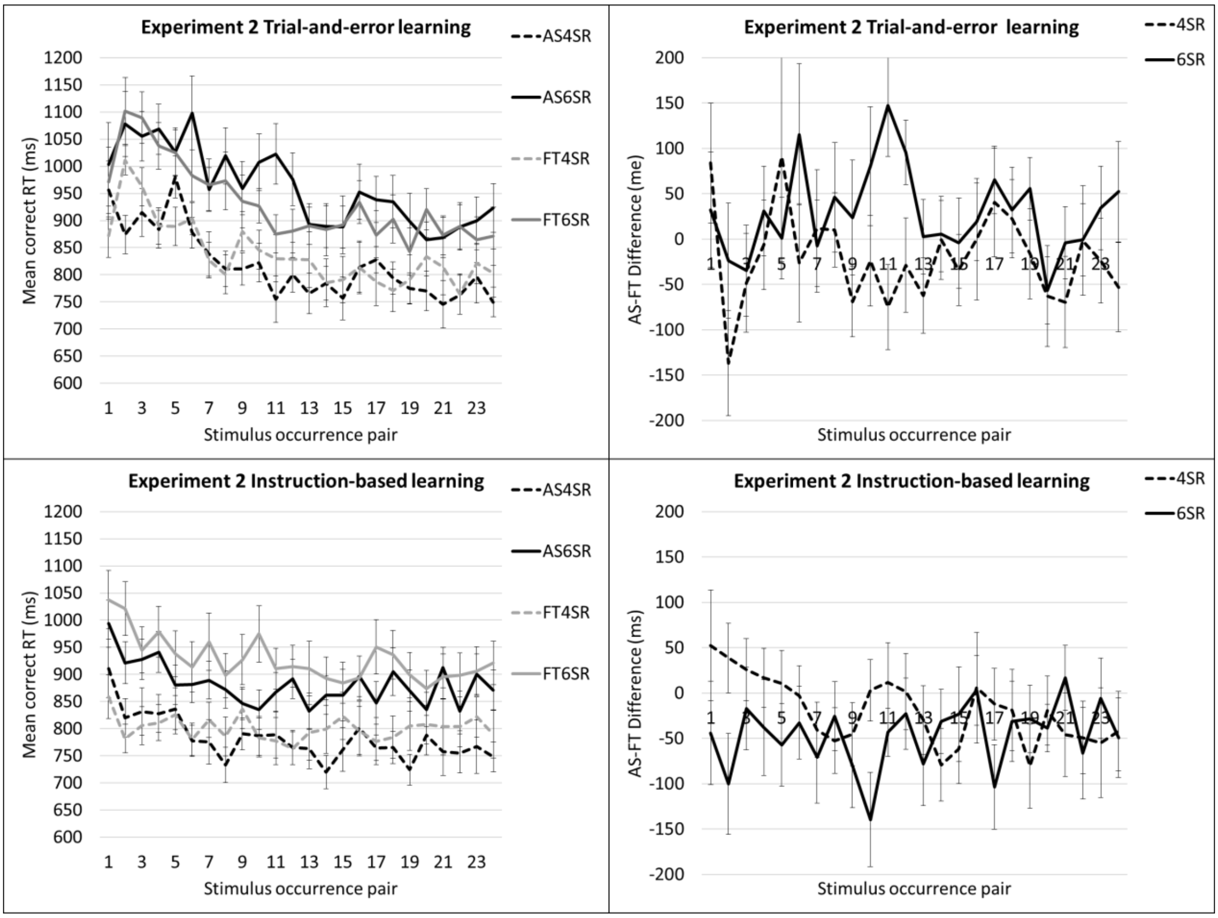

Figure 4. Mean correct RT data (±SEM) for Experiment 2 in the trial-and-error condition (top) and the instruction-based learning condition (bottom). On the left, data are plotted separately for the distractor task (AS or FT) and SR (4SR or 6SR) conditions, as a function of stimulus occurrence pair. On the right, the same data are plotted as a difference score (AS-FT).

The same 2 x 2 x 2 x 24 repeated measures ANOVA was run on the mean correct RT data (see Figure 4). Note that for this analysis, nine participants were removed due to empty cells[8]. This analysis

---

[8] As for Experiment 1, the same analysis was also run including all participants. For that analysis, missing values (10 values across 9 participants) were interpolated (condition mean RT – (participant mean RT - group mean

produced significant main effects of manner of acquisition, F(1,38)=6.51, p=.015, η_p^2=.146, complexity, F(1,38)=103.71, p<.001, η_p^2=.732, and stimulus occurrence, F(23,874)=11.17, p<.001, η_p^2=.227 (H-F). These main effects indicate that overall, responses were significantly slower in the trial-and-error condition (compared to instruction-based learning), and slower with six S-R rules (compared to 4 S-R rules); and that RTs decreased as a function of stimulus occurrence (see Figure 4). A significant manner of acquisition by occurrence interaction, F(23,874)=4.07, p<.001, η_p^2=.097 (H-F), reflected the fact that RT decreased more steeply as a function of practice under trial-and-error learning. A further marginally significant two-way interaction (manner of acquisition x distractor task, F(1,38)=3.41, p=.073, η_p^2=.082) and three-way interaction (manner of acquisition x distractor task x complexity, F(1,38)=7.35, p=.010, η_p^2=.162) reflected the fact that RTs were somewhat faster under AS than FT in instruction based learning (AS-FT difference of -20ms in 4SR; -46ms in 6SR); but less so in trial-and-error learning, where the opposite was true only in the 6SR condition (AS-FT difference of 30ms; -20ms in 4SR). For all the other main effects and interactions, F<1.16.

6.3 Experiment 2 summary

Experiment 2 investigated whether the role of language in novel task learning is modulated by task complexity, and by the manner of acquisition (trial-and-error learning or instruction-based learning). For the trial-and error condition, Experiment 2 confirmed the results of Experiment 1: a detrimental effect of AS on performance was found only when the task was more complex (six S-R rules per

---

RT)). This analysis produced very similar results to the one excluding participants with missing values, except that the manner of acquisition by distractor task interaction was now significant, F(1,47)=4.14, p=.048, η_p^2=.081, rather than marginally significant. The same main effects and interactions (as for the analysis excluding participants with missing cells) remained significant: Manner of acquisition main effect; F(1,47)=9.16, p=.004, η_p^2=.163; complexity main effect, F(1,47)=125.70, p<.001, η_p^2=.728; stimulus occurrence main effect, F(23,1081)=12.58, p<.001, η_p^2=.211 (H-F); manner of acquisition x occurrence interaction, F(23,1081)=4.51, p<.001, η_p^2=.087 (H-F); manner of acquisition x distractor task x complexity, F(1,47)=6.15, p=.010, η_p^2=.116). For all the other main effects and interactions, F<1.54.

task), and not when it was less complex (three or four S-R rules per task in Experiments 1 and 2, respectively).

With regards to the manner of acquisition, Experiment 2 found that the effect of articulatory suppression on novel task learning differed significantly between trial-and-error learning and instruction-based learning. When learning a more complex task by trial-and-error, the detrimental effect of AS on accuracy increased, and then decreased again with practice. This pattern of results replicates the findings of Experiment 1, and those of Van ‘t Wout and Jarrold (2020). On the contrary, when learning a novel task by instructions, the effect of articulatory suppression on accuracy was much more short-lived (with six S-R rules) or absent (with four S-R rules).

Finally, although the observation that RTs were faster under AS (compared to FT) in instruction-based learning might appear counterintuitive, this could imply participants are less cautious under articulatory suppression in instruction-based learning, because the S-R rules are not as well established. If the encoding or retrieval of an S-R rule during performance is reliant on language, and AS interferes with this, then participants may sometimes be forced to guess the answer, which would lead to faster RTs than retrieving the correct SR rule. One observation which potentially argues against this explanation is that the RT difference in the AS and FT conditions appears to persist for longer than the difference in the AS and FT conditions in the error rates. Hence, further research is needed to fully understand why participants may sometimes respond faster in the AS condition. Either way, these RT findings do not negate the results of the accuracy analysis, as there were no significant interactions involving both practice (stimulus occurrence pairs) and distractor task in the RT analysis.

6.4 Additional analyses

In order to further elucidate the process by which a novel task is acquired, two additional analyses were conducted: A serial order analysis and a post-error analysis. The former analysis has the potential to reveal *how* participants acquire a novel task; specifically, whether they compile a mental representation of a task in a serial manner, from the leftmost response to the rightmost response. The latter post-error analysis asked what happens after participants make an error during the initial stages of learning. Specifically, it explores the counterintuitive prediction that participants may be slower following a *correct* response in the trial-and-error condition, as participants would use this additional time to update their mental representation of the task-set. Note that there were not enough data to conduct these analyses for each distractor task condition separately, and so the data were pooled across distractor task conditions.

6.4.1 Serial order analysis

To examine whether there were any serial acquisition effects, data from both experiments were analysed as a function of response finger (from left to right). Data were analysed separately for each SR condition; and as a function of experiment quarter (in order to examine whether any effects of serial position are modulated by practice). As the main effects of interest in both experiments were manifested in the accuracy data rather than the RT data, the serial order analysis was restricted to the accuracy data.

For Experiment 1, two quarter (4) by response finger (3 or 6) repeated measures ANOVAs were run separately for the 3SR and 6SR conditions. For the 6SR condition, this ANOVA found significant main effects of quarter, F(3,141)=345.33, p<.001, η_p^2=.880 (H-F), response finger, F(5,235)=8.360, p<.001,

η_p^2=.151 (H-F), and a quarter by response finger interaction, F(15,705)=4.190, p<.001, η_p^2=.082 (H-F). Four separate one-way ANOVAs (with response finger as within-subjects factor) for each quarter found that, in the first quarter, error rates increased linearly from left to right, F(1,47)=12.67, p=.001, η_p^2=.212; whereas the effect of response finger was marked by a quadratic trend in the second, F(1,47)=14.10, p=.001, η_p^2=.231, third, F(1,47)=45.48, p<.001, η_p^2=.492, and fourth quarter, F(1,47)=20.49, p<.001, η_p^2=.304 (see Figure 5).

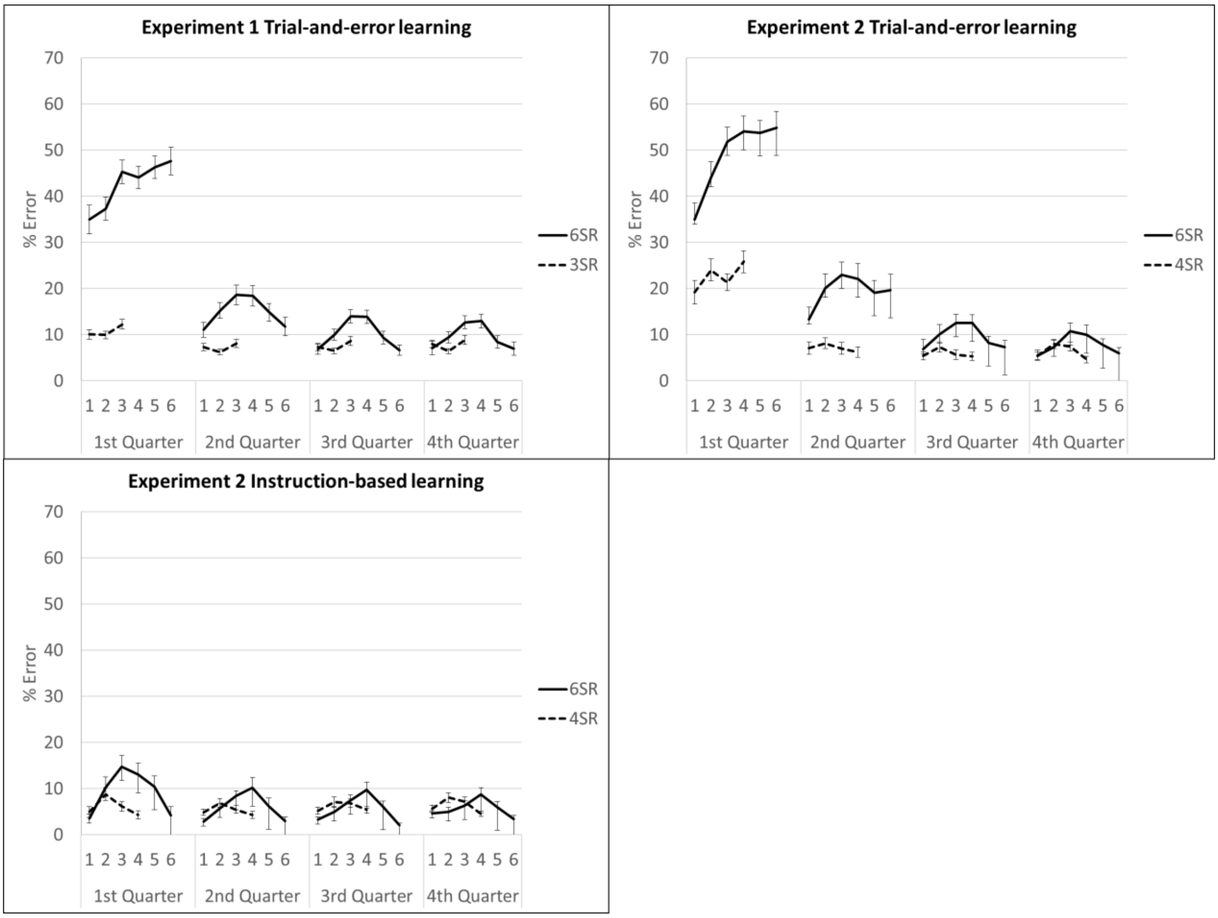

Figure 5. Mean % error data (±SEM) plotted as a function of SR condition, response finger, and practice (experiment quarter).

For the 3SR condition, there was a significant main effect of quarter, F(3,141)=32.44, p<.001, η_p^2=.408 (H-F), and of response finger, F(2,94)=6.29, p=.003, η_p^2=.118 (H-F), but no significant quarter by response finger interaction, F<1. As can be seen from Figure 5, error rates decreased with each quarter; and they were lower for the middle response finger compared to the outer two (quadratic trend of response, F(1,47)=9.17, p=.004, η_p^2=.163).

For Experiment 2, separate quarter (4) by response finger (4 or 6) repeated measures ANOVAs were run for the 4SR and 6SR trial-and-error and instruction-based learning conditions. Again, data were pooled across distractor task type, and only the first 192 trials of each task were analysed (this was done to make the analysis more comparable to Experiment 1; and Experiment 1 showed little change from the third to the fourth quarter, so culling these data was considered appropriate).

For the 6SR trial-and-error condition, this ANOVA replicated the results of Experiment 1: Errors increased linearly with response finger only in the first quarter, F(1,47)=12.99, p=.001, η_p^2=.216, and a quadratic trend was present in the second, F(1,47)=8.94, p=.004, η_p^2=.160, third, F(1,47)=13.65, p=.001, η_p^2=.225, and fourth quarter, F(1,47)=10.31, p=.002, η_p^2=.180 (two-way interaction, F(15,705)=4.26, p<.001, η_p^2=.083 (H-F)).

For the 4SR trial-and-error condition, the first quarter was characterized by a significant linear trend, F(1,47)=5.80, p=.020, η_p^2=.110, and the final quarter by a significant quadratic trend, F(1,47)=12.43, p=.001, η_p^2=.209 (two-way interaction, F(9,423)=3.15, p=.004, η_p^2=.063, H-F). No significant trends for response finger were observed in the middle two quarters (F's < 1.07).

For the 6SR instruction-based learning condition, all four quarters were marked by a significant quadratic trend only (in order: F(1,47)=40.83, p<.001, η_p^2=.465, F(1,47)=23.74, p<.001, η_p^2=.336, F(1,47)=22.89, p<.001, η_p^2=.328, F(1,47)=18.44, p<.001, η_p^2=.282). A significant two-way interaction between quarter and response (F(15,705)=2.24, p=.008, η_p^2=.046; H-F) likely indicates a more marked quadratic trend in the first quarter compared to the other three quarters. For the 4SR instruction-based learning condition, only the main effect of response was significant, reflecting a significant quadratic trend for response finger across all quarters, F(1,47)=21.12, p<.001, η_p^2=.310.

In summary, the serial order analysis showed that errors increased linearly with response finger in the first quarter, but only in the trial-and-error condition, and only if the task was sufficiently complex (4SR and 6SR conditions). In the instruction-based learning condition the pattern of results look quite different, with the quadratic function already emerging from the first quarter. The linear trend under trial-and-error learning suggests that participants are compiling (or accessing) the task-set in a serial manner, from left to right. The quadratic function which is seen in the remaining three quarters, and from the beginning under instruction-based learning, potentially suggests improved discriminability for the outer response fingers (Kent & Lamberts, 2005)

6.4.2 Post-error analysis

One common finding within the cognitive psychology literature is that participants are typically slower following an error (Rabbitt, 1966). However, there is reason to believe this may not happen when participants are learning by trial-and-error: In that case, RTs may be expected to be slower following a correct response right at the beginning of each task, as participants take time to update the representation of the task-set. To examine this possibility we analysed reaction times as a function of the previous trial accuracy (correct or incorrect), and as a function of SR condition. As such post-correct slowing would be most likely to present at the beginning of each task (when the task is not yet well-practiced), only the first 6 occurrences of each stimulus[9] were included in this analysis (any fewer would result in the exclusion of too many participants due to empty cells); however for completeness the data are plotted as a function of stimulus occurrence pair in Figure 6.

---

[9] As this is a sequential analysis, the first trial of each task was excluded from this analysis.

For Experiment 1,a 2 (3SR or 6SR) x 2 (previous correct or incorrect) repeated measures ANOVA found significant main effects of complexity, F(1,47)=136.39, p<.001, η_p^2=.744, and previous trial accuracy, F(1,47)=5.23, p=.027, η_p^2=.100; and a significant two-way interaction, F(1,47)=33.78, p<.001, η_p^2=.418. Further one-way ANOVAs (previous correct or incorrect) for each SR condition separately showed that in the 6SR condition, RTs were 161±29ms slower following a correct response, F(1,47)=31.29, p<.001, η_p^2=.400. For the 3SR condition, participants were marginally (59±30ms) slower following an error, F(1,47)=3.99, p=.052, η_p^2=.078.

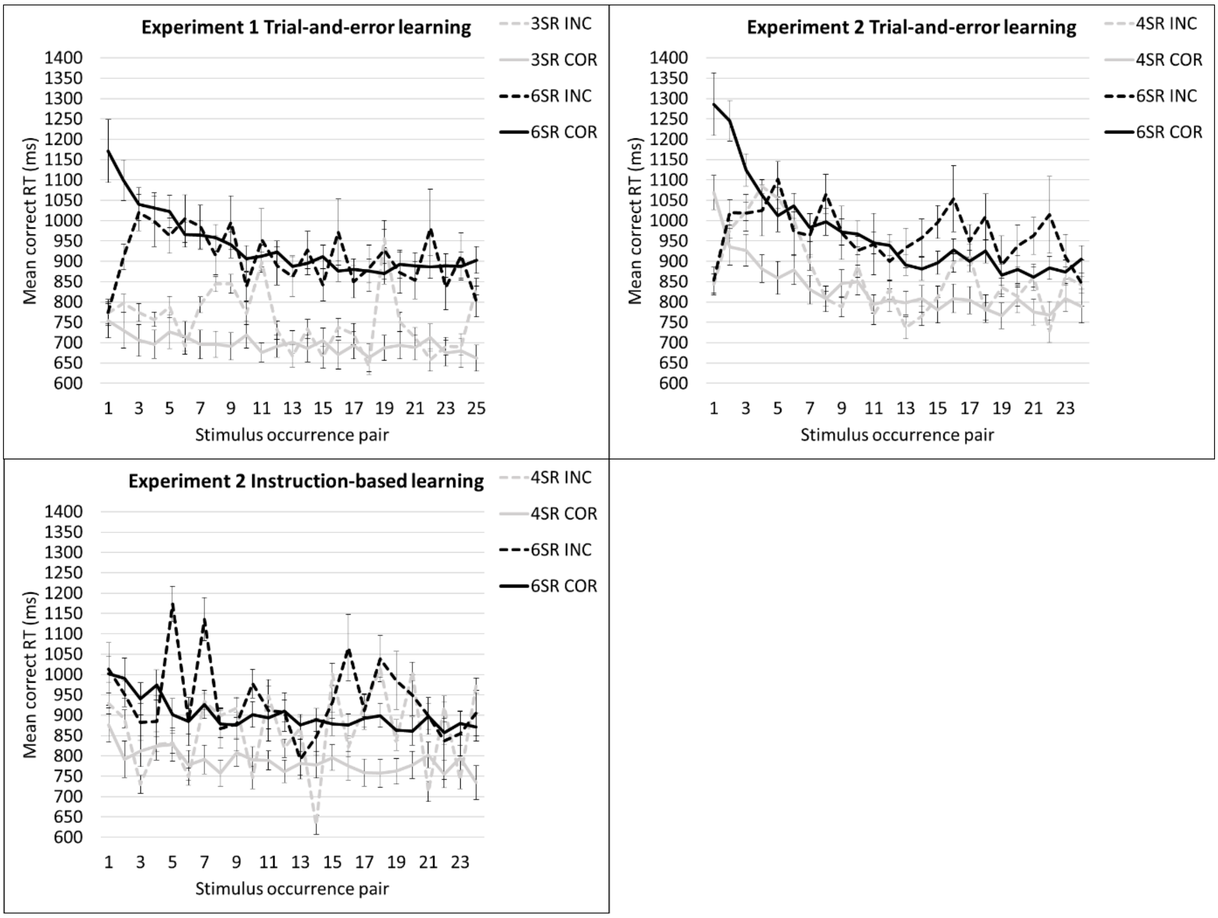

Figure 6. Mean correct RT data (±SEM) as a function of stimulus occurrence, plotted separately as a function of previous trial accuracy (correct or incorrect) and SR condition (3, 4 or 6 S-R rules).

The same 2 (4SR or 6SR) x 2 (previous correct or incorrect) repeated measures ANOVA was run separately for the trial-and-error and instruction-based learning conditions of Experiment 2. For the trial-and-error learning data, this analysis replicated the results of Experiment 1: There was again a significant two- way interaction between complexity and previous trial accuracy, F(1,47)=13.34, p=.001, η_p^2=.221. Further one-way ANOVAs (previous correct or incorrect) for each SR condition separately found that participants were slower following a correct response, but only in the 6SR condition (difference of 220±46ms; F(1,47)=22.98, p<.001, η_p^2=.328), and not significantly so in the 4SR condition (16±34ms; F(1,47)=0.23, p=.634, η_p^2=.005).

In the instruction-based learning condition, however, no post-correct slowing was observed (note that for this analysis, twelve participants were excluded due to empty cells). For this analysis, the main effect of complexity approached significance, F(1,35)=3.32, p=.077, η_p^2=.087. Additionally, RTs were numerically *faster* following a correct response in the 4SR (115±90ms) but less so in the 6SR condition (5±40ms), though neither the main effect of previous trial accuracy, F(1,36)=1.43, p=.249, η_p^2=.039, nor the interaction, F(1,36)=1.26, p=.269, η_p^2=.035, were significant. In summary, there was evidence of post-correct slowing in both experiments, but only when the task was acquired by trial-and-error, and only when it was sufficiently complex.

7. Discussion

The two experiments reported here sought to elucidate the role of language in acquiring simple cognitive tasks. Previous research (Van ‘t Wout & Jarrold, 2020) found that when participants acquire novel tasks by trial-and-error, more errors were made under articulatory suppression (AS; which disrupts the use of language) compared to foot tapping (FT; a non-verbal distractor task). Moreover, this effect of articulatory suppression was restricted to the first half of each novel task;

suggesting task performance only relies on language when a task is new, and not when it is well-practiced.

The experiments reported here were designed to investigate which factors influence the contribution of language to novel task learning, focussing specifically on task complexity (defined by the number of S-R rules), and on the manner of acquisition (trial-and-error versus instruction-based learning). Experiment 1 manipulated the complexity of the task, so that each task consisted of either three or six S-R rules. Experiment 1 showed that AS resulted in more errors during the early stages of learning, suggesting that language plays a crucial role in the acquisition of novel tasks, as it is used to establish and maintain novel S-R rules in verbal working memory during the beginning of each task. Importantly, this detrimental effect of AS was found only when the task is more complex (six S-R rules per task). When the task was less complex (three S-R rules per task) participants were no worse under articulatory suppression compared to foot tapping. Experiment 2 also manipulated the complexity of the task (four versus six S-R rules), but increased the number of S-R rules in the less complex condition from three to four. The results were very similar to those of Experiment 1: again articulatory suppression resulted in significantly more errors only when the task was more complex, and not when it was less complex.

There are two possible explanations for this pattern of results: Either language does not play an important role in the acquisition of less complex tasks by trial-and-error; or the role of language is so short-lived that the current study failed to detect it. One observation from Experiment 2 argues against this latter explanation. Specifically, Experiment 2 also compared the effects of AS and FT on performance for the first stimulus occurrence pairs, and found no significant difference between AS and FT even for the first two stimulus occurrences under trial-and-error learning (interestingly, the results were different for instruction-based learning, as described below). This result suggests that for complex tasks, language may be used to establish and maintain S-R rules in verbal working memory, whereas for tasks of simpler description, proceduralisation might not depend on language. Instead, it is possible that in the less complex (3SR and 4SR) conditions, participants used a visual strategy when articulatory suppression prohibited the use of language. One way in which language may aid the proceduralisation of S-R rules is through the use of rehearsal. As is well known, when rehearsal is available as a strategy, short-term memory typically holds around seven items (e.g., Miller, 1956). However, the capacity of visual short-term memory is thought to be restricted to three chunks (Zhang & Simon, 1985) or four objects (Luck & Vogel, 1997). This difference in verbal and visual short-term memory capacity could explain why such a visual strategy might only be effective in the less complex conditions (evidenced by the absence of a detrimental effect of AS), and not in the more complex conditions.

The comparison between tasks of greater and lesser complexity can also provide further insights into to precise role of language in novel task learning. With regards to the role of language, there are several possibilities: Firstly, it is possible that participants are using language to label the stimuli in a task in relation to their position within the sequence of responses (consistent with the results of Van ‘t Wout et al., (2013) and Monsell and Graham (2021) who found that performance was affected by the phonological similarity of the stimulus names). Another possibility is that language specifically supports the binding of action-effect associations (e.g., see Kray et al., 2006). Although the results of the current study do not unanimously distinguish between these two possibilities, they do rule out one possible explanation for the role of language in learning: Namely, the idea that an effective verbal representation of the task-set is solely dependent on having labelled each stimulus a fixed number of times. If this were the case, then the current study should have found no difference between the effect of AS on the less and more complex conditions, because in both experiments the data were plotted as a function of stimulus occurrence (rather than trial number). This analysis allowed us to determine whether the role of language in learning changes as a function of how many times each individual stimulus was encountered. For example, for the 2nd stimulus occurrence pair, each stimulus has been encountered 4 times regardless of the complexity of the task. If stimulus frequency – and specifically the frequency with which each stimulus was labelled – was a determining factor in the role of language in task learning, then there should be no difference between the 3 (or 4) and 6SR conditions when the data are plotted as a function of stimulus occurrence. This is not the case, suggesting that it is the overall complexity of the task-set (or perhaps the process of integrating different task elements into a coherent task representation) which modulates the role of language, not the number of times each individual stimulus is encountered (and labelled).

In addition to investigating the role of task complexity, this study also sought to investigate how the role of language in novel task learning might be modulated by the manner in which the task is acquired. To this end, Experiment 2 contrasted a trial-and-error condition (employed in Experiment 1, and in Van ‘t Wout & Jarrold, 2020) with another condition in which participants received instructions on the correct S-R mappings prior to the start of each task.

In Experiment 2, the difference in results for the instruction-based learning and trial-and-error conditions was striking. In the trial-and-error condition with six S-R rules, the pattern directly replicated Experiment 1, and Van ‘t Wout & Jarrold (2020): The detrimental effect of AS on accuracy was initially absent (suggesting that participants do not yet have a linguistic representation of the task rules to begin with), after which it increased, and then decreased again. This inverted U-shape in the data indicated that as participants are creating a verbal representation of the task-set, the detrimental effect of AS increases; and then as the role of language ebbs away, the effect of AS decreases gradually (consistent with theories of skill acquisition, Anderson, 1982). This U-shaped pattern of results was completely absent in the instruction-based conditions: Instead, in the 6SR instruction-based learning condition AS was found to have a detrimental effect on accuracy (compared to FT) for the first two stimulus occurrences only. By contrast, no significant difference between FT and AS for first stimulus occurrence pair was found in the other three conditions (4SR instruction-based learning and 4SR and 6SR trial-and-error learning). These results clearly demonstrate that the effect of AS on performance in instruction-based learning is extremely short-lived; and moreover that it is dependent on the complexity of the task. It is highly likely that participants are able to achieve near-effective proceduralisation during the instruction-phase of the experiment ("near" because an effect of AS remained for the first two stimulus occurrences in the more complex condition). Although it is possible that participants are compiling an effective representation of the task during the instruction-phase without the use of language, this seems unlikely given that an effect of AS was still obtained for the first two stimulus occurrences. Hence, there is strong evidence that when learning via instructions a verbal representation of the task was established during the instruction phase.

Further evidence for the idea that a linguistic representation of the S-R rules can be established during the instruction phase was provided very recently by Monsell and Graham (2021), who manipulated the phonological similarity of the object names in a choice RT task. They found that the detrimental effect of phonological similarity on performance was much more short-lived when participants acquired each task via instructions (Experiments 1 & 2), compared to a trial-and-error learning condition (Experiment 3). Consistent with our interpretation, Monsell and Graham (2021) conclude that for when learning via instructions, "…. considerable proceduralisation must have been accomplished during the instruction phase" (p. 11). The current study further adds to that observation, by demonstrating that this process of proceduralisation is affected by the complexity of the task.

In addition to the two main analyses described above, which found that the role of language in novel task learning is modulated by both task complexity and the manner of acquisition, two further analyses were conducted across both experiments to shed light on the cognitive mechanisms that support the acquisition of novel tasks. The first of these two analyses asked whether participants acquire new tasks in a serial manner (from left to right) with regards to the representation of S-R rules. The second set of analyses asked how post-error performance differs in trial-and-error learning compared to instruction-based learning.

With regards to the first set of analyses, data from the current study were analysed as a function of response finger to investigate whether participants acquire new tasks in a serial manner (from left to right). Some evidence from the task switching literature has found support for such serial representation for well-practiced tasks (e.g., Lien, Ruthruff, Remington & Johnston, 2005), although other experiments with a similar structure have failed to replicate this pattern of results (e.g., Monsell & Mizon, 2006; Lindsen & De Jong, 2010). Moreover, to date no studies have investigated how the serial representation of S-R rules might be modulated by practice.

Both of the current experiments showed evidence of serial representation of S-R rules at the beginning of each task – namely, a linear increase in error rates with response finger from left to right (i.e., fewer errors for the left most response fingers), but only in the trial-and-error condition, and not when the task was acquired by instructions[10]. The effect was furthermore modulated by the complexity of the task, as this linear trend was only observed when the task was more complex (four or six S-R rules), and not when the task had only three S-R rules. For the more complex trial-and-error conditions, the linear trend was only observed in the first quarter: for the remainder of each task a quadratic function was found, suggesting better encoding and/or recall for the leftmost and rightmost S-R rules (akin to serial position effects in free recall, Murdock, 1962); or increased discriminability for the outer response fingers (Kent & Lamberts, 2005). The most interesting finding

---

[10] Note that this latter finding contrasts with the result of a serial position analysis reported by Monsell and Graham (2021), who did find a linear increase in error rates with response finger at the beginning of each task for instruction-based learning. This difference in results could be explained by the fact that Monsell and Graham (2021) observed this linear trend for the first four encounters with each stimulus, whereas in the current study the first quarter included the first eight encounters with each stimulus.

is the linear trend at the start of learning, as it suggests participants can sometimes acquire novel tasks in a serial manner (from left to right). The fact that this pattern was only observed under some conditions (it was modulated by practice and task complexity) could explain why previous studies have found mixed results; sometimes observing evidence for serial representation (e.g., Lien et al., 2005), and sometimes not (e.g., Monsell & Mizon, 2006).

Given the crucial role of language in novel task learning (as demonstrated in the current study; also see Van ‘t Wout & Jarrold, 2020) it is certainly possible that the serial acquisition of S-R rules is verbally mediated. For example, participants may rehearse the stimulus names in a serial manner. Unfortunately, the current study was unable to speak to that possibility as a lack of data prevented the analyses from being conducted for each distractor task condition separately. Future studies (with increased power) could focus on the extent to which such serial effects are mediated by language. Nevertheless, this is the first convincing demonstration of how serial order effects are modulated by practice, and it is vital that theories of task acquisition are able to accommodate such effects.

The final set of analyses focussed on post-error performance. One of the most robust findings in the cognitive psychology literature is that responses are typically slower following an incorrect response than following a correct response. Such “post error slowing” (e.g., Rabbitt, 1966) has been interpreted by cognitive control theories as a strategic adjustment to encourage more careful behaviour and avoid future errors (e.g., Botvinick et al., 2001). Strikingly, the current study observed the opposite pattern under select circumstances: In the early stages of trial-and-error learning of a more complex task, RTs were in fact slower following a *correct* response. There are at least three possible explanations for such post-correct slowing: Variations in the response-stimulus interval (RSI; e.g., Jentz & Dudschig, 2009), the ratio of correct to incorrect responses (Notebaert et al., 2009), and time taken to update the task-set representation.

With regards to the first explanation, there is some evidence to suggest that variations in RSI can modulate post-error slowing: Jentz and Dudschig (2009) found increased slowing in short RSI conditions. At first sight this could potentially explain the post-correct slowing observed in the current study, as the RSI was always shorter following a correct response (no feedback message) than following an error (1000 ms error message). However, the RSI account of slowing cannot explain why in the current study, post error performance was modulated by task complexity and instruction method.

Another possible explanation of the post-correct slowing observed in the current study is related to the frequency of correct and incorrect responses. According to Notebaert et al.’s (2009) orienting account, post-error slowing is the result of the relative infrequency of errors. They provided some evidence for this theory, by demonstrating that post-error slowing could be observed when errors are infrequent, however post-correct slowing was observed when correct responses were infrequent.

Although Notebaert et al.’s (2009) orienting account therefore provides a potential explanation of the post-correct slowing observed in the current study (as post-correct slowing was only observed in conditions where correct responses were relatively infrequent), there is one other explanation of the data: It is possible that the post-correct slowing in the current study reflects the time required to update the mental representation of the task-set. In trial-and-error learning, correct responses enable participants to establish an accurate and complete representation of the task-set, and this process of updating the task-set representation might take time. As the pattern of data observed in the current study (post-correct slowing occurred only under trial-and-error learning, and only when the task was more complex) is consistent with both Notebaert et al.’s (2009) orienting account, and with the task-set updating account; future studies must attempt to distinguish between these explanations of post-correct slowing.

In summary, the current study provides further evidence for the role of language in novel task learning, by showing that language plays a crucial role in the acquisition of novel tasks, but only if that task is more complex. It furthermore found a striking difference between the contribution of language to learning via-trial-and-error, where language is used to establish a verbal representation of the S-R rules during task performance; and instruction-based learning, where language is likely used to encode novel S-R rules during the instruction phase. Additional analyses furthermore suggest that novel tasks are acquired in a serial manner (from left-to-right), but only when the task is more complex, and only when that task is acquired by trial-and-error. Future research must determine whether this serial acquisition of tasks is facilitated by verbal mediation, and should also attempt to further uncover the cognitive mechanism responsible for the "post-correct slowing" observed in the current experiments. Additionally, it remains unknown whether there are long-term differences between tasks acquired via trial-and-error learning, and those acquired via instruction-based learning. For example, it is possible that trial-and-error learning results can enhance memory through richer encoding (e.g., Cyr & Anderson, 2012). Future research could usefully investigate this possibility. Finally, future research could provide a more fine-grained analysis of the effect of complexity. Such research may provide further insights into what happens when a task exceeds working memory capacity (would participants still be able to use language to maintain a subset of the S-R rules?). A fine-grained analysis of set-size would also be of interest from a developmental perspective, as it could investigate whether older children's improved ability to maintain and execute sets of S-R rules (e.g., Van 't Wout & Jarrold, 2019) is driven by age-related improvements in the use of verbal strategies (e.g., Tam, Jarrold, Sabatos-DeVito, & Baddeley, 2010). Most importantly, this study has shown that it is essential that theories of cognitive skill acquisition and instruction following not only take into account the crucial role of language, but also the factors that modulate the contribution of language to learning novel tasks.

References

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89:* 369-406.

Baddeley, A. D., Chincotta, D. M. & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General, 130*: 641-657.

Bates, E., D'Amico, S., Jacobsen, T., Szekely, A., Andonova, E., Devescovi, A., Herron, D., Lu, C. C., Pechmann, T., Pleh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., Tzeng, A., & Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review, 10 (2)*: 344-380.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., and Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*: 624-652.

Brass, M., Liefooghe, B., Braem, S., & De Houwer, J. (2017). Following new task instructions: Evidence for a dissociation between knowing and doing. *Neuroscience and Biobehavioral Reviews, 81*: 16-28.

Brass, M., Wenke, D., Spengler, S., & Waszak, F. (2009). Neural correlates of overcoming interference from instructed and implemented stimulus-response associations. *Journal of Neuroscience, 29*: 1766-1772.

Cohen-Kdoshay, O., & Meiran, N. (2009). The representation of instructions operates like a prepared reflex: Flanker compatibility effects found in first trial following S-R instructions. *Experimental Psychology, 56*: 128-133.

Cyr, A-A., & Anderson, N. D. (2012). Trial-and-error learning improves source memory among young and older adults*. Psychology and Aging, 27*: 429-439.

Ferdinand, N. K., & Kray, J. (2017). Does language help regularity learning? The influence of verbalizations on implicit sequential regularity learning and the emergence of explicit knowledge in children, younger and older adults*. Developmental Psychology, 53(3)*: 597-610.

Jentzsch, I., & Dudschig, C. (2009). Why do we slow down after an error? Mechanisms underlying the effects of posterror slowing. *Quarterly Journal of Experimental Psychology, 62*: 209-218.

Kent, C., & Lamberts, K. (2005). An exemplar account of the bow and set-size effects in absolute identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(2)*: 289-305.

Kray, J., Eber, J., & Karbach, J. (2008). Verbal self-instructions in task switching: a compensatory tool for action-control deficits in childhood and old age? *Developmental Science, 11 (2):* 2236-236.

Kray, J., Eenshuistra, R., Kerstner, H., Weidema, M., & Hommel, B. (2006). Language and action control. *Psychological Science, 17 (9)*: 737-741.

Lien, M. C., Ruthruff, E., Remington, R. W., & Johnston, J. C. (2005). On the limits of advance preparation for a task switch: Do people prepare all the task some of the time or some of the task all the time? *Journal of Experimental Psychology: Human Perception and Performance, 31 (2):* 299-315.

Lindsen, J. P., & de Jong, R. (2010). Distinguishing between de partial-mapping preparation hypothesis and the failure to engage hypothesis of residual switch costs. *Journal of Experimental Psychology: Human Perception and Performance, 36 (5):* 1207-1226.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390 (6657)*: 279-281.

Lupyan, G., & Swingly, D. (2012). Self-directed speech affects visual search performance. *The Quarterly Journal of Experimental Psychology, 65 (6*): 1068-85.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information*. Psychological Review, 63 (2):* 81-97

Miyake, A., Emerson, M. J., Padilla, F., & Ahn, J. C. (2004). Inner speech as a retrieval aid for task goals: The effects of cue type and articulatory suppression in the random task cuing paradigm. *Acta Psychologica, 115 (2-3)*: 123-142.

Monsell, S., & Graham, B. (2021). Role of verbal working memory in rapid procedural acquisition of a choice response task. *Cognition*, 214: 1-13.

Monsell, S., & Mizon, G. A. (2006). Can the task-cuing paradigm measure an endogenous task-set reconfiguration process? *Journal of Experimental Psychology: Human Perception and Performance, 32 (3):* 493-516.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64(5)*: 482-488.

Nakahara, K., Hayashi, T. Konishi, S., & Miyashita, Y. (2002). Functional MRI of macaque monkeys performing a cognitive set-shifting task. *Science, 295 (5559)*: 1532-1536.

Notebaert, W., Houtman, F., Opstal, F. V., Gevers, W., Fias, W., and Verguts, T. (2009). Post-error slowing: an orienting account. *Cognition, 111*: 275-279.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology, 4*: 272-281.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*: 195-203.

Rabbitt, P. M. (1966). Errors and error correction in choice-response tasks. *Journal of Experimental Psychology, 71*: 264-272.

Ruge, H., Karcz, T., Mark, T., Martin, V., Zwosta, K., & Wolfensteller, U. (2018). On the efficiency of instruction-based rule encoding. *Acta Psychologica, 184*: 4-19.

Stoet, G., & Snyder, L. H. (2003). Executive control and task switching in monkeys. *Neuropsychologia, 41 (10)*: 1357-1364.

Tam, H., Jarrold, C., Sabatos-DeVito, M., & Baddeley, A.D. (2010). The development of memory maintenance: Children's use of phonological rehearsal and attentional refreshment in working memory tasks. Journal of Experimental Child Psychology, 107, 306-324.

Van ‘t Wout, F. (2018). The contribution of stimulus frequency and recency to set-size effects. *Psychonomic Bulletin & Review, 25(3)*: 1123-1128.

Van ‘t Wout, F., & Jarrold, C. (2019). An investigation of children’s working memory capacity for task rules. *Cognitive Development, 51*: 14-31.

Van ‘t Wout, F., & Jarrold, C. (2020). The role of language in novel task learning. *Cognition, 194*: 1-7.

Van ‘t Wout, F., Lavric, A., & Monsell, S. (2013). Are stimulus-response rules represented phonologically for task-set preparation and maintenance*? Journal of Experimental Psychology: Learning, Memory and Cognition, 39*: 1538-1551.

Verbruggen, F., McLaren, R., Pereg, M., & Meiran, N. (2018). Structure and implementation of novel task rules: A cross-sectional developmental study. *Psychological Science, 29(7)*: 1113-1125.

Zhang, G., & Simon, H. A. (1985). STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses*. Memory & Cognition, 13(3)*: 193-201.