Van 'T Wout, F. M., & Jarrold, C. R. (2020). The role of language in novel task learning. *Cognition*, *194*, [104036]. https://doi.org/10.1016/j.cognition.2019.104036

The role of language in novel task learning

Abstract: 148 words; Text: 5523 words; 20 references; 1 table; 3 figures

This manuscript was originally submitted to Cognition on the 4th of March 2019.

A revised manuscript was submitted on the 10th of June 2019.

The work described in this article was funded by a research grant from the Leverhulme Trust (RPG-2014-414; Experiment 1); and an Experimental Psychology Society Small Grant (Experiment 2).

Abstract

The ability to rapidly acquire novel cognitive skills is a hallmark of human cognition. Theories of skill acquisition assume that this process is reliant on language, but to date this assertion has not been conclusively supported by empirical evidence. In two experiments participants (total N = 68) were required to learn, by trial-and-error, the correct response to sets of five object stimuli. To investigate the contribution of language to this process, participants performed a verbal (articulatory suppression), a non-verbal (foot tapping), or no distractor task during the first or second half of each task. In both experiments, articulatory suppression resulted in increased error rates (compared to foot tapping), but only during the first (and not the second) half of each task. These results constitute the first convincing evidence for the diminishing role of language in novel task learning and are discussed in relation to theories of skill acquisition.

Introduction

Daily life frequently requires us to learn novel tasks or skills. Whether a participant in a cognitive psychology experiment learning a set of arbitrary stimulus-response mappings for the first time, or out in the real world learning how to drive, the ability to rapidly acquire a novel skill with relative ease is crucial. One of the cognitive processes that is thought to play a vital role in the acquisition of novel tasks is language. In particular, according to theories of skill acquisition (e.g., Anderson, 1982), the acquisition of a novel task or skill begins with a "declarative phase", during which language is required to maintain the task rules in working memory. More recent models of "instruction following" (e.g., Brass, Liefooghe, Braem, & De Houwer, 2017) similarly include an initial stage during which linguistic information is transformed into a procedural representation, which guides behavior. In both accounts, use of language is particularly important during the early stages of novel task learning; although this declarative representation may continue to exist beyond the initial declarative phase, it is the procedural representation that is thought to govern action as the task becomes more practiced.

Despite this, there is currently no convincing evidence for the diminishing role of language with practice. Three studies have provided relevant (though not conclusive) evidence; the results of these studies are briefly summarized here[1]. Firstly, Kray, Eenshuistra, Kerstner, Weidema and Hommel (2006) found that 4-year old children benefited from verbal labelling when learning novel action-effect associations. Although these results suggest language can help children to learn novel tasks, they do not demonstrate that the role of language diminishes with practice, as the theories described above would predict. Another study by Kray and colleagues (Kray, Eber, & Karbach, 2008) examined the contribution of language to task switching performance, by requiring children and adults to switch between two tasks whilst engaging in either a task-relevant verbalization (naming the next task), a task-irrelevant verbalization (saying an over-learned three-word sequence, one word per trial), or no verbalization. Although the effect of a task-irrelevant verbalization on the mixing cost (difference between single-task blocks and mixed-task blocks) decreased with practice, significant effects of verbalization on the mixing cost remained even after extensive practice (>1000 trials). While this result is seemingly inconsistent with the abovementioned notion that language does not support the performance of well-practiced tasks, it is possible that the function of language

[1] It should be noted that there are many more studies that have investigated the contribution of language to task switching performance, but because the current study is specifically interested in how the contribution of language changes with practice, we have reviewed only those studies which analysed performance as a function of practice or those that have focussed on the learning of novel S-R associations.

in a task switching scenario differs from the role that language plays in transforming the declarative representation of a simple task rule into a procedural one. Specifically, the former involves additional processes (including keeping track of the task sequence, and retrieving a task-set from long term memory) which may rely on aspects of language beyond the initial practice phase (cf. Miyake, Emerson, Padilla, & Ahn, 2004). Finally, one other task switching study designed to investigate the contribution of language to task-set control manipulated the phonological similarity of the stimulus terms (Van 't Wout, Lavric, & Monsell, 2013). This study did not find any effect of phonological similarity on task switching performance once the tasks were well-practiced. However, right at the beginning of the experiment (when participants were practicing the tasks in single task blocks), performance was worse for phonologically similar sets than for phonologically dissimilar sets of stimuli, suggesting that participants rely on a phonological representation of the stimulus-response (S-R) rules when learning a novel task. But there are other important differences between single task and task switching blocks (e.g., Monsell, 2003) that complicate the interpretation of that result. Furthermore, Van 't Wout et al.'s (2013) study was not designed to investigate the diminishing role of language as a function of practice (it contained only one phonologically similar and one dissimilar task per experiment), and therefore a systematic investigation into this process is still required.

To summarize, although these studies suggest that language plays a role when learning novel tasks, they do not provide conclusive evidence for the diminishing role of language with practice. This would require an experimental paradigm with the following properties: 1) inclusion of an appropriate non-verbal dual task control condition; 2) several novel S-R tasks (to achieve sufficient power; also see Cole, Laurent, & Stocco, 2013); and 3) sufficient trials per task so that performance can be analyzed as a function of practice (when the task is completely novel, versus when the task is well-practiced). The paradigm employed in this study was designed with those conditions in mind. Specifically, it required participants to learn, by trial-and-error, the correct response to novel sets of line drawings. A trial-and-error based paradigm was used rather than instruction-based learning, because recent studies have shown that under some conditions, participants are able to "proceduralize" task instructions prior to performance (e.g., Cohen-Kdoshay & Meiran, 2009). Using a trial-and-error procedure allowed us to capture this process of proceduralization as it occurred during task performance. To investigate the contribution of language to this process, participants learnt each novel task whilst performing either a verbal distractor task (articulatory suppression; AS), a non-verbal distractor task (foot tapping; FT), or no distractor task (control condition). Importantly, we also manipulated the order of these distractor tasks (they could each be performed during the first or the second half of each novel S-R task). We hypothesized that if language is especially important during the early phases of novel task learning, then AS should disrupt performance more than FT when it is performed during the first half of each task, but not when it is performed during the second half of each task.

Experiment 1 Method

Subjects

Thirty-six participants (aged between 18 and 32 [mean age = 20], 30 female) provided informed consent prior to taking part. All participants were awarded 1 course credit for taking part. Experiments 1 and 2 were approved by the University of Bristol's School of Psychological Science Human Research Ethics Committee (ID 75541 and ID 78862, respectively).

Procedure

The experimental task required participants to learn, by trial-and-error, the correct response to novel sets of five black and white line drawings. In total, nine sets of five stimuli were selected from the International Picture Naming Project (IPNP; Bates et al., 2003). Within a set, stimuli were selected as to avoid phonological, semantic or visual similarity (see Table 1). Across sets, stimuli were matched for average naming latency, which served as an indirect measure of frequency (Oldfield & Wingfield, 1965); and percent naming agreement (proportion of all trials on which participants produced the dominant target name), which was at least 98% for all stimuli.

#	Set 1	RT	%	#	Set 2	RT	%	#	Set 3	RT	%
1	book	656	100	**6**	hat	684	98	**11**	ear	681	100
2	car	751	100	**7**	spoon	777	100	**12**	watch	780	100
3	tree	796	100	**8**	tent	744	100	**13**	bus	771	100
4	fan	865	98	**9**	box	753	100	**14**	leaf	848	100
5	sun	762	100	**10**	pig	855	100	**15**	pen	753	100
	Mean	**766**	**100**		**Mean**	**763**	**100**		**Mean**	**766**	**100**
#	**Set 4**	**RT**	**%**	**#**	**Set 5**	**RT**	**%**	**#**	**Set 6**	**RT**	**%**
16	key	738	100	**21**	foot	758	98	**26**	chair	732	100
17	dog	702	100	**22**	moon	804	100	**27**	hand	723	98
18	cake	789	100	**23**	house	745	98	**28**	train	838	100
19	heart	720	100	**24**	bread	773	98	**29**	snake	775	100
20	ball	886	100	**25**	frog	751	100	**30**	kite	796	100
	Mean	**767**	**100**		**Mean**	**766**	**99**		**Mean**	**773**	**100**
#	**Set 7**	**RT**	**%**	**#**	**Set 8**	**RT**	**%**	**#**	**Set 9**	**RT**	**%**
31	bed	706	100	**36**	eye	700	98	**41**	bell	703	100
32	fish	777	100	**37**	door	719	100	**42**	flag	847	100
33	cheese	843	100	**38**	broom	821	100	**43**	horse	809	100
34	clock	772	98	**39**	saw	863	100	**44**	comb	717	100
35	nose	721	100	**40**	dress	840	100	**45**	sock	712	100
	Mean	**764**	**100**		**Mean**	**789**	**100**		**Mean**	**757**	**100**

Table 1. Picture names for each of the nine (Experiment 1; Sets 1-9) or eight (Experiment 2; Sets 1-8) stimulus sets used. Stimuli were matched for percent name agreement (%) and RT target mean (mean latency for dominant responses only).

For each task, stimuli were presented centrally on the computer screen, one at a time. Participants were instructed to place five fingers (three from the left hand, two from the right for half the participants; vice versa for the other half) on the x, c, v, b and n keys (covered by black stickers) of a standard QUERTY keyboard. The stimulus remained on the screen until a response was made. An error message (“Error!”) was displayed for 1000ms following an incorrect response. Prior to each block, participants received the following instructions: “You are going to sort five different pictures. You will have to learn the correct response to each picture. To begin with, you will have to guess (I won't tell you what the correct responses are). Remember to try your best (even though you will make some mistakes to begin with)!”. These generic instructions were followed by a sentence instructing participants which distractor task (see below) should be performed for the duration of each block.

Participants performed 200 trials with each of the nine novel S-R sets; this trial sequence was split into two halves (“blocks”) of 100 trials (total of 1800 trials). To examine the effect of language on novel task learning, performance (mean correct RT and % error) was examined under three conditions: articulatory suppression (AS; which has been shown to disrupt the use of inner speech; e.g., Baddeley, Chincotta, & Adlam, 2001), foot tapping (FT; which does not interfere with language but is well-matched to AS in terms of difficulty; Miyake et al., 2004), or no distractor task (control condition). For the AS condition, participants were required to say “tick, tick, tick” to a metronome. For the FT condition, participants were asked to tap one foot to the beat of the metronome. During the control condition (no distractor task), the metronome remained on, but participants were instructed to ignore it. For all conditions, the metronome was set to 100 beats per minute. To ensure participants performed the distractor tasks correctly, an experimenter was present at all time. On the rare occasion that a participant forgot to engage in AS or FT, the experimenter would immediately remind the participant (e.g., “Don’t forget to tap your foot!”).

To examine whether the role of language in novel task learning is restricted to the early phases of learning, the following conditions were compared: 1) AS in the first half followed by FT in the second half (AS-FT); 2) FT in the first half followed by AS in the second half (FT-AS); 3) no distractor task in the first or the second half (none-none). Each participant performed all three conditions three times (resulting in 9 novel tasks of 200 trials each; a total of 1800 trials). The order of conditions, the assignment of stimuli to conditions, and the assignment of responses to stimuli were balanced between subjects. The presentation of stimuli was pseudorandomized so that there were no immediate stimulus repetitions, and each stimulus occurred 4 times within a subblock of 20 trials.

At the start of each block, participants were instructed to perform a distractor task (AS or FT) or not. They were not shown the stimuli; instead they were instructed to learn by trial-and-error the correct response for a set of 5 stimuli. The trial sequence was as follows: a 250 ms centrally presented fixation cross was followed by the stimulus, which remained on screen until a response was made. Feedback (a 1000ms “Error!” message) only occurred on incorrect trials (see Figure 1 for an example of the trial procedure).

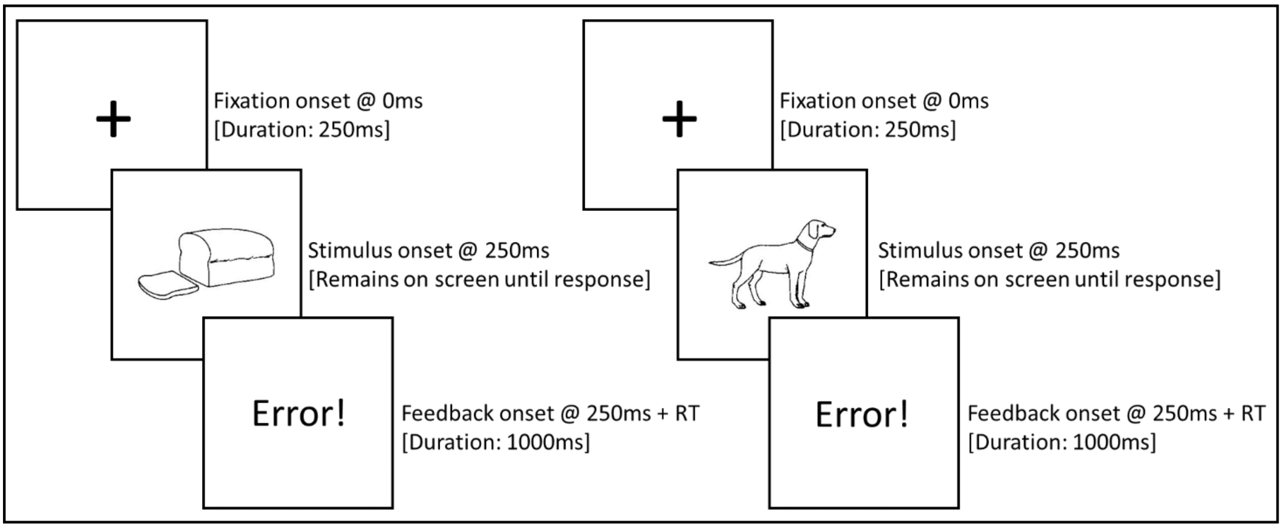


Figure 1. Example of a trial sequence (displaying two consecutive trials) in Experiments 1 and 2.

Prior to the start of the experimental session, all participants also completed a baseline task, which was designed to examine the effect of both distractor tasks on performance (to ensure that they were well matched in terms of difficulty). In this baseline task, participants were required to respond to a separate set of five line drawings selected from the IPNP. Each line drawing could appear in green or blue; and participants were instructed to respond to the green stimuli by pressing the left key (a), and the blue stimuli by pressing the right key (l). The trial sequence in the baseline task was identical to the trial sequence in the experimental task; except that in the baseline task, a green circle was presented in the left lower corner of the screen; and a blue circle was presented in the right lower corner of the screen. This was done to minimize any memory load and obtain a "pure" measure of the difficulty of both distractor tasks. Each participant completed 20 practice trials, followed by 40 trials of each distractor task type (AS, FT or none), resulting in a total of 140 trials for the baseline task. The order of distractor task types was balanced between participants.

In total, the experiment lasted approximately 1 hour. The experiment was programmed in Psychopy (Peirce et al., 2019) and run on a Toshiba laptop. Participants were tested one at a time, and the experimenter remained present to ensure that the participant was performing the distractor task as required. On completion of the experiment, all participants were thanked and debriefed.

Experiment 1 Results

Prior to conducting the analyses described below, reaction times (RTs) greater than 5000ms (0.4% of correct responses) were removed from the data set. Throughout, the number following the ± symbol indicates the 95% confidence interval, which was computed using the Cousineau-Morey method (Morey, 2008) for calculating confidence intervals for within-subjects designs.

Baseline task

To examine whether AS and FT were well-matched in terms of difficulty, a one-way repeated measures ANOVA with the factor distractor task type (AS, FT or none) was run on the accuracy (% error) and mean correct RT (ms) data from the baseline condition.

For the % error data, a significant main effect of distractor task type reflected increased error rates under AS (5.5±1.2%) and FT (5.5±1.0%), compared to the control condition (3.8±1.1%), F(2,70)=3.36, p=.040, η_p^2=.088. Further ANOVAs revealed that error rates were significantly increased under AS compared to the control condition, F(1,35)=4.48, p=.041, η_p^2=.113; and under FT compared to the control condition, F(1,35)=6.49, p=.015, η_p^2=.156; but that the difference between AS and FT was not significant, F(1,35)<0.011, p>.999, η_p^2<.001.

Similarly, for the mean correct RTs, a significant main effect of distractor task type reflected increased RTs under AS (520±15ms) and FT (539±18ms), compared to the control condition (481±16ms), F(2,70)=13.84, p<.001, η_p^2=.283. Again, further ANOVAs revealed that RTs were significantly greater under AS than in the control condition, F(1,35)=14.88, p<.001, η_p^2=.298; and under FT than in the control condition, F(1,35)=23.30, p<.001, η_p^2=.400; but that the difference between AS and FT was not significant, F(1,35)=2.85, p=.100, η_p^2=.075. Note that an increase in RT under FT compared to AS should not negate any detrimental effect of AS on performance in the experimental task (if anything, the increased RTs under FT should make it more difficult to observe such an effect).

Experimental task

A 3 (distractor task type: AS, FT or none) x 2 (half: first versus second half) repeated measures ANOVA was conducted on the accuracy data (% error) and the mean correct RT data (ms). The % error analysis (see Figure 2, left panel) found significant main effects of half, F(1,35)=270.43, p<.001, η_p^2=.885, and distractor task type, F(2,70)=12.93, p<.001, η_p^2=.270. Importantly, the two-way interaction between distractor task type and half was also significant, F(2,70)=9.22, p=.001, η_p^2=.209. Further ANOVAs found that participants made more errors under AS than under FT, but only in the first half of each task, F(1,35)=10.79, p=.002, η_p^2=.236 (AS: 24.9±2.6%; FT: 19.5±1.8%), and not in the second half, F(1,35)=4.07, p=.052, η_p^2=.104, during which participants made marginally more errors under FT (9.0±1.1%) than AS (7.7±1.2%). One-way ANOVAs comparing AS and the control condition found that participants made more errors under AS (24.9±2.6%) than in the control condition (18.9±1.4%) in the first half, F(1,35)=18.96, p<.001, η_p^2=.351, and in the second half, F(1,35)=5.86, p=.021, η_p^2=.143 (AS: 7.7±1.2%; none: 6.7±1.3%). Finally, the difference between FT and the control condition was not significant in the first half, F(1,35)=0.36, p=.552, η_p^2=.010 (FT: 19.5±1.8%; none: 18.9±1.4%), but it was significant in the second half, F(1,35)=10.16, p=.003, η_p^2=.225 (FT: 9.0±1.1%; none: 6.7±1.3%).

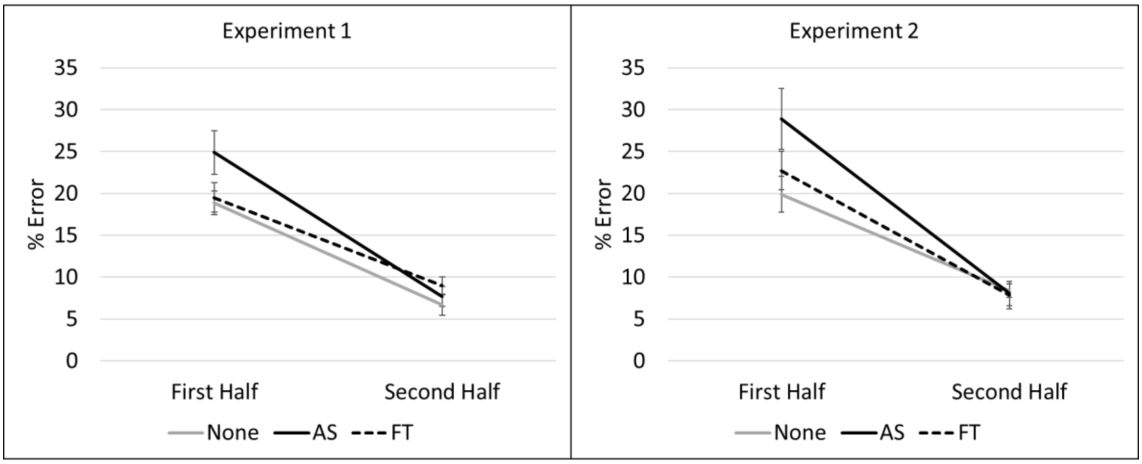


Figure 2. Accuracy data (mean % error) in Experiment 1 (left) and 2 (right) for each distractor task type (AS, FT or none) plotted as a function of task half (first 100 trials versus second 100 trials). Error bars reflect 95% confidence intervals.

For the mean correct RT data, the same 3 (distractor task type: AS, FT or none) x 2 (half: first versus second half) repeated measures ANOVA found a significant main effect of condition, F(2,70)=6.61, p=.003, η_p^2=.159, and a significant main effect of half, F(1,35)=54.42, p<.001, η_p^2=.609, but no significant interaction between condition and half, F(2,70)=1.46, p=.241, η_p^2=.040 (also see Appendix). With regards to the main effect of half, participants were slower in the first half (911±17ms) than in the second half (823±17ms). With regards to the main effect of distractor task type, three further ANOVAs contrasting the three distractor task types (averaged over half) found that RTs were faster in the control condition (843±19ms) than under AS (872±19ms), F(1,35)=4.11, p=.050, η_p^2=.105, and FT (886±14ms), F(1,35)=15.10, p<.001, η_p^2=.301; but that the difference between AS and FT was not significant, F(1,35)=1.83, p=.184, η_p^2=.050.

Experiment 1 Summary

The results of Experiment 1 clearly show that AS led to more errors than FT, but only when AS was performed during the first (and not the second) half of each task.

However, although the effect of distractor task type during the first half of each task is unambiguous, differences between the AS and FT conditions in the second half are more difficult to interpret, because performance in that half is likely to be influenced by performance in the first half. Specifically, error rates in the FT condition during the second half may be inflated because that condition was always preceded by AS in the first half, and participants made more errors when performing AS during the first half. Hence, although the data of Experiment 1 show that AS affects performance more than FT when the task is novel, they do not conclusively demonstrate that there are no detrimental effects of AS on performance once the task is well-practiced.

To obtain a purer estimate of the effects of AS and FT on performance once the task is well-practiced, a second experiment was run in which the effect of AS and FT on performance in the second half was assessed when participants performed no distractor task in the first half (none-AS versus none-FT). Experiment 2 also included two further conditions (AS-none and FT-none), with the aim of replicating the results of Experiment 1. Aside from this change to the order of the distractor tasks (and the resulting change to the overall number of trials), Experiment 2 was identical to Experiment 1, as described below.

Experiment 2 Method

Subjects

Thirty-two participants (aged between 18 and 28 [mean age = 20], 28 female) provided informed consent prior to taking part. None of the participants had taken part in Experiment 1, and all were awarded 1 course credit for taking part.

Procedure

In Experiment 2, participants learned 8 novel tasks, each consisting of 5 sets of black and white line drawings (Sets 1-8 in Figure 1). Each task was performed under one of the following four conditions:

1) AS in the first half followed by no distractor task in the second half (AS-None); 2) No distractor task in the first half followed by AS in the second half (None-AS); 3) FT in the first half followed by no distractor task in the second half (FT-None); 4) No distractor task in the first half followed by FT in the second half (None-FT). As in Experiment 1, the order of distractor task type, the assignment of stimulus sets to distractor task type and the response assignments were counterbalanced between participants. Each participant performed each of the four conditions twice, resulting in a total of 1600 trials. Again, participants performed a baseline condition (identical to Experiment 1) prior to the experimental condition. Experiment 2 lasted approximately 50 minutes.

Experiment 2 Results

As in Experiment 1, reaction times (RTs) greater than 5000ms (0.3% of correct responses) were removed from the data set prior to conducting the analyses described below.

Baseline task

A one-way ANOVA with the repeated measures factor distractor task type (AS, FT or none) showed that participants made fewer errors when there was no distractor task (3.8±1.3%), compared to the AS (7.0±1.5%) and FT (5.9±1.0%) conditions, F(2,62)=6.69, p=.003, η_p^2=.178. Consistent with Experiment 1, further ANOVAs demonstrated that the difference between AS and the control condition and the difference between FT and the control condition were both significant (F(1,31)=9.38, p=.005, η_p^2=.232 and F(1,31)=8.88, p=.006, η_p^2=.223, respectively); whilst the difference between AS and FT was not, F(1,31)=1.36, p=.252, η_p^2=.042.

With regards to the mean correct RT data in the baseline condition, the same one-way ANOVA found a significant main effect of distractor task (AS, FT or none), F(2,62)=8.76, p=.001, η_p^2=.220. Further ANOVAs revealed that RTs were significantly greater under FT (570±29ms) compared to both the AS (515±20ms) and the control condition (509±19ms), F(1,31)=8.92, p=.005, η_p^2=.223, and F(1,31)=11.22, p=.002, η_p^2=.266, respectively. The difference between the AS and control condition was not significant, F(1,31)=0.39, p=.538, η_p^2=.012.

Experimental task

A 2 (distractor task type: AS or FT) x 2 (half: first or second) repeated measures ANOVA was run on the accuracy data (% error) and the mean correct RT data (ms). Note that data from the blocks in which participants did not perform a distractor task (control condition), though displayed in Figure 2 (right panel), were not included in this analysis. This was because the main aim of this analysis was to compare the effects of FT and AS on performance once the task was well-practiced; and because the control condition in the second half was always preceded by either FT or AS in the first half, unlike in the first experiment.

For % error data, the abovementioned ANOVA showed a significant interaction between distractor task type and half, F(1,31)=9.53, p=.004, η_p^2=.235. Two further one-way ANOVAs contrasting the distractor tasks (AS or FT) within each half showed that in the first half, participants made more errors under AS (28.9±3.6%) than under FT (22.7±2.3%), F(1,31)=10.52, p=.003, η_p^2=.253, but that there was no significant difference in the second half (AS: 8.1±1.4%; FT: 7.9±1.7%), F(1,31)=0.24, p=.626, η_p^2=.008. As the effects of AS and FT were expected to be equivalent after practice, a Bayesian one-way ANOVA with default priors was conducted (JASP Team, 2018) to compare the effect of AS and FT on performance in the second half of each task. This analysis found "positive" evidence (BF_{01} = 3.685; Raftery, 1995) in support of the null hypothesis of no meaningful difference under the two types of distraction once the tasks were well-practiced.

For the mean correct RT (ms) data, the same 2 x 2 repeated measures ANOVA found only a significant main effect of half, F(1,31)=10.22, p=.003, η_p^2=.248, reflecting increased RTs in the first half (970±26ms) compared to the second half (912±26ms). The main effect of distractor task type (FT or AS) and the interaction between half and distractor task type were not significant, F(1,31)=0.27, p=.609, η_p^2=.009, and F(1,31)=0.14, p=.712, η_p^2=.004, respectively (also see Appendix).

Effect of articulatory suppression versus foot tapping within a block across Experiments 1 and 2

The separate analyses described above for each of Experiments 1 and 2 compared the effect of distractor task type (AS, FT or none) in the first and second half of each task. However, a more detailed analysis can also be performed, in which the data are plotted separately for each distractor task type (AS, FT or none) as a function of stimulus occurrence (each stimulus occurred 20 times within each half). This analysis was run on the data from Experiments 1 and 2 combined to increase power (as there were only 15 trials per stimulus occurrence per subject for each distractor task type) and to reduce the likelihood of Type I errors.

One prediction with regards to this analysis is that the effect of distractor task type (AS or FT) on accuracy should not be linear (i.e., largest at the start of each task and then steadily decreasing). Instead, one might expect AS to have no adverse effects on performance (above and beyond that of FT) at least the very first time participants encounter a stimulus because at that point in time, there is no linguistic representation of the S-R rule yet. If so, then one might expect the effect of AS (compared to FT) on accuracy to follow an inverted U-shaped trend, where the effect of AS on accuracy increases at first (as participants use language to compile the task-set), and then decreases thereafter (as performance becomes more automatic and the role of language diminishes).

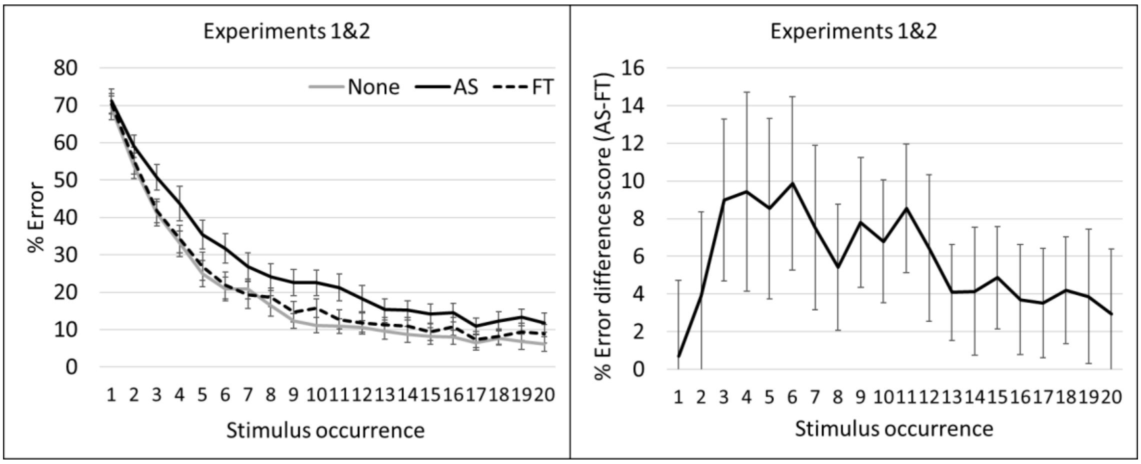


Figure 3. Accuracy data (mean % error) as a function of stimulus occurrence (1-20); plotted separately for each distractor task type (AS, FT or none; left) and as a difference score (AS minus FT; right). Error bars reflect 95% confidence intervals.

A 2 (distractor task type: AS or FT) x 20 (stimulus occurrence: 1-20) repeated measures ANOVA with Experiment (1 or 2) as a between-subjects factor confirmed this prediction, $F(1,66)=5.86$, $p=.018$, $\eta_p^2=.082$ (two-way interaction between distractor task type and the quadratic component of occurrence). As predicted, the effect of AS on the error rate increases, and then decreases, as a function of stimulus occurrence (see Figure 3).

Experiment 2 Summary

The results of Experiment 2 replicate those of Experiment 1: AS again resulted in increased error rates compared to FT, but only during the first half of each task. Furthermore, Experiment 2 was able to determine that when the task is well-practiced (AS or FT preceded by no distractor task in the first half), AS did not have a detrimental effect on performance compared to FT. This observation was further supported by a Bayesian analysis, which found "positive evidence" in support of the null hypothesis, according to which the effect of AS and FT on performance is equivalent once the task is well-practiced.

Discussion

The two experiments reported here demonstrate for the first time that participants use language when learning a novel cognitive task by trial-and-error; and that the role of language diminishes with practice. After 100 trials of practice, error rates were no longer significantly increased under articulatory suppression (AS) compared to foot tapping (FT). This profound effect of AS on initial task performance cannot be attributed to an increased difficulty of this distractor task per se , as data from the baseline task showed that AS and FT yield comparable error rates when the task does not require the participant to learn novel S-R mappings (in the baseline task the correct S-R mappings remained on screen throughout). In fact, RTs were significantly greater under FT compared to AS in the baseline task of Experiment 2 (the same trend was observed in Experiment 1), making the detrimental effect of AS on novel task learning even more striking.

A more detailed analysis of performance as a function of practice (in which error data were plotted as a function of stimulus occurrence) shed further light on the role of language in novel task learning. Specifically, at the very beginning of each task, performance was no worse under AS than under FT; a difference between the two conditions then appeared, and then disappeared towards the end of the first half of each task. This result is entirely consistent with an account of novel task learning which stipulates that people use language to compile a mental representation of the task. According to such an account, when each stimulus is first encountered, participants do not yet have a linguistic representation of the S-R rule, because in our trial-and-error paradigm that would require (at least) one correct response. Consequently, AS does not disproportionally disrupt performance the first time a stimulus is encountered, but it does thereafter. Then, as performance becomes more automatic, language ceases to support performance, and the difference between AS and FT disappears again. These results constitute the first convincing demonstration in support of the diminishing role of language in skill acquisition (Anderson, 1982). They are also consistent with theories of working memory (Oberauer, 2009) and task-set control (Monsell, 2017), which assume that although a novel task may be represented linguistically during acquisition, performance is ultimately governed by a non-linguistic, procedural representation of the task.

One other noteworthy finding of the present study is that in both experiments, the detrimental effect of AS was restricted to the error rates (in the RTs, the interference caused by AS and FT was equivalent). This pattern of data sheds light on the specific role that language might play in novel task learning. In particular, it suggests that AS affects participants' ability to form an accurate representation of the relevant S-R rules; but it does not affect the time required to retrieve or implement an S-R rule once an accurate representation of that S-R rule has been established. The former would result increased error rates under AS (which were found in both experiments), whereas the latter would result in increased RTs under AS (which were not found). Finally, it is worth noting that in Kray et al.'s (2006) study, effects of verbalisation also manifested in the error (and not the RT) data, further confirming that the difference between the RT and accuracy patterns in the current study are not an anomaly, but rather a reflection of the specific role that language plays in novel task learning.

In relation to previous findings, our results are consistent with those of Van 't Wout et al. (2013), who found that S-R rules are not represented phonologically once the task is well-practiced. Conversely, the current results may at first appear to contradict Kray et al.'s (2008) finding of a significant effect of AS on the mixing cost even after extensive practice (>1000 trials). However, as Kray et al. (2008) did not use a non-verbal distractor task for comparison, it is possible that at least some of the residual effect of AS after practice is due to generic dual task demands. Additionally, there are other differences between Kray et al.'s (2008) and the current study, which could explain this apparent discrepancy. Specifically, it is possible that the role played by language in a task switching scenario (e.g., Kray et al., 2008) differs from the contribution of language to the acquisition on novels sets of simple S-R rules (e.g., Van 't Wout et al., 2013). For example, task switching involves additional processes, such as retrieving the task goal from long term memory; and it is possible these processes continue to be supported by language beyond the initial practice phase (cf., Miyake et al., 2004). Indeed, the idea that the contribution of language to task performance may vary according to the task demands could explain why in some studies, the role of language is restricted to the early stages of practice (e.g., Van 't Wout et al., 2013), whereas in other studies, beneficial effects of verbal labelling are observed throughout the experimental session (Kray et al., 2008), or even increase during the course of the task (Lupyan & Swingley, 2012; Ferdinand & Kray, 2017).

In the current study, language appears to be supporting the acquisition of novel sets of S-R rules, though some questions remain with regards to the precise nature of this process. For example, are participants using language to label the stimuli, or the responses, or both? The results of Van ‘t Wout et al. (2013), which found that performance was affected by the phonological similarity of the stimulus names, would suggest that language is being used for the phonological recoding of visual stimuli. Additionally, the results of Kray et al. (2006) suggest that language specifically supports the binding of action-effect associations. Future research could further explore these possibilities by requiring participants to selectively label the stimulus and/or response components of task-set (cf. Kray et al., 2006), whilst also examining whether and how the effects of labelling are modulated by practice.

The current results also speak to recent models of instruction following (Brass et al., 2017), according to which language helps to transform the instructions into a procedural representation that guides behavior. Future experiments will have to confirm whether language also supports the acquisition of novel tasks in instruction-based learning paradigms. The effect of AS on novel task learning by instruction may differ from the effect described here in two ways: Firstly, one might expect the effect of AS on novel task learning to be less pronounced in instruction-based learning, as participants may proceduralise the task prior to task performance (Cohen-Kdoshay & Meiran, 2009). Alternatively, in an instruction-based learning paradigm, participants might be able to rely on a linguistic representation of the S-R rules right from the beginning, and if so one would expect to see differences between performance under AS and FT even for the first stimulus occurrence, in contrast to the data described here.

Finally, one area in which this paradigm may prove useful is within developmental psychology. The ability to use language to guide behavior is thought to improve throughout childhood (e.g., Cragg & Nation, 2010). Consequently, it is possible that age differences in novel task learning are the result of developmental differences in the use of verbal strategies. If this is the case, then one might expect age differences in novel task learning to be eliminated under AS. Our paradigm could enable future studies to answer these (and other) questions, and would help to reveal the precise contribution of language to novel task learning. Those findings will in turn have important consequences for theories of skill acquisition and novel task learning; and they could be used to help determine the best way to teach novel tasks and skills to adults and children, both within and beyond educational settings.

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.

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.

Cole, M. W., Laurent, P., & Stocco, A. (2013). Rapid instructed task learning: A new window into the human brain's unique capacity for flexible cognitive control. *Cognitive, Affective & Behavioral Neuroscience, 13 (1)*: 1-22.

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.

Cragg, L., & Nation, K. (2010). Language and the development of cognitive control. *Topics in Cognitive Science, 2 (4)*: 631-642.

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.

JASP Team, 2018. JASP (Version 0.9)[Computer software].

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.

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

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7 (3)*: 134-140.

Monsell, S. (2017). Task Set Regulation. In T. Egner (Ed), *The Wiley Handbook of Cognitive Control* (pp. 29-49). Wiley-Blackwell.

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.

Oberauer, K. (2009). Design for a working memory. *Journal of learning and motivation, 51:* 45-100.

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. Advance online publication.

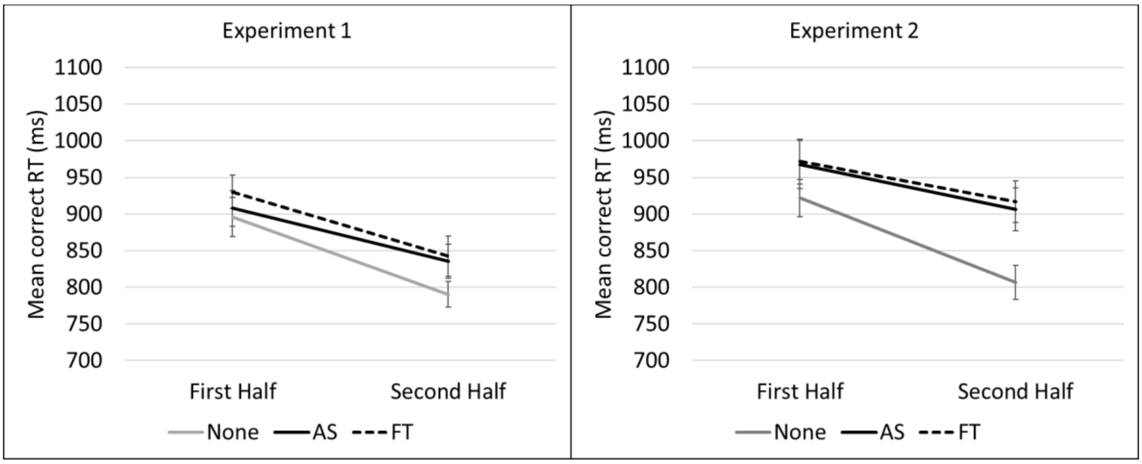
Raftery, A.E. (1995). Bayesian model selection in social research. In: Marsden, P.V. (Ed.), Sociological methodology 1995. Blackwell, Cambridge, MA, pp. 111–196.

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.

Author contributions

F. van 't Wout and C. Jarrold jointly designed the study. F. van 't Wout performed the data analyses and drafted the manuscript. C. Jarrold provided critical revisions. Both authors approved the final version of the manuscript for submission.

Appendix



Mean correct RT (ms) in Experiment 1 (left) and 2 (right) for each distractor task type (AS, FT or none) plotted as a function of task half (first 100 trials versus second 100 trials). Error bars reflect 95% confidence intervals.