There are limits to the effects of task instructions: Making the automatic effects of task instructions context-specific takes practice. — Source link ↗

Senne Braem, Baptist Liefooghe, Jan De Houwer, Marcel Brass ...+1 more authors

Institutions: Ghent University

Related papers:

- The Power of Instructions: Proactive Configuration of Stimulus-Response Translation

- Instruction-based task-rule congruency effects.

- Instruction-based response activation depends on task preparation

- When planning results in loss of control: intention-based reflexivity and working-memory.

- Automatic motor activation by mere instruction

There Are Limits to the Effects of Task Instructions: Making the Automatic Effects of Task Instructions Context-Specific Takes Practice

Senne Braem, Baptist Liefooghe, Jan De Houwer, Marcel Brass, and Elger L. Abrahamse
Ghent University

Unlike other animals, humans have the unique ability to share and use verbal instructions to prepare for upcoming tasks. Recent research showed that instructions are sufficient for the automatic, reflex-like activation of responses. However, systematic studies into the limits of these automatic effects of task instructions remain relatively scarce. In this study, the authors set out to investigate whether this instruction-based automatic activation of responses can be context-dependent. Specifically, participants performed a task of which the stimulus-response rules and context (location on the screen) could either coincide or not with those of an instructed to-be-performed task (whose instructions changed every run). In 2 experiments, the authors showed that the instructed task rules had an automatic impact on performance—performance was slowed down when the merely instructed task rules did not coincide, but, importantly, this effect was not context-dependent. Interestingly, a third and fourth experiment suggests that context dependency can actually be observed, but only when practicing the task in its appropriate context for over 60 trials or after a sufficient amount of practice on a fixed context (the context was the same for all instructed tasks). Together, these findings seem to suggest that instructions can establish stimulus-response representations that have a reflexive impact on behavior but are insensitive to the context in which the task is known to be valid. Instead, context-specific task representations seem to require practice.

Keywords: cognitive control, task sets, instructions, context-sensitivity

The ability to use and share symbolic representations is commonly thought to separate humans from other animals, equipping us with unique language abilities that guide our everyday action and perception (Deacon, 1997). For example, using language we can learn without trial and error the route to a new city, how to build a cabinet, or how to prepare a meal. Although verbal instructions are omnipresent in daily life and psychological research, the mechanisms via which they influence behavior are still poorly understood. In line with this observation, Cole, Laurent, and Stocco (2013) recently stressed the need for a more systematic investigation into this unique ability, to help accelerate the insights in what is now still a relatively small research domain.

Intriguingly, recent expeditions into the domain of instruction learning already demonstrated how the presentation of a

Correspondence concerning this article should be addressed to Senne Braem, Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, B–9000 Ghent, Belgium. E-mail: senne.braem@ugent.be

stimulus can result in the automatic activation of a response that was assigned to that stimulus via instructions, even when the stimulus-response (S-R) instruction was never before executed (e.g., Liefooghe, Wenke, & De Houwer, 2012; Liefooghe, De Houwer, & Wenke, 2013; Meiran, Pereg, Kessler, Cole, & Braver, 2015a; Theeuwes, Liefooghe, De Houwer, 2014; Wenke, Gaschler, & Nattkemper, 2007). Mere S-R instructions were also shown to influence automatic motor activation during stimulus presentation as indexed via electroencephalography and electromyography (Bardi, Bundt, Notebaert, & Brass, 2015; Everaert, Theeuwes, Liefooghe, & De Houwer, 2014; Meiran, Pereg, Kessler, Cole, & Braver, 2015b). These findings suggest that instructions can establish S-R associations that are activated in an almost reflexive manner upon stimulus presentation (Meiran et al., 2015b). Although evidence suggests that practice does further strengthen these instructed associations (Wenke, De Houwer, De Winne, & Liefooghe, 2015), the boundary conditions of the automatic impact of task instructions remain largely unknown. In the current study, we set out to investigate to which extent instruction-based automatic behavior is restricted to the context in which the instructions are said to be valid.

In our everyday lives, we are often required to learn that certain actions or tasks are only applicable in certain contexts. For example, the sound of your doorbell might trigger an automatic tendency to stand up when you hear the sound in your own house, but not when hearing it at a friend's house. This impact of task context has received ample attention in the domain of trained (rather than merely instructed) task sets, demonstrating that responses can be automatically activated in a stimulus-driven and context-dependent manner (e.g., Abrahamse

& Verwey, 2008; Braem, Abrahamse, Duthoo, & Notebaert, 2014; Bugg, Jacoby, & Toth, 2008; Cañadas, Rodríguez-Bailón, Milliken, & Lupiáñez, 2013; Chun & Jiang, 1998; Crump, Gong, & Milliken, 2006; Crump, Vaquero, & Milliken, 2008; Crump & Logan, 2010; King, Korb, & Egner, 2012; Mayr & Bryck, 2005, 2007; Reuss, Desender, Kiesel, & Kunde, 2014; Rubin & Koch, 2006; Ruitenberg, Abrahamse, De Kleine, & Verwey, 2012; Schouppe, de Ferrerre, Van Opstal, Braem, & Notebaert, 2014). In fact, context-dependence seems to be an almost automatic consequence of systematically repeating a certain task within a specific context, and applies, among others, to the processing of response conflict (e.g., Crump et al., 2006), the deployment of spatial attention (e.g., Chun & Jiang, 1998), or the learning of sequential motor skills (e.g., Abrahamse & Verwey, 2008). These studies often differ both in terms of paradigm and specific effect of interest. However, the manipulations and general motivation behind these studies is very similar. These studies have in common that they seem to evidence how task or goal representations can become bound to the context in which they were most frequently experienced (as reviewed in Abrahamse, Braem, Notebaert, & Verguts, 2016). In this study, we are particularly interested in the context-specificity of (instructed) S-R representations (e.g., Mayr & Bryck, 2007).

In the lab, this context-dependency is often tested by presenting a certain task or stimulus more in one over the other context (e.g., location on the screen, background color). After a certain amount of practice, it is then observed that context starts to function as a cue for the application of the task. For example, in the study by Mayr and Bryck (2007) participants alternated between two tasks in which they judged either the orientation (i.e., when vertical press left, when horizontal press right), or the color (i.e., when green press left, when red press right) of a rectangle. Note that stimuli could vary on both dimensions of which only one was relevant. This way, congruent (i.e., a vertical green rectangle) and incongruent (i.e., a horizontal green rectangle) stimuli were created. Crucially, participants responded more slowly to incongruent stimuli, but only when both tasks were always presented at the same location. When the location on the screen was predictive of the task, the congruency effect disappeared, suggesting that the location started indicating the applicability of the task rules. Interestingly, this context-dependent task performance is often hypothesized to be a product of implicit, habitual learning, which on its turn is thought to require practice (Crump et al., 2008; Reuss et al., 2014). Based on these studies, one might postulate that context-dependent task performance could not be achieved on the basis of mere instructions (without any prior task experience). This would suggest that context-dependence requires actual experience. Yet, such a potential boundary condition of what can be achieved on the basis of mere verbal instructions remains to be tested. Importantly, earlier studies on the context-dependence of task sets only used paradigms in which the context was not instructed, and thus had to be learned through practice.

Second, similar to how automaticity was thought to result exclusively from practice but can actually be established by instructions (see above), we hypothesize that instructions can establish context-dependence too. Specifically, the above-mentioned instruction studies (e.g., Liefooghe et al., 2012) suggest that working memory is capable of maintaining and implementing stimulus-response mappings, but it remains unclear to what extent working

memory can also integrate contextual cues into task-set representations. If working memory can bind task-sets to contextual features, then environmental cues that match or mismatch these contextual features may control whether the contents of working memory interfere with performance. Therefore, the present study aims to be a first test of whether people can integrate contextual cues into their task-set representations on the basis of verbal instructions alone.

We adapted the procedure introduced by Liefooghe and colleagues (Liefooghe et al., 2012, 2013) to include a context manipulation (see Figure 1). As in previous work, the paradigm consisted of an inducer task and a diagnostic task. In a series of runs, each separate run started with the presentation of the instructions for the inducer task, which consisted of two highly frequent Dutch four-letter words (see the appendix) and their left versus right response button assignment (e.g., "if *lamp* press left, if *wolf* press right"). Participants were instructed to only respond to the word identity when the word was presented in green (i.e., inducer task). In the retention interval between the inducer task instructions and the actual inducer task, several trials of a diagnostic task were performed. Diagnostic trials were indicated by the fact that the four-letter words were presented in a white font type (as opposed to the green font inducer trials). While the instructions for the inducer task changed each run, the instructions for the diagnostic task were always the same: "When presented in upright font, press left; when presented in italic font, press right." Importantly, the identity of the words was irrelevant for the diagnostic task.

Using this paradigm, Liefooghe and colleagues (2012, 2013) observed congruency effects that resulted from the instructions of the inducer task during performance on the diagnostic task. For example, the word *lamp* presented in italic requires a right button response in the diagnostic task, but a left button response in the later-to-be-performed inducer task, thereby slowing down performance in the diagnostic task—even though the inducer instructions were formally irrelevant at that point in the run. As argued above, these findings suggest that merely instructed, but never executed, S-R mappings can automatically influence behavior. To investigate whether context instructions could modulate this effect, we added the instruction that the inducer task would be presented either above or below a central fixation cross. The diagnostic trials, however, were always randomly presented above or below the fixation cross. This design allowed us to investigate whether the instructed task context (i.e., stimulus location) would impact the automatic interference from merely instructed task rules. Specifically, if context-dependence can be instructed, we would expect to observe the impact of the instructed task-rules on diagnostic trials only at the location of the later-to-be-performed inducer trial.

Experiment 1

Method

Participants. Thirty-two students (range = 18–20 years, 29 women, 28 right-handed) took part in return for course credits or 10€. Although earlier studies on the context-sensitivity of trained task sets (e.g., Crump et al., 2006; Mayr & Bryck, 2005, 2007) or the instruction-based congruency effect (e.g., Liefooghe et al., 2012) all used around 14 to 18 subjects per experiment, we decided to test 32 subjects for our first experiment to ensure sufficient
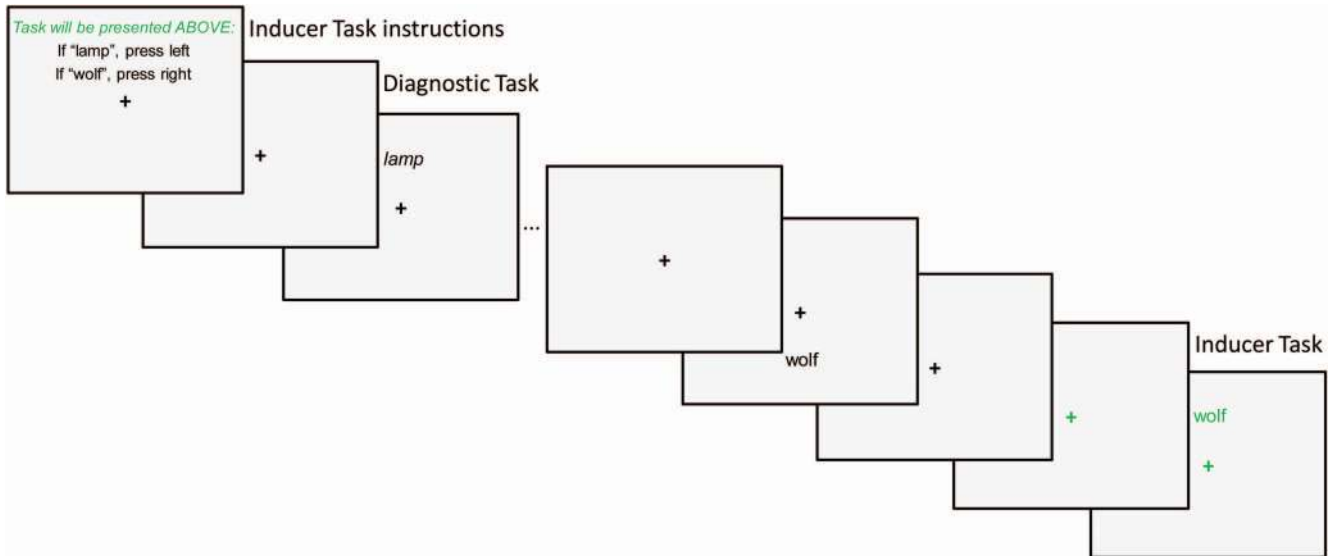
Figure 1. General paradigm and trial procedure. At the start of each run, participants were instructed about the inducer task which they had to perform whenever the word appeared in green. After 750 ms the first of four, eight, 12, or 16 trials of the diagnostic task appeared. The duration of the inter-trial-interval for the diagnostic task randomly varied between 500 and 1,000 ms and was indicated by the presence of a black fixation cross. During this task participants had to respond to the orientation of the font type, while the meaning of the word was irrelevant. The inducer task was preceded by a 500-ms presentation of a green fixation cross. In Experiment 2, the inducer task could appear at the alternative location in 12.5% of the trials, indicating that participants should not respond to the inducer task. In Experiment 3, 69 trials of the inducer task preceded the diagnostic task. In Experiment 4, the location of the inducer task was the same for all runs and was counterbalanced across participants. In Experiments 2 to 4 a rectangle surrounded each stimulus to further emphasize its location. See the online article for the color version of this figure.

power to detect context effects. For all of the hereafter presented experiments we recruited 20 subjects.

Stimuli and material. A list of 98 highly frequent four-letter Dutch nouns (see the appendix) were selected from the SUBTLEX-NL database (Keuleers, Brysbaert, & New, 2010). For each participant, a randomly determined set of 49 unique pairs of words was chosen, of which one was assigned to the practice run and the other 48 to the six blocks of experimental runs (blocks were interspersed with self-paced brakes). Specifically, eight pairs per block were randomly assigned to the eight different runs within a block, of which two runs counted either four, eight, 12, or 16 trials of the diagnostic task, which were presented in a random order.

The words were presented either above or below the centrally presented fixation cross, halfway between the fixation cross and the edge of the screen, in Arial 24-point font. The words in the diagnostic task could be presented either in italic or in upright font which determined the correct response (see above). Importantly, the location and congruency of the words were randomized but balanced across trials (all participants received an equal number of congruent and incongruent trials at either location, per run). The instructions for the inducer task always consisted of three lines and were similarly presented above or below the fixation cross. The location of these instructions corresponded to the location on which the stimulus for the inducer task would be presented. This location was determined randomly across runs. All instructions and stimuli were presented in white on a black background, except

for the words during the inducer task which were presented in green. Stimuli were presented on a CRT monitor located 60 cm away from the eyes using Tscope software (Steven, Lammertyn, Verbruggen, & Vandierendonk, 2006). The left and right response key were the letters F and J, respectively, on a standard QWERTY keyboard.

Procedure. Participants read the general instructions, after they were seated behind the computer in a dimly lit room and filled in the informed consent form. The instructions explained the general procedure as well as the specific response rules for the diagnostic task. These response rules were the same for all participants: Participants had to press left when the stimulus was presented in upright font and right when presented in italic. The instructions further emphasized to use the location information of the inducer task instructions to respond as fast as possible to the inducer task. After a first example run with 12 diagnostic trials, the participants read the instructions once more, and completed six blocks of eight runs.

The trial procedure is visualized in Figure 1. First, the inducer task instructions were presented on the screen above or below the fixation cross, depending on the inducer task location. The first line mentioned whether the task would be presented above or below the fixation cross with the location printed in uppercase letters (i.e., the Dutch translation of "The task will be presented ABOVE"), under which the next two lines indicated the run-specific response mapping (e.g., the Dutch translation of "if wolf, press right"). The order of the lines indicating the left or right

button response was randomized. These instructions remained on the screen for 20 s or until the participant pressed the spacebar. After the instructions and a 750-ms interval, the first stimulus of the diagnostic task appeared. Each stimulus of the diagnostic task was presented until response or for 2,000 ms. When the participant responded incorrectly, a red screen appeared for 200 ms. In between trials, there was a random intertrial interval that was between 500 and 1,000 ms long. Following the final diagnostic trial of each run, the same intertrial interval was presented after which the fixation cross turned green for another 500 ms. Finally, the inducer task stimulus appeared which was also presented in green and remained on screen until response or for 2,000 ms. A new run started 1,500 ms after the participant performed the inducer task. The fixation cross remained on the screen throughout the entire run.

Results

Three participants were excluded from further analyses because they performed at chance level on either the diagnostic task ($n = 2$), or the inducer task ($n = 1$). Mean accuracy for the remaining 29 participants was 92.3% ($SD = 6.0\%$) on the diagnostic task and 86.5% ($SD = 9.7\%$) on the inducer task. All further analyses focused on the diagnostic task only. Trials following an error, or trials that were part of a run on which the inducer task was performed inaccurately, were removed from analyses. The former is to avoid posterror (and postfeedback) slowdown effects in our data, whereas the latter is to exclude trials where we cannot guarantee that participants successfully implemented the inducer task. For the reaction time (RT) analysis, only RTs of correct responses that were slower than 200 ms were considered. As a result of these exclusion criteria, 24.1% of the diagnostic trials were removed for RT analyses. Median RT and mean accuracy results were analyzed statistically in repeated-measures analyses of variance (rANOVA) with two within-subject factors for congruency (congruent vs. incongruent) and context (same vs. different location as inducer task). To evaluate whether the following results depended on the experimenter's choice of using median RTs instead of the more conventional mean RTs, we also analyzed mean RT data using an outlier criterion of 2.5 SD (see also, Everaert et al., 2014; Theeuwes et al., 2014). Importantly, similar significance levels were obtained, which rendered the same statistical conclusions as below. All rANOVA and follow-up tests were two-tailed. To determine whether a nonsignificant finding could be considered support for the null hypothesis (i.e., the location relevance does not affect the automatic activation of task instructions), we also performed Bayesian analyses on null findings. Specifically, we computed the Bayes factor BF_{01}, which quantifies the evidence for the null hypothesis against the evidence for the alternative hypothesis. To this end, we used the open source statistical program JASP Version 7.1 (Love et al., 2015) and ran one-sided Bayesian tests (with Cauchy prior width = 1, as specified by Rouder et al., 2009).

RTs. Overall, there was a significant congruency effect, $F(1, 28) = 9.55$, $p < .005$. Moreover, a significant context-effect was observed, $F(1, 28) = 4.93$, $p < .05$, indicating a RT benefit when the diagnostic task occurred on the same, relative to the alternative, location as the later-to-be performed instructed inducer task. Importantly, despite this evidence that context was processed during the diagnostic task, the factors congruency and context did not interact, $F(1, 28) = 0.58$, $p = .452$, indicating that the congruency effect was not modulated by the context in which it occurred. As can be seen on Figure 2, the effect, if anything, was numerically larger at the alternative location.

Error rates. The error rates analysis only showed a significant congruency effect, $F(1, 28) = 16.551$, $p < .001$, indicating more accurate responses on congruent trials. The other effects did not reach significance, both $Fs(1, 28) < 1$.

Bayesian analyses. The BF_{01} for the interaction between congruency and context was 11.499 in the RT analyses (and 4.113 in the error analyses), suggesting that these data are 11.499 more likely to be observed under the null hypothesis. This is considered strong evidence for the null hypothesis that the automatic location-specific activation of task sets cannot be observed on the basis of mere instructions (Jeffreys, 1961).

Discussion

The results are in line with earlier observations showing that instructed, but never executed, S-R mappings can have an automatic impact on task performance (Liefooghe et al., 2012, 2013). Interestingly, the instructed inducer task context had an impact on diagnostic task performance, as evidenced by the faster RTs at the location where the inducer task was expected to occur. This suggests that participants' (spatial) attention was modulated by the context instruction. However, although participants clearly encoded the instructions about both the S-R mapping and task context, the results suggest that both types of information were not integrated. Specifically, the impact of the task-irrelevant instructed S-R rules was not modulated by location.

These results demonstrate that, unlike to what is the case for practiced S-R mappings (e.g., Cañadas et al., 2013; Crump et al., 2006, 2008; Crump & Logan, 2010; Mayr & Bryck, 2005, 2007; Reuss et al., 2014; Rubin & Koch, 2006; Schouppe et al., 2014), tasks cannot be triggered in a context-specific manner on the mere basis of instructions. Possibly, the integration between task sets and their context is something that requires practice and is not possible on the basis of mere instructions. This interpretation would be in line with the idea that the context-dependency of task
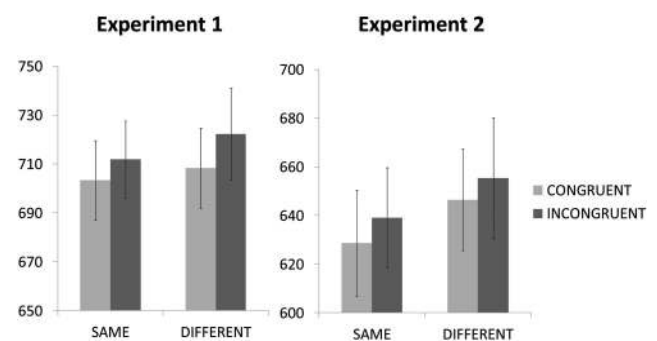

Figure 2. Mean median reaction times for Experiments 1 and 2. Reaction times are depicted for each congruency and context condition separately. As reported in the results sections, the congruency and context effects both reached significance, but there was no interaction. All error bars are ± 1 standard error of the mean.

sets is acquired through implicit learning mechanisms that require practice (Crump et al., 2008; Reuss et al., 2014). However, before drawing this conclusion we wanted to further probe the impact of instructed task contexts.

Experiment 2

Although the main effect of context in Experiment 1 seems to suggest that participants did implement the context information, this information was not necessary for accurate task performance. At best, remembering this information could speed up inducer task performance. Therefore, in our second experiment, we further stressed the role of task context by instructing that participants should withhold from responding to the inducer task, whenever it appeared at the uninstructed location. In line with this general instruction, we presented one in eighth of the inducer trials at the alternative location. Note that, here, our context manipulation is different from the other experiments, as well from previous studies, in that other context studies focused on the effects of contexts that were nonessential for accurate task completion. However, we reasoned that making context task relevant could increase its saliency and therefore increase the chances of observing a modulatory effect of context instructions.

Method

Participants.   Twenty students (range = 18–36 years, 14 women, 19 right-handed) took part in return for course credits or 10€.

Stimuli and material.   The stimuli and material were exactly the same as Experiment 1 (see Figure 1), except that the task stimuli were now surrounded by a rectangle that further emphasized the location of the stimulus. Centered around the stimulus, the rectangle was 60% of the screen wide, and 20% of the screen high. Similar to the word stimuli, the rectangle was white for the diagnostic task, and green for the inducer task.

Procedure.   The procedure was similar to Experiment 1. However, the participants were now instructed that the inducer stimulus would not always appear at its instructed location. Consistent with this instruction, and in contrast to the other experiments, one out of eight inducer trials (randomly determined) did not appear at its instructed location, but at the alternative location instead. Furthermore, whenever the inducer stimulus appeared at the alternative location, participants were explicitly instructed to withhold their response. This way, the location information was also necessary to perform the task accurately.

Results

Four participants were excluded from further analysis, because they either performed at chance level on the inducer task ($n = 1$), or ignored the instruction to withhold from responding when the inducer task was at the wrong location ($n = 3$). Mean accuracy for the remaining 16 participants was 94.7% ($SD = 2.6\%$) on the diagnostic task and 90.2% ($SD = 6.6\%$) on the inducer task. Moreover, the remaining 16 participants were able to successfully withhold their response when the inducer task was presented at the wrong location (inhibition accuracy = 85.8%; $SD = 13.3\%$). The data preparation procedure was the same as for Experiment 1.

Similar to Experiment 1, we only analyzed trials that were part of a run on which the inducer task was performed accurately. Runs where the inducer task was presented on the wrong location (and inducer task performance could thus not be evaluated), were also excluded from the analyses. As a result of these exclusion criteria, 27.9% of the diagnostic trials were removed for RT analyses.

RTs.   Similar to Experiment 1, we observed significant main effects of congruency, $F(1, 15) = 5.46$, $p < .05$, and context, $F(1, 15) = 16.01$, $p < .005$. Importantly, the factors congruency and context again did not interact, $F(1, 15) = 0.02$, $p = .892$, indicating that also in Experiment 2 the congruency effect was not modulated by the context in which it occurred (see Figure 2).

Error rates.   None of the effects in the error rates analysis reached significance, all $Fs(1, 15) < 1$.

Bayesian analyses.   The BF_{01} for the interaction between congruency and context was 4.747 in the RT analyses (and 2.151 in the error analyses), suggesting that these data are 4.747 more likely to be observed under the null hypothesis. This is considered substantial evidence for the null hypothesis (Jeffreys, 1961). Pooling the data of Experiment 1 and 2 together, we observed a BF_{01} of 12.048 in the RT analyses (and 2.993 in the error analyses) for the null hypothesis that the automatic location-specific activation of task sets cannot be observed on the basis of mere instructions.

Discussion

The findings of Experiment 2 replicated those of Experiment 1. Again, we observed that both the instructed task rule and task context from the inducer task had an automatic impact on the RTs of the task-irrelevant diagnostic task, but did not interact. Together with the results of Experiment 1, our findings suggest that participants did not integrate the task rules with their task context, as the automatic impact of task rules was independent of the context in which they appeared. These results seem to contrast with what has consistently been found with trained task sets, suggesting that the context-sensitivity of task sets requires practice. If practice is indeed a crucial ingredient for observing a context-dependent automatic activation of task sets, we should be able to observe a context-modulated congruency effect after the inducer task has been practiced in its context. This would confirm that the absence of context-specific effects in Experiments 1 and 2 is not simply due to some unique feature of our procedure.

Experiment 3

In Experiment 3, we tested an alternative version of our paradigm in which the diagnostic trials were preceded by at least 69 trials of inducer task trials. Based on the prior literature and the idea that training allows for context-specific impact of S-R associations, we expected to observe a context-specific congruency effect in the present experiment.

Method

Participants.   Twenty students (range = 18–29 years, 19 women, 19 right-handed) took part in return for course credits or 10€.

Stimuli and material.   The stimuli and material were exactly the same as Experiment 2.

Procedure. The instructions and trial procedure were highly similar to Experiment 1 (see Figure 1). However, the diagnostic task was now preceded by the inducer task. Each block consisted of only a single run, in which one randomly chosen (without replacement) pair of words was used. The practice run consisted of 16 trials of the inducer task preceding four trials of the diagnostic task. In the next 10 blocks, each instruction was first followed by 69 trials of the inducer task (note that in the other experiments the diagnostic task always preceded the inducer task). After these 69 trials of the inducer task, we presented the first 10 diagnostic trials, followed by another 13 inducer task trials, another six diagnostic trials, and a final four inducer task trials, summing up to a total of 102 trials per block. This alternating sequence, which always ended with the inducer task, was implemented to ensure that the participants kept the S-R mappings of the inducer task active (Meiran, Cole, & Braver, 2012). Moreover, the trial order was fixed to ensure the comparability across the short amount of runs (i.e., one per block). This way, each block counted 16 diagnostic trials which were preceded by 69 or more trials of inducer task training.

Results

The data of all participants were included in the analysis. Mean accuracy was 92.1% ($SD = 4.1\%$) on the diagnostic task and 96.2% ($SD = 1.7\%$) on the inducer task. The data preparation procedure was the same as for Experiment 1. As a result of these exclusion criteria, 24.0% of the diagnostic trials were removed for RT analyses.

RTs. Again, we observed a main effect of context, $F(1, 19) = 17.52$, $p < .01$, and a main effect of congruency, $F(1, 19) = 5.61$, $p = .029$. Moreover, the factors congruency and context interacted significantly, $F(1, 19) = 5.18$, $p < .05$. Planned comparisons indicated that the congruency effect was only significant when the location was the same as that of the inducer task, $t(19) = 3.069$, $p < .05$, but not when the location was different, $t(19) = 0.476$, $p = .640$ (see Figure 3).

Error rates. The error rates analysis also showed a significant interaction, $F(1, 19) = 4.778$, $p < .05$, again showing a significant

congruency effect on the same location, $t(19) = 2.517$, $p < .05$, but not on the different location, $t(19) = 0.172$, $p = .865$. The main effects did not reach significance, both $Fs(1, 19) < 2.36$, $p > .141$.

Bayesian analyses. The BF_{10} (the evidence for the alternative hypothesis against the evidence for the null hypothesis) for the interaction between congruency and context was 3.054 in the RT analyses (and 2.605 in the error analyses), suggesting that these data are 3.054 more likely to be observed under the alternative hypothesis. This is considered moderate evidence for the alternative hypothesis (Jeffreys, 1961), and merely anecdotal evidence in the error rates. However, note that this is an underestimation of the effect as this Bayesian test does not take into account the abundant previous evidence for the automatic location-specific activation of trained task sets (e.g., Crump et al., 2006, 2008; Crump & Logan, 2010; King, Korb, & Egner, 2012; Mayr & Bryck, 2007). Still, we decided to test the effect of training on the automatic location-specific activation of trained task sets once more in Experiment 4.

Discussion

The results of Experiment 3 suggest that a small amount of training on the inducer task (69 or more trials) can be sufficient to induce a context-specific congruency effect. These findings fit well with prior work demonstrating the context-sensitivity of task representations (e.g., Cañadas et al., 2013; Crump et al., 2006, 2008; Crump & Logan, 2010; Mayr & Bryck, 2005, 2007; Reuss et al., 2014; Rubin & Koch, 2006; Schouppe et al., 2014), and further suggest that although instructions are insufficient to bind task rules to their context, practice might be necessary to make this process possible.

Given this finding, we reasoned that although people appear to bind S-R associations to specific contexts only after practice, people might still be able to form such representations with merely instructed S-R mappings when they practiced other S-R mappings in the crucial context. That is, perhaps a joint effect of S-R instructions and context instructions is possible without prior practice of the S-R mappings in that context, provided that participants have previously learned that the context is relevant. Specifically, in Experiments 1 and 2 the context varied on a run-by-run basis. The context, much like the S-R instructions, could differ from run to run. However, if we would keep the inducer task context constant, participants might be able to learn this context-relevance in a mode that allows them to facilitate integration with the merely instructed S-R mappings.

Figure 3. Mean median RTs for Experiments 3 and 4. Reaction times are depicted for each congruency and context condition separately. As reported in the results sections, both experiments showed a significant interaction between congruency and context, demonstrating that when trained on both the task and context (Experiment 3), or the context only (Experiment 4), a context-specific congruency effect can be observed. All error bars are ± 1 standard error of the mean (SEM).

Experiment 4

In this final experiment, we decided to use the same design as Experiment 1, but keep the context (i.e., inducer task location) constant throughout the entire experiment (i.e., for each participant the inducer task was always presented either above or below, counterbalanced across participants). If training on the context alone allows for context-dependent instructed task representations, we should observe a context-specific congruency effect in the diagnostic task, especially later on in the experiment (when context could be sufficiently trained). As such, we explored the impact of context for each half of the experiment separately.
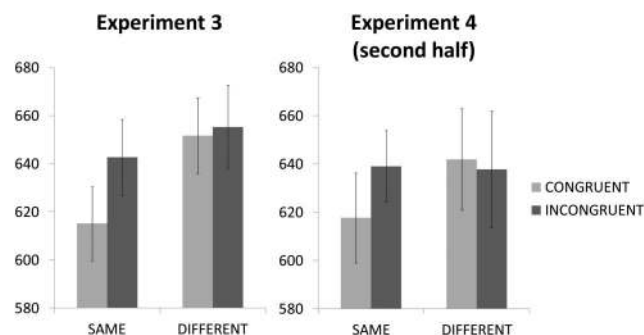
Method

Participants. Twenty students (range = 18–37 years, 17 women, 16 right-handed) took part in return for course credits or 10€.

Stimuli and material. The stimuli and material were exactly the same as Experiment 2, that is, rectangles were printed around the word stimuli.

Procedure. The procedure was the same as in Experiment 1 (see Figure 1). Only now, the location of the inducer task was always the same for all runs (counterbalanced across participants).

Results

Four participants were excluded from further analysis, because they performed at chance level on the diagnostic task ($n = 1$), on the inducer task ($n = 2$), or both ($n = 1$). Mean accuracy for the remaining 16 participants was 92.0% ($SD = 4.8\%$) on the diagnostic task and 89.4% ($SD = 6.9\%$) on the inducer task. The data preparation procedure was the same as for Experiment 1. As a result of these exclusion criteria, 22.3% of the diagnostic trials were removed for RT analyses. However, different to the previous experiments, we included the within-subject factor experiment half (first vs. second half) in our analysis, because we wanted to explore the hypothesis that only after training, context exhibits a mediating effect on the automatic interference of instructed S-R mappings.

RTs. Similar to Experiment 3, we observed a significant main effect of context, $F(1, 15) = 8.13$, $p < .05$, but not congruency, $F(1, 15) < 1$. Importantly, however, congruency significantly interacted with context and experiment half, $F(1, 15) = 11.57$, $p < .01$. All other effects did not reach significance, all $Fs(1,15) < 1$. When looking at both experiment halves separately, congruency did not interact with context in the first half, $F(1, 15) = 2.21$, $p = .158$, but did interact with context in the second half, $F(1, 15) = 8.49$, $p < .05$. As can be seen in Figure 3, planned comparisons indicated that the congruency effect was significant only when the location was the same as that of the inducer task, $t(19) = 3.189$, $p < .05$, but not when the location was different, $t(19) < 1$.

Error rates. The error rates showed a significant main effect of congruency, $F(1, 15) = 7.79$, $p < .05$, and a marginally significant effect of context, $F(1, 15) = 3.81$, $p = .07$, indicating more accurate responses on congruent or same context trial conditions. All other effects did not reach significance, all $Fs(1, 15) < 1.38$, $ps > .259$.

Bayesian analyses. The BF_{10} for the interaction between congruency and context (in the second half) was 10.276 in the RT analyses (and 0.118 in the error analyses), suggesting that these data are 10.276 more likely to be observed under the alternative hypothesis. This is considered strong evidence for the alternative hypothesis (Jeffreys, 1961), and evidence for a substantial null effect in the error rates. Taking the data of Experiment 3 and 4 together, we observed a BF_{10} of 50.788 in the RT analyses (and 2.932 in the error analyses) strongly supporting the hypothesis that a trained context can modulate the automatic activation of task sets (in the RTs only).

Discussion

Similar to the results of Experiment 3, Experiment 4 showed a context-specific congruency effect. Importantly, this effect developed only in the second phase of the experiment, after the participants had a sufficient amount of training with the context of the inducer task. These findings suggest that training on the context can be sufficient to evoke context-specific instruction-based task representations.

General Discussion

In this study, we explored the limits of what can be achieved with verbal instructions by testing whether instructions can automatically influence behavior in a context-specific way. To this end, we used a recent procedure developed by Liefooghe and colleagues (2012, 2013) that probes the automatic activation of instructed knowledge, and paired it with an instructed context manipulation by indicating the conditions under which the task would have to be performed (i.e., when an inducer stimulus was presented at a certain location on the screen). Within the boundaries of our paradigm, we conclude that whereas instructions can both (a) bias (spatial) attention toward the relevant context and (b) allow for the automatic activation of responses, they are insufficient to allow for a context-dependent automatic activation of responses. Conversely, actual practice on either the task in its context, or the context alone, was successful in inducing context-dependent congruency effects.

The current study offers important new insights in the possibilities and limitations of instruction-based learning. In line with previous observations, we demonstrated how instructed S-R rules can result in the automatic activation of responses upon stimulus presentation, even when irrelevant (Liefooghe et al., 2012, 2013; Meiran et al., 2015a; Theeuwes et al., 2014; Wenke et al., 2007). This observation is in line with theories on the prepared reflex (Exner, 1879; Woodworth, 1938; Hommel, 2000), suggesting that instructions allows us to prepare a state of high readiness in which we can execute a response to a stimulus with very little additional effort. In their recent review, Meiran, Cole, and Braver (2012) suggested that this reflexive nature of instruction-based intentions is also what might make them rigid, relative to practiced plans. Similarly, the present findings seem to demonstrate how instructions per se can be a very powerful tool to establish S-R associations, whereas, at the same time, this powerful impact cannot be easily fine-tuned by contextual aspects that are part of the same instructions. Concordant with this observation, Liefooghe and colleagues (2012) demonstrated how instructed S-R associations impact task performance even when the irrelevant task was assigned to different response keys, further suggesting that instructed S-R associations have an unspecific impact on behavior.

Interestingly, also instructed task location seemed to have a general effect on performance. This effect can be considered surprising given that, in our task, the locations on which two different tasks could be presented generated faster RTs relative to locations on which only a single task was possible (i.e., the diagnostic task). However, although the instructed task location was home to two different tasks, these two tasks were always presented in a fixed order (diagnostic task first, instructed task second) and clearly cued (i.e., the fixation cross turned green 500 ms before the instructed task target word), minimizing possible

slowdowns due to task confusion. Instead, we believe that the anticipation of the instructed task location might have biased participants' spatial attention toward that location, even before the actual occurrence of the instructed task, and therefore facilitated any task occurring on that location. Still, a full understanding of this effect will require further studies.

The inability to establish these context-dependent S-R mappings via instructions, could be due to the very limited resources in working memory (Oberauer, 2009; as also reviewed by Meiran et al., 2012). In his working memory model, Oberauer (2009) dissociates between three different components of working memory: (a) the activated long-term memory (ALTM), (b) a component responsible for building new structural representations (called the region of direct access, RoDA), and (c) a selection mechanism (focus of attention). Importantly, the RoDA is thought to be highly context-sensitive, while the ALTM is not (Oberauer, 2001). The retaining of a new set of instructions (e.g., if *wolf*, press right) requires both the ALTM (e.g., long term memory traces of right button presses, memories of the word *wolf*) and the RoDA (e.g., the new association between the word 'wolf' and a right button press), whereas the activation of trained instructions relies mostly on the ALTM alone, thereby freeing up the context-sensitive RoDA. Given the severely limited resources of the RoDA, the present results could be explained by its unavailability to form context-dependent associations when retaining new instructions. In contrast, when these new instructions are trained and have had the time to form LTM traces, the RoDA is now available to benefit from its context-sensitivity. Alternatively, it is possible that people process instructed location information and S-R rules in different processing stages or systems that do not yet allow a full integration or interaction. For example, the retention of a spatial location and a verbal S-R instruction could rely on different neural structures that each suffer their own capacity limitations (e.g., a declarative vs. procedural working memory; Oberauer, 2009). A future experiment in which context and S-R information are more similar in terms of modality could determine whether this is a criterion to achieve context-specific effects of task representations on the basis of instructions alone.

This powerful but inexact nature of instructed S-R mappings that our findings seem to hint at is also consistent with recent observations by Theeuwes, De Houwer, Eder, and Liefooghe (2015). In their study, they investigated whether the retention of response-effect contingencies (in contrast to the S-R contingencies investigated in previous studies) would also trigger automatic response activation upon stimulus presentation (i.e., the effect) during the diagnostic task. Their results demonstrated that this type of response activation could indeed be observed, despite the explicit instructions about the unidirectional nature of the relation. Inducer task performance indicated that participants clearly remembered both the relation and its direction (i.e., response-effect), while diagnostic task performance showed how the presentation of the effect was sufficient to induce the response. Concordant to our findings, it seems like the instructed knowledge was retained in such a manner that both types of information (i.e., the association and its unidirectional relation in Theeuwes et al., 2015; or the S-R associations and their context in our study) were stored separately.

We are the first to have investigated the learning of task contexts via instructions. While our results offer important initial insights, the robustness and generalizability of our findings remain to be determined. First, we used a location manipulation because it is the most popular and most widely used context manipulation. However, other context manipulations, such as changing the font (Bugg, Jacoby, & Toth, 2008) or surrounding the task stimuli by different shapes (Crump et al., 2006; Schouppe et al., 2014), should be explored as well. Second, the effects of instructions are, naturally, highly dependent on the precise wording of the instructions. Perhaps more persuasive or motivating instructions might be able to induce context-dependent task-rule effects after all. This would mean that in our experiment, participants simply did not bother to integrate the location and S-R instructions. Third, future studies should establish whether other types of second-order relations (e.g., if A than B, but only if C) can be implemented in an integrated manner. As suggested by an anonymous reviewer, one specific idea for a future study could be to instruct participants that a set of S-R mappings on one location should be reversed on the other (e.g., "if *lamp* press left, if *wolf* press right", when presented above, but "if *lamp* press right, if *wolf* press left" when presented below). This manipulation is rather different from previous studies into the context-specificity of task control, but could provide a strong test of whether people are actually capable of implementing this type of second-order task representations.

Our data seem to suggest that only after practice, we can observe context-specific automatic effects of instructed task rules. Notably, this interaction appears to be driven by RT differences on congruent trials. This pattern can indicate either a RT increase for congruent trials at the irrelevant context, or a RT decrease for congruent trials at the relevant context (or both). The latter would be most in line with the idea that, only after practice, task representations become bound to their relevant context. After practice, the relevant context will further prime the task representations, whereas the irrelevant context will not (or to a lesser degree). However, when considering the former possibility that RT increased for congruent trials at the irrelevant context, an alternative interpretation for the observed interaction would be that the automatic activation of the inducer task is still present at both locations, but now gets more easily suppressed by the irrelevant context. Following this interpretation, the irrelevant context further inhibits task relevant representations after practice. Although our study indicates that practice plays an important role in bringing about the context-specific interference effect from instructed task sets, future studies should determine whether these effects are the result of context-specific activation or inhibition mechanisms, or a combination of the two.

In sum, the present study builds on the findings of a relatively new but growing research field (e.g., Cole et al., 2013) that focuses on the possibilities and limitations of learning via instructions. Although instructions can be a powerful tool to quickly share and acquire new action plans or knowledge about the world, our findings suggest that learning via instructions has its limitations. Instructed S-R mappings can result in the automatic activation of responses, but this activation will not be context-dependent. However, this limitation might be overcome via a small amount of practice of the S-R mapping within a specific context or via practice with the relevant context only.

References

Abrahamse, E., Braem, S., Notebaert, W., & Verguts, T. (2016). Grounding cognitive control in associative learning. *Psychological Bulletin, 142,* 693–728. http://dx.doi.org/10.1037/bul0000047

Abrahamse, E. L., & Verwey, W. B. (2008). Context dependent learning in the serial RT task. *Psychological Research, 72,* 397–404. http://dx.doi.org/10.1007/s00426-007-0123-5

Bardi, L., Bundt, C., Notebaert, W., & Brass, M. (2015). Eliminating mirror responses by instructions. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 70,* 128–136. http://dx.doi.org/10.1016/j.cortex.2015.04.018

Braem, S., Abrahamse, E. L., Duthoo, W., & Notebaert, W. (2014). What determines the specificity of conflict adaptation? A review, critical analysis, and proposed synthesis. *Frontiers in Psychology, 5,* 1134. http://dx.doi.org/10.3389/fpsyg.2014.01134

Bugg, J. M., Jacoby, L. L., & Toth, J. P. (2008). Multiple levels of control in the Stroop task. *Memory & Cognition, 36,* 1484–1494. http://dx.doi.org/10.3758/MC.36.8.1484

Cañadas, E., Rodríguez-Bailón, R., Milliken, B., & Lupiáñez, J. (2013). Social categories as a context for the allocation of attentional control. *Journal of Experimental Psychology: General, 142,* 934–943. http://dx.doi.org/10.1037/a0029794

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36,* 28–71. http://dx.doi.org/10.1006/cogp.1998.0681

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,* 1–22. http://dx.doi.org/10.3758/s13415-012-0125-7

Crump, M. J., Gong, Z., & Milliken, B. (2006). The context-specific proportion congruent Stroop effect: Location as a contextual cue. *Psychonomic Bulletin & Review, 13,* 316–321. http://dx.doi.org/10.3758/BF03193850

Crump, M. J., & Logan, G. D. (2010). Contextual control over task-set retrieval. *Attention, Perception, & Psychophysics, 72,* 2047–2053. http://dx.doi.org/10.3758/BF03196681

Crump, M. J., Vaquero, J. M., & Milliken, B. (2008). Context-specific learning and control: The roles of awareness, task relevance, and relative salience. *Consciousness and Cognition: An International Journal, 17,* 22–36. http://dx.doi.org/10.1016/j.concog.2007.01.004

Deacon, T. W. (1997). *The symbolic species.* New York, NY: Norton.

Everaert, T., Theeuwes, M., Liefooghe, B., & De Houwer, J. (2014). Automatic motor activation by mere instruction. *Cognitive, Affective & Behavioral Neuroscience, 14,* 1300–1309. http://dx.doi.org/10.3758/s13415-014-0294-7

Exner, S. (1879). Physiologie der Grosshirnrinde. In L. Hermann (Ed.), *Handbuch der Physiologie* (Vol. 2, Pt.2; pp. 189–350). Leipzig, Germany: Vogel.

Hommel, B. (2000). The prepared reflex: Automaticity and control in stimulus response translation. In S. Monsell & J. Driver (Eds.), *Attention and performance: Control of cognitive processes* (pp. 247–273). Cambridge, MA: MIT Press.

Jeffreys, H. (1961). *Theory of probability.* Oxford, UK: Oxford University Press.

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods, 42,* 643–650. http://dx.doi.org/10.3758/BRM.42.3.643

King, J. A., Korb, F. M., & Egner, T. (2012). Priming of control: Implicit contextual cuing of top-down attentional set. *The Journal of Neuroscience, 32,* 8192–8200. http://dx.doi.org/10.1523/JNEUROSCI.0934-12.2012

Liefooghe, B., De Houwer, J., & Wenke, D. (2013). Instruction-based response activation depends on task preparation. *Psychonomic Bulletin & Review, 20,* 481–487. http://dx.doi.org/10.3758/s13423-013-0374-7

Liefooghe, B., Wenke, D., & De Houwer, J. (2012). Instruction-based task-rule congruency effects. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, 38, 1325–1335. http://dx.doi.org/10.1037/a0028148

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., . . . Wagenmakers, E.-J. (2015). *JASP* (Version 0.7) [Computer software]. Amsterdam, the Netherlands: JASP.

Mayr, U., & Bryck, R. L. (2005). Sticky rules: Integration between abstract rules and specific actions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 337–350. http://dx.doi.org/10.1037/0278-7393.31.2.337

Mayr, U., & Bryck, R. L. (2007). Outsourcing control to the environment: Effects of stimulus/response locations on task selection. *Psychological Research, 71,* 107–116. http://dx.doi.org/10.1007/s00426-005-0039-x

Meiran, N., Cole, M. W., & Braver, T. S. (2012). When planning results in loss of control: Intention-based reflexivity and working-memory. *Frontiers in Human Neuroscience, 6,* 104. http://dx.doi.org/10.3389/fnhum.2012.00104

Meiran, N., Pereg, M., Kessler, Y., Cole, M. W., & Braver, T. S. (2015a). The power of instructions: Proactive configuration of stimulus-response translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41,* 768–786. http://dx.doi.org/10.1037/xlm0000063

Meiran, N., Pereg, M., Kessler, Y., Cole, M. W., & Braver, T. S. (2015b). Reflexive activation of newly instructed stimulus-response rules: Evidence from lateralized readiness potentials in no-go trials. *Cognitive, Affective & Behavioral Neuroscience, 15,* 365–373. http://dx.doi.org/10.3758/s13415-014-0321-8

Oberauer, K. (2001). Removing irrelevant information from working memory: A cognitive aging study with the modified Sternberg task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 948–957. http://dx.doi.org/10.1037/0278-7393.27.4.948

Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation, 51,* 45–100. http://dx.doi.org/10.1016/S0079-7421(09)51002-X

Reuss, H., Desender, K., Kiesel, A., & Kunde, W. (2014). Unconscious conflicts in unconscious contexts: The role of awareness and timing in flexible conflict adaptation. *Journal of Experimental Psychology: General, 143,* 1701–1718. http://dx.doi.org/10.1037/a0036437

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16,* 225–237. http://dx.doi.org/10.3758/PBR.16.2.225

Rubin, O., & Koch, I. (2006). Exogenous influences on task set activation in task switching. *The Quarterly Journal of Experimental Psychology, 59,* 1033–1046. http://dx.doi.org/10.1080/02724980543000105

Ruitenberg, M. F., Abrahamse, E. L., De Kleine, E., & Verwey, W. B. (2012). Context-dependent motor skill: Perceptual processing in memory-based sequence production. *Experimental Brain Research, 222*(1–2), 31–40. http://dx.doi.org/10.1007/s00221-012-3193-6

Schouppe, N., de Ferrerre, E., Van Opstal, F., Braem, S., & Notebaert, W. (2014). Conscious and unconscious context-specific cognitive control. *Frontiers in Psychology, 5,* 539. http://dx.doi.org/10.3389/fpsyg.2014.00539

Steven, M., Lammertyn, J., Verbruggen, F., & Vandierendonck, A. (2006). Tscope: A C library for programming cognitive experiments on the MS windows platform. *Behavior Research Methods, 38,* 280–286. http://dx.doi.org/10.3758/BF03192779

Theeuwes, M., De Houwer, J., Eder, A., & Liefooghe, B. (2015). Congruency effects on the basis of instructed response-effect contingencies. *Acta Psychologica, 158,* 43–50. http://dx.doi.org/10.1016/j.actpsy.2015.04.002

Theeuwes, M., Liefooghe, B., & De Houwer, J. (2014). Eliminating the Simon effect by instruction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 1470–1480. http://dx.doi.org/10.1037/a0036913

Wenke, D., De Houwer, J., De Winne, J., & Liefooghe, B. (2015). Learning through instructions vs. learning through practice: Flanker congruency effects from instructed and applied SR mappings. *Psychological Research, 79,* 899–912. http://dx.doi.org/10.1007/s00426-014-0621-1

Wenke, D., Gaschler, R., & Nattkemper, D. (2007). Instruction-induced feature binding. *Psychological Research, 71,* 92–106. http://dx.doi.org/10.1007/s00426-005-0038-y

Woodworth, R. S. (1938). *Experimental psychology.* New York, NY: Holt, Rinehart and Winston.

Appendix

The Highly Frequent Four-Letter Dutch Words Used for the Present Study (English Translations Are Presented in Parentheses)

huis (house)	auto (car)	stad (city)	deur (door)	hand (hand)	hart (heart)	hond (dog)
baby (baby)	boek (book)	vuur (fire)	boot (boat)	bank (bank)	kerk (church)	raam (window)
jurk (dress)	been (leg)	slot (lock)	ring (ring)	trap (stairs)	boom (tree)	voet (foot)
rook (smoke)	fles (bottle)	brug (bridge)	ster (star)	maan (moon)	tent (tent)	doos (box)
blik (can)	hoed (hat)	graf (grave)	berg (mountain)	neus (nose)	muur (wall)	wijn (wine)
glas (glass)	kast (closet)	kist (case)	bord (plate)	heks (witch)	touw (rope)	beer (bear)
bril (glasses)	klok (clock)	hout (wood)	kaas (cheese)	wolf (wolf)	tank (tank)	vlag (flag)
stok (stick)	riem (belt)	lamp (lamp)	pijp (pipe)	kooi (cage)	haak (hook)	hemd (shirt)
roos (rose)	trui (sweater)	blok (block)	ezel (donkey)	nest (nest)	blad (leave)	muis (mouse)
helm (helmet)	munt (coin)	duim (thumb)	ober (waiter)	haai (shark)	bijl (axe)	worm (worm)
eend (duck)	tand (tooth)	geit (goat)	vest (vest)	spin (spider)	hals (neck)	teen (toe)
pijl (arrow)	wiel (wheel)	poot (paw)	pomp (pump)	hert (deer)	wolk (cloud)	duif (dove)
vork (fork)	vaas (vase)	rits (zipper)	mand (basket)	tang (pliers)	haan (rooster)	slee (sledge)
krab (crab)	noot (nut)	zaag (saw)	veer (feather)	hooi (hay)	kruk (stool)	mier (ant)