Original article

# Attentional prioritization reconfigures novel instructions into action-oriented task sets

Carlos González-García[a,*], Silvia Formica[a], Baptist Liefooghe[b], Marcel Brass[a]

[a] *Department of Experimental Psychology, Ghent University, Belgium*
[b] *Department of Experimental-Clinical & Health Psychology, Ghent University, Belgium*

## ABSTRACT

An astonishing aspect of human cognitive flexibility concerns the ability to efficiently convert complex symbolic instructions into novel behaviors. In such ability, the rapid transformation of relevant content into action plans is particularly crucial as it allows for reflexive, automatic-like execution of merely instructed task sets. However, little is known about the mechanisms that give rise to this transformation. In the current study, we test the hypothesis that novel instructions held in working memory are reformatted into action-oriented representations when selective attention prioritizes their content. To do so, we devised a paradigm in which participants first encoded 4 S-R mappings and later, a retro-cue selected two of them. We first found that participants can benefit from retro-cues during the implementation of novel task-sets. Then, across two preregistered experiments, we observed that cued mappings (but not uncued ones) induced intention-based reflexivity, suggesting that only these entered an action-oriented state. Altogether, our results reveal that selective attention prioritizes relevant novel instructed content, playing an important role in its prospective reformatting into an action-bound task set.

## 1. Introduction

The ability to rapidly learn new behaviors based on instructions is a distinctive aspect that separates humans from other species (Cole, Laurent, & Stocco, 2013), allowing us to achieve some of the most advanced forms of collaborative behavior. Consider the interaction between air traffic controllers and pilots, in which a mutual understanding of complex symbolic instructions is required to ensure successful functioning of congested air spaces. In such a situation, efficient performance depends on several steps. At least, the content and meaning of the instructions have to be encoded first, then rapidly transformed into action codes, and finally executed, all in a matter of seconds. The transformation from encoded content to action plans, a process often called *proceduralization* (Brass, Liefooghe, Braem, & De Houwer, 2017), is particularly crucial, as most of what makes instructions following unique (e.g. speed and reflexivity) critically depends on it. However, given the challenging nature of its computations, how this transformation takes place is currently unknown (Brass et al., 2017).

In the last years, significant effort has been devoted to characterize the phenomenon of Intention-Based Reflexivity (IBR; Braem, Liefooghe, De Houwer, Brass, & Abrahamse, 2017; Liefooghe & De Houwer, 2018; Liefooghe, De Houwer, & Wenke, 2013; Liefooghe, Wenke, & De Houwer, 2012; Meiran, Cole, & Braver, 2012; Meiran, Pereg, Kessler,

Cole, & Braver, 2015b). The IBR effect refers to the reflexive activation of responses on the basis of merely instructed stimulus-response (S-R) mappings, even when these are task-irrelevant and they have not been overtly executed before (Cohen-Kdoshay & Meiran, 2007; Liefooghe et al., 2012; Liefooghe et al., 2013; Meiran et al., 2015b; Wenke, Gaschler, Nattkemper, & Frensch, 2009). IBR also impacts brain activity, as revealed by electroencephalography studies showing that novel S-R associations induce reflexive motor activation (Everaert, Theeuwes, Liefooghe, & De Houwer, 2014; Meiran, Pereg, Kessler, Cole, & Braver, 2015a). Altogether, these findings suggest that the intention to execute a recently encoded instruction configures this instruction in an action-bound format that is activated in an almost reflexive manner when the associated stimulus is encountered again. However, several studies have shown that IBR effects depend at least partially on strategic processes (Whitehead & Egner, 2018). For instance, IBR primarily takes place when participants intend to implement upcoming instructions, but not when they prepare to simply memorize the contents of S-R mappings (Liefooghe et al., 2012; but see Liefooghe & De Houwer, 2018), and the extent of this effect depends on the degree of proactive task preparation (Liefooghe et al., 2013; Meiran et al., 2015a, 2015b, Experiment 4; Wenke et al., 2009). Neuroimaging studies have also shown that when preparing to implement, in contrast to just memorize, different brain regions coordinate to bridge novel stimuli and responses

(Demanet et al., 2016; González-García, Arco, Palenciano, Ramírez, & Ruz, 2017; Hartstra, Kühn, Verguts, & Brass, 2011; Palenciano, González-García, Arco, & Ruz, 2018), and that the intention to implement transforms the underlying neural code (Bourguignon, Braem, Hartstra, De Houwer, & Brass, 2018; Muhle-Karbe, Duncan, De Baene, Mitchell, & Brass, 2017). This dissociation between "knowing" (retaining task knowledge) and "doing" (successfully executing an instruction) is further supported by goal-neglect effects in patients and healthy population (Bhandari & Duncan, 2014; Duncan, Emslie, Williams, Johnson, & Freer, 1996). Altogether, these findings suggest that preparing to implement a novel instruction constitutes a separate neurocognitive state that, beyond just maintaining task knowledge in a declarative-like state, activates procedural (action-oriented) representations (Brass et al., 2017; Wenke et al., 2009), thus leading to IBR effects. Still, the precise mechanisms underlying the transformation of encoded information into a procedural state remain unexplored.

In parallel, in the neighboring field of working memory (WM), theoretical models and experimental results suggest that WM systems operate under two functional states (Oberauer, 2009), similar to the ones proposed to underlie instructions following: 1) information held in mind without an action plan (i.e. declarative), and 2) action-oriented prioritized content (i.e. procedural). Recent work in this field has put forward some suggestive ideas about how information can be reformatted from declarative to procedural WM. Much of this work has benefitted from the use of retro-cues, that is, a cue that selects retrospectively a subset of items already encoded in WM (Souza & Oberauer, 2016). Importantly, in contrast with pre-cues (which appear before the occurrence of relevant stimuli), retro-cues do not bias the detection nor the initial (e.g. perceptual) processing of cued information and therefore operate exclusively on WM contents. These accounts propose that attentional selection (for instance, elicited by retro-cues) might work not only as a prioritization tool that leverages the capacity limits of WM systems, but also as a bridge between WM states. Specifically, attention directed towards specific WM content by retro-cues would not only select relevant declarative representations but also reconfigure them into a *behavior-optimized* format (Myers, Chekroud, Stokes, & Nobre, 2018; Myers, Stokes, & Nobre, 2017). Given the resemblance of this process to task set formation we reasoned that, within this framework, the representation of novel task sets in an action-oriented format could be conceptualized as the *behavior-optimized* state that allows for the successful execution of instructions. In this regard, the fact that proceduralization seems to depend on capacity-limited WM systems (Brass et al., 2017; Liefooghe et al., 2012) that benefit from prioritization mechanisms (Oberauer, 2013) hints a particularly crucial role of attentional selection. Thus, an important question that remains unanswered is whether attentional selection constitutes a fundamental step within the proceduralization chain, which likely encompasses other processes, such as the intention to implement (Brass et al., 2017). That is, whether attentional selection contributes in the initial prioritization of instructed contents for their transformation into action codes.

In the present study, we tested the hypothesis that selective attention over instructed task sets, via retro-cues, would reformat them into an action-oriented representation, therefore enhancing IBR effects. In Experiment 1, we sought to obtain initial evidence that participants could benefit from the use of retro-cues in a context of implementation of novel instructions. To do so, we devised a paradigm that combined retro-cues with the traditional inducer tasks used in previous IBR experiments (Liefooghe et al., 2012; Meiran et al., 2015b). In this experiment, participants encoded 4 novel S-R associations at the beginning of each trial. Subsequently, a valid or neutral retro-cue could select a subset of these S-R associations. Finally, a stimulus would appear as a target, prompting participants to provide the associated response. This setting allowed us to test whether valid, compared to neutral retro-cues, enhanced execution of novel instructions. Next, in Experiments 2 and 3, we introduced a secondary task (i.e. diagnostic trials) between the retro-cue and the target, requiring responses that would or would not

overlap with the instructed S-R mappings, thus leading to compatibility effects (i.e. IBR). This allowed us to assess the effect of attentional selection in proceduralization. Specifically, if attentional selection helps in prioritizing information into a *behavior-optimized* state, then we should observe an enhanced IBR effect for those mappings selected by the retro-cue. Importantly, the fact that retro-cues operate on WM contents ensured identical processing of instructed mappings and in principle a similar declarative representation of such content regardless of the selection instantiated by the retro-cue. Therefore, any modulation of the IBR effect due to selection could arise only as a consequence of a transformation of already encoded content.

## 2. Experiment 1

### 2.1. Method

#### 2.1.1. Participants

A sample of 36 Dutch-native, undergraduate students (2 males, mean age = 19.2, *SD* = 3.6), with normal or corrected-to-normal vision, from Ghent University participated in exchange for course credits. Sample size was estimated a priori (PANGEA; https://jakewestfall.shinyapps.io/pangea/) to detect a medium effect size (d = 0.40) with a power of 0.80 using two repeated measurements (valid and neutral trials). Four participants were excluded from the analysis due to low accuracy (> 2 standard deviations of the across-participants mean) in the inducer task and/or memory probes.[1] Participants signed an informed consent form before participating in the study, which was carried out in accordance with the Declaration of Helsinki. Data are available at https://osf.io/u8y9c.

#### 2.1.2. Materials

S-R associations were created combining images with words that indicated the response finger. Given the prerequisite of novel S-R associations on each trial (see below), images of animate (non-human animals) and inanimate (vehicles and instruments) items were compiled from different available databases (Brady, Konkle, Alvarez, & Oliva, 2008; Brady, Konkle, Alvarez, & Oliva, 2013; Brodeur, Guérard, & Bouras, 2014; Griffin, Holub, & Perona, 2006; Konkle, Brady, Alvarez, & Oliva, 2010), creating a pool of 1550 unique pictures (770 animate items, 780 inanimate). To increase perceptual similarity and facilitate recognition, the background was removed from all images, items were centered in the canvas, and images were converted to black and white.
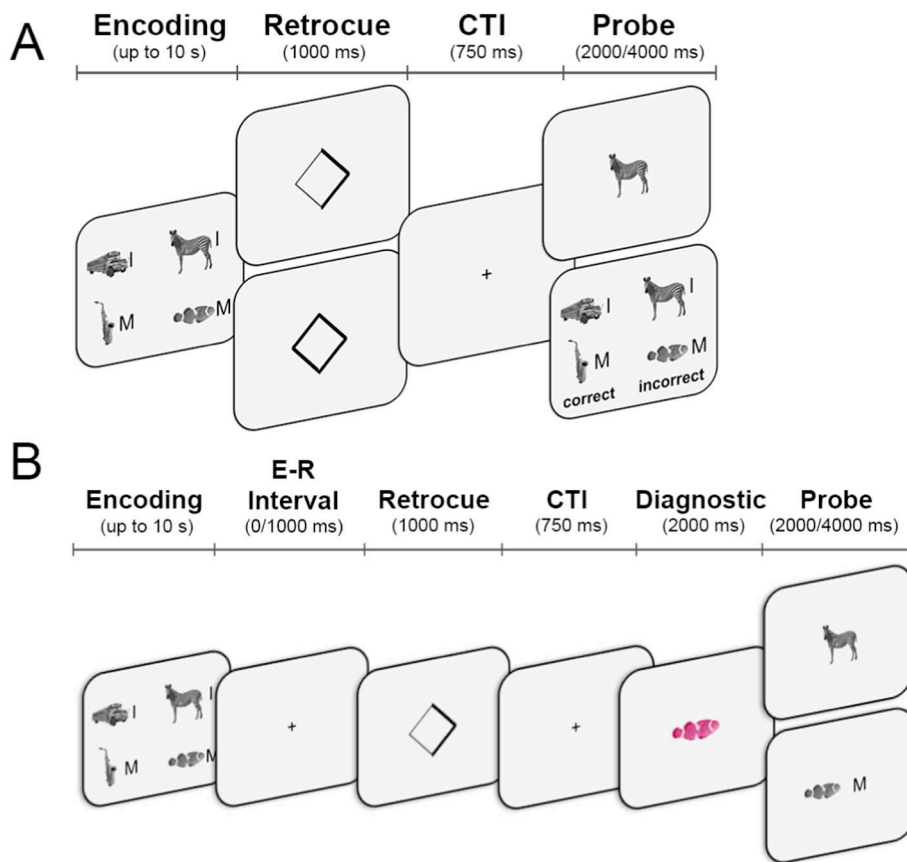
#### 2.1.3. Procedure

The study was individually carried out in a dimly illuminated and soundproof room. The task was presented to participants in Psychopy (Peirce, 2007). Participants first performed a practice session in which they were trained on the main task. The mean accuracy was computed every 16 trials, and participants had to repeat the practice (with new S-R mappings) until their mean accuracy was above 85%, or after a total of 64 trials. The unique mappings used during the practice were not repeated during the main task. Immediately after the practice session, participants performed the main task, which had 4 blocks of 64 trials each. Half of the trials included 100% valid retro-cues, while in the other half, neutral retro-cues appeared. Between each block, participants were allowed to take short breaks. The duration of the study was ~75 min.

Each trial of the task (for a visual representation see Fig. 1A) started with a white fixation cross prompting participants to hold the spacebar with their thumbs to see 4 new S-R mappings.[2] Two of these mappings

---

[1] The final sample thus was valid to detect a medium effect size (d = 0.40) with a power of 0.76.

[2] We decided to use 4 S-R mappings for several reasons. IBR is thought to depend on a capacity-limited system and therefore is primarily observed when the number of instructed mappings is restricted to a relatively low amount. Specifically, previous studies have not observed IBR effects when

**Fig. 1.** General paradigm and trial procedures. *(A)* Trial procedure for Experiment 1. At the beginning of each inducer trial, participants were instructed 4 novel S-R associations. Immediately after, a retro-cue appeared. Valid retro-cues (50% of trials) selected one side of the screen (e.g. right), and therefore cued 2 of the encoded associations. Neutral retro-cues selected the 4 mappings. In 75% of trials, a target consisting of one of the selected stimuli was displayed, prompting participants to execute its associated response. In the remaining trials, a memory probe appeared. (B) Trial procedure including diagnostic trials, used in Experiments 2 and 3. In 75% of trials, the retro-cue was followed by a 750 ms CTI and a diagnostic image. During diagnostic trials, participants had to respond to the color of the image with responses that could overlap with the instructed S-R mapping. After the diagnostic trial, or immediately after the CTI in the remaining 25% of trials, the probe appeared. In Experiment 3, to reduce the demands of memory probes, participants were probed on only 1 S-R mapping (as depicted in the figure). Also in Experiment 3, the duration of the encoding-retro-cue interval was either 0 or 1000 ms.

included animate items and the other 2 inanimate objects. The four mappings referred to bimanual responses ("middle" or "index" fingers from both hands, corresponding to "d", "k", and "s", "l" keys, respectively).[3] Each mapping (i.e. the association between responses and specific images) was presented just once during the entire experiment to prevent the formation of long-term memory traces (Meiran et al., 2015b). Moreover, the specific response could not be predicted based on the spatial location of the S-R association (i.e. mappings could appear in any of the four possible locations regardless of the required response). The encoding phase on each trial was self-paced, and therefore mappings remained on screen for as long as participants held the spacebar or for a maximum of 10 s. Immediately after the encoding screen, the next screen displayed a retro-cue (arrow centered in the middle of the screen, pointing to the left, to the right, or both sides). The retro-cue was on the screen for 1 s and followed by a fixation point, which was displayed for 750 ms (CTI; cue-target interval). Directly after the CTI, an inducer task target (2 s; 75% of trials) or a memory probe (4 s; 25% of trials) was displayed. Target screens displayed the image of one of the selected mappings, which prompted participants to execute the response associated with that image. Memory probes displayed again 4 S-R mappings and prompted participants to answer whether these were identical to the ones presented during the encoding phase of

that trial by pressing the two left fingers for *yes* (keys "s" and "d") or the two right fingers for *no* (keys "k" and "l"). These probes were included to force participants to hold a declarative-like representation of all 4 S-R mappings, regardless of the retro-cue validity.[4] Importantly, the retro-cue was relevant for the inducer task target but never for memory probes. Last, a red fixation cross appeared between trials for 2 s (ITI; inter-trials interval).

### 2.1.4. Design

A within-participant design with one factor (Validity; neutral vs. valid retro-cue; 50% of trials each) was employed. Reaction Times (RT) and error rates during the target and memory screens were used as dependent variables. To test the effect of validity, paired *t*-tests (or Wilcoxon tests when the normality assumption was violated) were carried out for each dependent measure. All analyses (also in Experiments 2 and 3) were performed in JASP (JASP Team, 2018).

### 2.2. Results

A paired Wilcoxon test[5] on the RTs of the inducer task revealed faster RTs after valid (*M* = 943 ms, *SD* = 153), compared to neutral
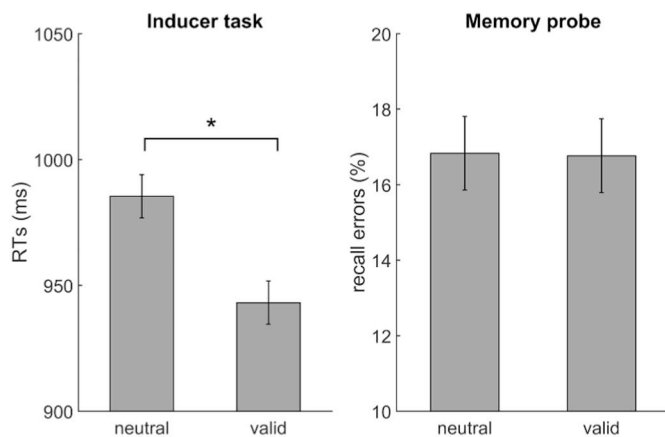
---

**Fig. 2.** Mean reaction times for probes (left) and recall error rate (right). Results are depicted for each validity condition separately. As reported in the results sections, RTs differed significantly in neutral and valid trials, whereas recall was equated in these two conditions. Stars denote significant differences ($p < 0.05$, paired Wilcoxon test). All error bars represent within-participant standard error mean (Morey, 2008).

($M = 985$ ms, $SD = 144$) retro-cues ($W_{31,1} = 78$, $p < 0.001$, effect size $= 0.7$; Fig. 2, left). In contrast, no differences were found on error rates ($M$ valid $= 7.67\%$, $SD$ valid $= 5.74\%$; $M$ neutral $= 7.74\%$, $SD$ neutral $= 5.5\%$; $t_{31,1} = 0.12$, $p = 0.45$, Cohen's d $= 0.02$). To test the reliability of such null effect, we performed equivalence tests using the methodology proposed in Lakens (2017). Specifically, we followed a TOST procedure, in which we first set upper and lower bounds based on standardized effect size (Cohen's d $= 0.5$) and considered results within these bounds as candidates for the absence of an effect. Then, two one-sided $t$-test were performed, comparing the observed value against the two boundaries (Lakens, 2017). An effect is considered absent when it is statistically below the higher bound and above the lower bound. When looking at error rates during the inducer task, this is precisely what we observed (upper $t_{31,1} = -2.95$, $p = 0.003$; lower $t_{31,1} = 2.70$, $p = 0.006$), suggesting statistically equivalent error rates in valid and neutral trials. Regarding memory probes, reaction times were faster for neutral ($M = 2587$ ms, $SD = 277$) than for valid trials ($M = 2634$, $SD = 218$; $t_{31,1} = 2.26$, $p = 0.03$, Cohen's d $= 0.4$). However, a paired t-test revealed no differences in recall performance after valid ($M$ error rate $= 17.8\%$, $SD = 10\%$) and neutral ($M = 17.8\%$, $SD = 9\%$) trials ($t_{31,1} = 0.047$, $p = 0.96$, Cohen's d $= 0.008$; Fig. 2 right). A TOST equivalence test revealed that recall performance after valid and neutral trials were statistically equivalent (upper $t_{31,1} = -2.78$, $p = 0.005$; lower $t_{31,1} = 2.88$, $p = 0.004$).

### 2.3. Discussion

These results are in line with previous observations showing that participants can benefit from retro-cues (see Souza & Oberauer, 2016) and, for the first time, extend these effects to the implementation of novel S-R mappings, as revealed by faster responses to targets in valid trials. Still, recall accuracy was not modulated by cue validity. This pattern of results suggests that retro-cues elicit the selection of the cued representations, which speeds up performance but still does not impair the initial representation of uncued mappings (Myers et al., 2018). This could be at least partially due to task requirements, which render un-cued mappings potentially relevant, therefore making it inefficient to drop them from memory. Still, one might wonder why we did not ob-serve a benefit in recall for valid trials, consistent with previous reports in the retro-cue literature. As explained before, retro-cues in this ex-periment refer to targets and not to memory probes and therefore retro-cue information can be rendered irrelevant in memory screens. In other words, whatever representation should be selected to observe such a

recall benefit needed not be prioritized by the retro-cue in our experi-ment. This issue is further discussed in Experiments 2 and 3.

In sum, Experiment 1 reveals an effect of validity in the im-plementation of novel S-R mappings. In Experiment 2, we sought to assess the specific effect of retro-cues on the representation of these mappings. More specifically, we wanted to investigate whether retro-cues establish intention-based reflexivity.

## 3. Experiment 2

Although Experiment 1 shows that the retrospective selection of novel instructions benefits execution, it is agnostic regarding the effect of retro-cues on the representational state of instructed mappings. In Experiment 2, we specifically tested the hypothesis that internal at-tention selectively prioritizes relevant instructed content, enhancing the transformation into an action-oriented representation. To do so, we extended the paradigm of Experiment 1 to include a secondary, diag-nostic task that induced compatible or incompatible responses with the instructed mappings. Thus, this experiment capitalized on 1) retro-cues to prioritize specific representations from WM, and 2) compatibility effects to track the degree of reflexivity of selected vs. unselected re-presentations. Specifically, in this experiment, we set out to assess whether IBR effects arise primarily for those instructions that are prioritized (i.e. selected) in working memory.

The research questions, hypotheses and analysis plan of Experiment 2 were pre-registered prior to data collection. The pre-registration re-port and data of this experiment are available at https://osf.io/u8y9c.

### 3.1. Methods

#### 3.1.1. Participants

A total of 34 undergraduate students (14 males, mean age $= 20.7$, $SD = 2.24$) from Ghent University participated in exchange for course credits. All participants were native Dutch speakers and had normal or corrected-to-normal vision. Given that the results of this experiment rely on compatibility effects between an inducer and a secondary task we expected a smaller effect size compared to Experiment 1. Thus, to achieve a power of 0.80 with an expected low-medium effect size of d $= 0.3$, the required sample was 34 participants. Three participants were excluded from the analysis due to low accuracy in the inducer task, diagnostic, and/or the memory probe. One additional participant failed to finish the task and therefore was excluded, resulting in a final sample of 30 participants.[6] The study was carried out in accordance with the Declaration of Helsinki and participants signed a consent form before participating in the study.

#### 3.1.2. Materials

As in Experiment 1, the same pool of images was used. Additionally, in diagnostic trials, participants had to respond to the color of the item (see *Procedure* below). Therefore, two color versions (one in pink, and one in yellow) of each of the images were created. To reduce the likelihood of processing the color without processing the identity of the image the color was limited to the item itself while keeping the back-ground of the image in white.

#### 3.1.3. Procedure

The inducer task remained identical to Experiment 1, except for the fact that only valid trials were used now. After the encoding phase and the retro-cue, a fixation point was displayed for 750 ms. In 75% of the trials, this interval was followed by a single diagnostic image (until participants' response or a maximum of 2 s). Diagnostic images consisted of color versions of the same images used in the encoding phase.

---

[6] The final sample thus was valid to detect a low-medium effect size (d $= 0.3$) with a power of 0.79.

Participants were instructed to always press both index fingers ("d", "k" keys) if the image was colored in pink, and both middle fingers ("s", "l" keys) if the image appeared in yellow. These responses overlapped with those of the inducer task, and therefore could be compatible or incompatible with the instructed mappings. In the remaining 25% of trials, no diagnostic tasks were included.[7] After the diagnostic task, or directly after the CTI in trials with no diagnostic tasks, a target (75% of trials) or a memory probe (25% of trials) appeared. As in Experiment 1, the location of two of the mappings (the selected or unselected subset) was shuffled in half of the memory screens. Given the absence of neutral trials in Experiment 2, this manipulation allowed us to measure the impact of the retro-cue on memory performance by comparing the recall rate when the shuffling occurred on the cued or the uncued side of the screen. Last, a red fixation cross was displayed between trials for 1 s (ITI).

Participants completed a total of 192 trials, from which 144 (75%) included one diagnostic task. Among all diagnostic trials, there were 36 observations per design cell (selected compatible, selected incompatible, unselected compatible, unselected incompatible). Of these 36 observations, 75% were followed by an inducer task probe, whereas 25% were followed by a memory probe.

Prior to the main task, two practice sessions took place. During the first one, participants practiced exclusively the inducer task, without diagnostic trials. After a maximum of 16 trials, the mean accuracy was computed. Participants had to repeat this practice block until their mean accuracy was above 85% or after a maximum of 64 trials. Afterward, participants performed the second practice session, which included both the inducer and the diagnostic tasks. Participants performed eight complete trials (including diagnostic tasks). This practice repeated until their accuracy was above 85% or after 4 blocks of eight trials. S-R mappings were always novel during the practice sessions and were not used during the main task. The total duration of the experiment was ~75 min.

### 3.1.4. Design

A within-participant design with two factors (Selection: diagnostic image corresponding to retro-cued (selected)/uncued mappings; Compatibility: compatible, incompatible responses between the inducer and the diagnostic tasks) was employed. As stated in the pre-registered report, based on previous studies (Liefooghe & De Houwer, 2018; Meiran & Pereg, 2017) and pilot data, we expected a compound measure such as Inverse Efficiency Scores (IES) to be a more sensitive measure of compatibility effects in this experimental setting, compared to RTs or error rate alone (Townsend & Ashby, 1983). Therefore, we focused on this dependent variable during diagnostic trials. To obtain IES, we divided each participant's mean trimmed RT (within $\pm 2$ s.d. of each participant's mean) of a design cell by the percentage of accurate responses. To test the effect of selection and compatibility, a repeated-measures ANOVA was carried out. To assess the effect of selection on probes, RTs and error rates in targets and memory screens were also used as dependent measures.

### 3.2. Results

### 3.2.1. Diagnostic trials

A repeated-measures ANOVA on the IES revealed a significant main effect of Selection ($F_{29,1} = 4.26$, $p = 0.048$, $\eta_p^2 = 0.13$) and Compatibility ($F_{29,1} = 5.77$, $p = 0.02$, $\eta_p^2 = 0.16$), whereas the Selection × Compatibility interaction did not reach significance ($F_{29,1} = 3.31$, $p = 0.08$, $\eta_p^2 = 0.1$). Numerically, similar results were found when using RTs or error rates alone (see Table 1 for descriptive statistics and Tables 2 and 3 for full ANOVA results using RTs and error rates).

### 3.2.2. Diagnostic trials – unplanned analysis

After following the preregistered analysis plan, we observed that the chosen cutoff of $\pm 2$ s.d. was filtering out a large portion of diagnostic trials (average of 18% of correct trials per participant). At this point, we decided to perform a new, unplanned analysis using a more lenient filter of $\pm 3$ s.d. (which removed an average of 3% of correct trials per participant). When trimming the data with this new cutoff, an ANOVA on the IES revealed a significant main effect of Selection ($F_{29,1} = 5.45$, $p = 0.027$, $\eta_p^2 = 0.16$), whereas the effect of Compatibility did not reach significance ($F_{29,1} = 3.96$, $p = 0.056$, $\eta_p^2 = 0.12$). Crucially, the interaction of both factors was significant ($F_{29,1} = 4.63$, $p = 0.04$, $\eta_p^2 = 0.14$; see Fig. 3). Planned comparisons further revealed significant differences between compatible ($M = 1062$ ms, $SD = 199$) and incompatible trials ($M = 1120$, $SD = 246$) of mappings selected by the retro-cue ($F_{29,1} = 8.7$, $p = 0.006$, $\eta_p^2 = 0.14$), whereas no differences were found between trials compatible ($M = 1129$, $SD = 273$) and incompatible ($M = 1129$, $SD = 257$) with unselected mappings (F < 1). When using RTs and error rates alone, the interaction effect did not reach significance, but the same pattern was found numerically (see Table 4 for descriptive statistics and Table 5 for full ANOVA results using RTs. For error rates, see Tables 1 and 3).

### 3.2.3. Inducer task

Participant performed overall well in targets ($M$ error rate $= 0.11$, $SD = 0.07$) and memory screens ($M = 0.18$, $SD = 0.09$). Critically, performance in all memory conditions was significantly higher than chance (one-sample $t$-test against 0.50; all $ts > 8.5$, all $ps < 0.001$, all $ds > 1.56$), suggesting that regardless of their selection status, all mappings were represented optimally. Despite the retro-cue did not predict the type of memory probe, a shuffling of the position of S-R associations could occur in the selected or the unselected subset of instructions. This allowed us to assess whether selection and the subsequent proceduralization (as revealed by diagnostic trials) had any effect on the declarative-like representation of mappings. To do so, we carried out repeated-measures ANOVAs with the factors Hit (whether the memory probe matched [hit] or not [no hit] the encoding screen) and Shuffle (whether the position of the selected or unselected mappings was shuffled compared to the encoding screen) on the error rates and RTs. As for error rates, neither the main effects of Hit ($F = 1.1$, $p = 0.303$) and Shuffle ($F = 1.05$, $p = 0.313$) nor the interaction component ($F < 1$) were significant. When using RTs, the analysis revealed faster reaction times for mismatching memory screens ($F_{29,1} = 49.36$, $p < 0.001$, $\eta_p^2 = 0.63$), probably reflecting the lower computational requirements of detecting a mismatch compared to exhaustively confirming a matching screen. Still, no differences were found between trials in which position shuffling occurred on the selected or the unselected mappings ($F < 1$). The interaction component was not significant either ($F < 1$). To explore the link between compatibility effects and memory performance, we extracted a compatibility index for each participant by subtracting the IES score in selected compatible trials to the score in selected incompatible trials. We then correlated this value with the individual mean memory performance, as well as with the mean encoding time of each participant (i.e. the time they held the spacebar during encoding), which in principle could modulate the fidelity of declarative representations. We found no significant correlations between the compatibility index and the remaining variables (all $ps > 0.05$). A Bayesian counterpart of this analysis provided anecdotal evidence for a correlation with error rates ($r = 0.34$, $BF_{10} = 1.19$), suggesting that, if anything, increased reflexivity was linked to overall worse memory performance.[8]

---

[7] Trials with zero diagnostic tasks were included to ensure that participants would actively try to implement the selected mappings. Otherwise, retro-cues would always be followed by a diagnostic screen and not the probe. In this case, participants could strategically decide to not prepare after the retro-cue and wait until the diagnostic task had finished (Meiran et al., 2015b).

[8] This analysis further provided support for a null correlation (Jeffreys, 1998) between compatibility index and RTs in memory probes ($r = -0.09$, $BF_{10} = 0.25$), and inconclusive evidence for the correlation with encoding time ($r = -0.27$, $BF_{10} = 0.6$).

**Table 1**
Descriptive statistics (mean and standard deviation) for diagnostic trials of Experiment 2.

| Measure | Selected | | Unselected | |
| --- | --- | --- | --- | --- |
| | Compatible | Incompatible | Compatible | Incompatible |
| RTs | 999 (171) | 1013 (165) | 1010 (187) | 1020 (183) |
| ER | 0.03 (0.03) | 0.06 (0.05) | 0.06 (0.06) | 0.06 (0.06) |
| IES | 1037 (191) | 1081 (225) | 1084 (249) | 1089 (232) |

**Table 2**
ANOVA results of Experiment 2 using RTs.

| | Sum of squares | df | Mean square | F | $p$ | $\eta_p^2$ |
| --- | --- | --- | --- | --- | --- | --- |
| Selection | 2445 | 1 | 2445 | 1.068 | 0.310 | 0.036 |
| Residual | 66,419 | 29 | 2290 | | | |
| Compatibility | 4364 | 1 | 4394 | 2.880 | 0.100 | 0.090 |
| Residual | 44,242 | 29 | 1525 | | | |
| Selection * Compatibility | 119 | 1 | 119 | 0.057 | 0.813 | 0.002 |
| Residual | 60,646 | 29 | 2091 | | | |

*Note.* Type III sum of squares.

**Table 3**
ANOVA results of Experiment 2 using error rates.

| | Sum of squares | df | Mean square | F | $p$ | $\eta_p^2$ |
| --- | --- | --- | --- | --- | --- | --- |
| Selection | 0.005 | 1 | 0.005 | 2.690 | 0.112 | 0.085 |
| Residual | 0.059 | 29 | 0.002 | | | |
| Compatibility | 0.003 | 1 | 0.003 | 2.319 | 0.139 | 0.074 |
| Residual | 0.033 | 29 | 0.001 | | | |
| Selection * Compatibility | 0.004 | 1 | 0.004 | 1.964 | 0.172 | 0.063 |
| Residual | 0.061 | 29 | 0.002 | | | |

*Note.* Type III sum of squares.
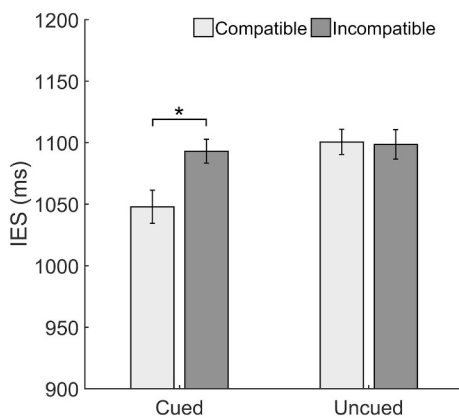


**Fig. 3.** Mean inverse efficiency scores during diagnostic trials of Experiment 2. Results are depicted for each condition separately. As reported in the results section, when using a more lenient filter, a significant Selection × Compatibility interaction was found, with significant differences between compatible and incompatible trials only for cued mappings. Stars denote significant differences ($p < 0.05$). Error bars represent within-participant s.e.m.

*3.3. Discussion*

In Experiment 2 we found initial evidence in line with the hypothesis that merely instructed novel tasks had a reflexive impact on behavior insofar they had been previously selected by a retro-cue, whereas compatible and incompatible responses with unselected task sets were effectively identical. These results suggest that selective attention, elicited by retro-cues, help in the prioritization of previously encoded task sets that would otherwise remain in an action-unspecific

**Table 4**
Descriptive statistics (mean and standard deviation) for diagnostic trials of Experiment 2 ( ± 3 s.d. filter).

| Measure | Selected | | Unselected | |
| --- | --- | --- | --- | --- |
| | Compatible | Incompatible | Compatible | Incompatible |
| RTs | 1010 (167) | 1025 (173) | 1027 (189) | 1030 (180) |
| IES | 1048 (187) | 1093 (232) | 1101 (247) | 1099 (228) |

**Table 5**
ANOVA results of Experiment 2 using RTs ( ± 3 s.d. filter).

| | Sum of squares | df | Mean square | F | $p$ | $\eta_p^2$ |
| --- | --- | --- | --- | --- | --- | --- |
| Selection | 3334 | 1 | 3334 | 1.539 | 0.225 | 0.050 |
| Residual | 62,840 | 29 | 2167 | | | |
| Compatibility | 2156 | 1 | 2156 | 1.229 | 0.277 | 0.041 |
| Residual | 50,895 | 29 | 1755 | | | |
| Selection * Compatibility | 1077 | 1 | 1077 | 0.500 | 0.485 | 0.017 |
| Residual | 62,523 | 29 | 2156 | | | |

*Note.* Type III sum of squares.

state. This prioritization, in turn, updates the format of selected tasks into a reflexive-like (Liefooghe et al., 2012; Meiran et al., 2015b), *behavior-optimized* representation (Myers et al., 2017). An outstanding question is whether such a mechanism implies a zero-sum game, in which benefits for cued mappings come at a cost for the uncued ones. The alternative is that prioritization renders cued items into a privileged state without hindering the representation of other items that might still be relevant (Myers et al., 2017). The results of Experiment 2 seem to be in line with the latter interpretation. First, results in the diagnostic trials suggest a benefit for selected compatible trials, rather than a cost for unselected mappings. Second, despite the specific IBR effects for selected mappings, recall was not modulated by selection, suggesting that participants can hold the initial, declarative-like representation of all (selected and unselected) instructions. Still, this pattern of results could partially be due to the complex nature of the memory probes, in which 4 mappings were presented at the same time, potentially masking any subtle influence of selection. In Experiment 3, we set out to explore this possibility and, importantly, replicate the selective IBR effect in the diagnostic trials.

**4. Experiment 3**

A further consideration of the previous experiment is that, given that the retro-cue appears immediately after the encoding screen, our results could be partially driven by iconic memory traces (Souza & Oberauer, 2016). To test whether selective prioritization of task sets can occur beyond the realm of iconic memory, we manipulated the retention interval between the encoding screen and the onset of the retro-cue. Based on the consistent finding that retro-cue benefits go above and beyond selective read-outs from iconic memory (Souza & Oberauer, 2016), we predicted that prioritization (i.e. compatibility effects) would also take place, perhaps even to a larger extent, in trials with larger retention intervals. Last, we modified the nature of the memory task, reducing its complexity, to test if the potential effects of selection of task sets on recall could be captured in such a setting.

The research questions, hypotheses and analysis plan of Experiment 3 were pre-registered prior to data collection. The pre-registration report and data can be found at https://osf.io/u8y9c.

*4.1. Methods*

*4.1.1. Participants*

A total of 40 undergraduate students (3 males, mean age = 21,

$SD$ = 1.38) from Ghent University participated in exchange for course credits. All participants were native Dutch speakers and had normal or corrected-to-normal vision. Given the relatively small effects in Experiment 2, we estimated the needed sample size assuming a low effect size ($d = 0.2$) and a power $> 0.8$. Two participants were excluded from the analysis due to low accuracy in the inducer task, diagnostic, and/or the memory probe. One additional participant failed to finish the task and therefore was excluded, resulting in a final sample of 37 participants.[9] The study was carried out in accordance with the Declaration of Helsinki and participants signed a consent form before participating in the study.

### 4.1.2. Materials

The same materials as in Experiment 2 were used.

### 4.1.3. Procedure

The task procedure remained identical to Experiment 2, except for 1) the configuration of memory probes and 2) the manipulation of the encoding - retro-cue interval. First, to lower the requirements of memory probes, these now consisted of only one image and one response. Still, the retro-cue was irrelevant regarding the specific mapping to be probed in memory screens. As in Experiment 2, participants had to press both left keys in matching and both right keys in mismatching screens. Second, to test whether the retention interval between the encoding and the retro-cue impacted the effect of the latter, we included a blocked manipulation. In long interval blocks, this interval lasted 1 s, whereas, in short interval blocks, the retro-cue was displayed immediately after the offset of the encoding screen. In total, participants performed 192 trials, from which 75% included one diagnostic task, resulting in 18 observations per design cell (selected compatible/incompatible, unselected compatible/incompatible, in short, and long intervals). The task was divided into 6 blocks of 32 trials each. Three of these blocks comprised short encoding-retro-cue intervals, while the other 3 contained long intervals. Prior to the main task, two practice sessions took place, mimicking the procedure of Experiment 2. In these sessions, the encoding-retro-cue interval was either 0 or 1 s, selected randomly.

### 4.1.4. Design

A $2 \times 2 \times 2$ within-participant design (Selection; Compatibility; and Interval [Short/Long]) was employed. IES during the diagnostic trials and RTs (within $\pm 2$ s.d. of each participant's mean) and error rates during targets and memory probes were used as dependent measures.

### 4.2. Results

#### 4.2.1. Diagnostic trials

A repeated-measures ANOVA on the IES revealed a significant main effect of Compatibility ($F_{36,1} = 16.72$, $p < 0.001$, $\eta_p^2 = 0.32$). Crucially, this analysis revealed again a significant Selection $\times$ Compatibility interaction ($F_{36,1} = 6.63$, $p = 0.014$, $\eta_p^2 = 0.16$; see Fig. 4), replicating results from Experiment 2. Specifically, we found significant differences between compatible ($M = 1009$, $SD = 154$) and incompatible trials ($M = 1072$, $SD = 205$) of mappings selected by the retro-cue ($F_{36,1} = 18.25$, $p < 0.001$, $\eta_p^2 = 0.34$), whereas no differences were found between trials compatible ($M = 1040$, $SD = 139$) and incompatible ($M = 1058$, $SD = 169$) with unselected mappings ($F_{36,1} = 2.55$, $p = 0.12$, $\eta_p^2 = 0.07$). Moreover, neither the main effect of Interval, nor its interaction with Selection or Compatibility was significant (all $p$s $> 0.19$). The Selection $\times$ Compatibility $\times$ Interval interaction was not significant either ($F = 1.01$, $p = 0.32$).

To determine the reliability of this non-significant finding, we performed a Bayesian ANOVA with the same factors as before (using the

methodology proposed in Rouder, Morey, Speckman, & Province, 2012). The $BF_{10}$ (evidence in favor of $H_1$ against evidence for $H_0$) for the Selection $\times$ Compatibility $\times$ Interval interaction was 0.42, after removing the main effect of compatibility. This constitutes close to moderate evidence for the null hypothesis (Jeffreys, 1998) that the Interval Factor does not modulate the Selection $\times$ compatibility interaction. Similar results were obtained when using RTs and errors alone (see Table 6 for descriptive statistics and Tables 7 and 8 for full ANOVA results).

#### 4.2.2. Inducer task

As in Experiment 2, we observed overall good performance in targets ($M$ error rate = 0.11, $SD = 0.07$) and memory screens ($M = 0.18$, $SD = 0.07$). Specifically, during memory probes, performance was above chance in all conditions (one-sample t-test against 0.50; all $ts > 5.54$, all $ps < 0.001$, all $ds > 0.97$). To explore the effect of selection on recall, we carried out a repeated-measures ANOVA with the factors Hit, Selection, and Encoding-Retro-Cue Interval, on the error rates and RTs. When using RTs, no significant effects were found (all $ps > 0.05$; although the Hit $\times$ Selection interaction yielded an $F = 4.17$ and $p = 0.052$, a Bayesian analysis revealed no reliable evidence ($BF_{10} = 0.82$) for such interaction). As for error rates, neither the main effects (all $ps > 0.27$) nor any interaction was significant (all $Fs < 1$). Last, we tested for any potential relationship between compatibility effects and memory recall by correlating a compatibility index (extracted as in Experiment 2) with overall memory performance and encoding time. As in Experiment 2, correlations with encoding time and RTs were not significant (all $ps \geq 0.05$). A Bayesian counterpart revealed anecdotal evidence for a positive correlation with error rates ($r = 0.32$, $BF_{10} = 1.3$).

### 4.3. Discussion

In Experiment 3, we replicated the main results from Experiment 2, obtaining specific IBR effects only for mappings that had been selected by a retro-cue. As in the previous experiment, the results point to a benefit for selected compatible mappings, rather than costs for incompatible or unselected mappings. Interestingly, we found evidence towards a null effect of the interval, revealing that the selective prioritization of novel task sets can take place beyond the realm of iconic memory. Regarding performance in memory probes, we again found similar recall rates for probes of selected and unselected mappings, even though the configuration of the probes was less demanding than in the previous experiments.

## 5. General discussion

In the present study, we describe for the first time the benefits of attentional selection in the implementation of novel instructions. By combining an IBR paradigm with retro-cues we first found that participants were able to use valid retro-cues to enhance the execution of novel task sets, while still holding in mind optimal representations of both selected and unselected mappings. Furthermore, in two pre-registered experiments, we found and replicated an interaction between selection and compatibility of diagnostic trials. Specifically, we observed a reflexive activation of responses of only those mappings that were previously selected by a retro-cue, whereas no evidence of IBR was found for uncued instructions. As in Experiment 1, the effective use of the retro-cue did not modulate recall rates. Altogether, these results extend retro-cues benefits previously reported in visual WM studies to the implementation of novel tasks and, crucially, reveal an important role of attentional selection in the emergence of IBR. We propose that attentional selection could work (potentially together with other processes) as a prioritization mechanism that reformats relevant encoded task knowledge into an optimal state that entails action-oriented representations (Myers et al., 2017).

Previous studies have extensively reported top-down, strategic control

---

[9] The final sample thus was valid to detect a low effect size ($d = 0.2$) with power = 0.84.
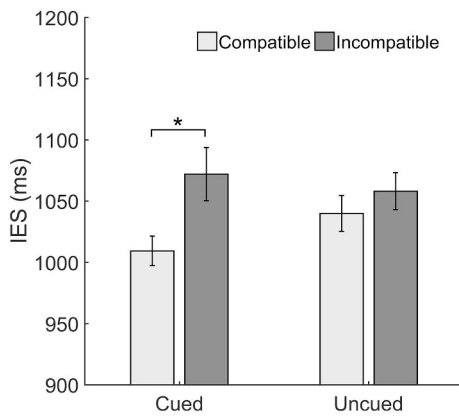
**Fig. 4.** Mean inverse efficiency scores during diagnostic trials of Experiment 3. Replicating results from the second experiment, we found a significant Selection × Compatibility interaction and differences between compatible and incompatible trials only for cued mappings. Stars denote significant differences ($p < 0.05$). Error bars represent within-participant s.e.m.

over the reflexive effects of instructed mappings (Liefooghe et al., 2013; Wenke et al., 2009; Whitehead & Egner, 2018). This has led to the speculation that instructions can be represented in two differentiated functional states, one related to the maintenance of declarative task knowledge, and another, linked to specific action-oriented task sets and in principle activated via the intention to implement (Brass et al., 2017). The current study reveals a mechanism through which relevant declarative items can be reformatted into these *condition-action* task sets. Our results can be interpreted within Oberauer's WM model (Oberauer, 2009). This model proposes an architecture with two subsystems, declarative (in charge of maintaining information without a specific action plan) and procedural WM (action-bound rules). Furthermore, the model predicts that several items can be held in declarative and procedural WM, but that only a number of relevant representations placed in the *region of direct access* (in

the case of declarative items) or the *bridge* (in case of procedural WM) can guide behavior at a certain time. In principle, the intention to implement a novel instruction would place the specific condition-action links of that instruction in the *bridge*, leading to the reflexive activation of responses (Liefooghe et al., 2012; Liefooghe & De Houwer, 2018; Meiran et al., 2015b). Within Oberauer's framework, the reported influence of internal selective attention in reformatting instructions could be conceptualized as one mechanism through which declarative condition-action links already present in declarative WM are recoded into the *bridge*.

One major question stemming from this interpretation is whether this mechanism is not only necessary but also sufficient. In other words, whether or not attention must be coupled with additional processes, such as the intention to implement (see also Liefooghe and De Houwer (2018), for a similar issue). In this regard, it should be noted that the current setting does not allow disentangling attentional selection and intention. Although we have demonstrated that participants benefit from valid retro-cues (which suggests that retro-cues are eliciting attentional processes that optimize behavior), the specific reflexive effect of selected mappings could be in principle also driven by the intention to implement such mappings. Therefore, whether IBR arises as a consequence of attentional selection exclusively or whether it interacts with additional processes remains unclear. Future studies should try to answer this question by, for instance, embedding a similar retro-cue procedure in a task where participants have to memorize (instead of implement) novel S-R mappings. This would allow observing if attention by itself generates IBR effects, even without the intention to implement.

Our results resonate with the general idea that attention and action are tightly intertwined (e.g. Allport, 1987) and, in particular, with theoretical accounts that propose a bidirectional influence of action information in attentional processing (Fagioli, Ferlazzo, & Hommel, 2007; Fagioli, Hommel, & Schubotz, 2007; Soto, Hodsoll, Rotshtein, & Humphreys, 2008). In this regard, previous studies have revealed that novel instructions can modulate early attentional processes (Tibboel, Liefooghe, & De Houwer, 2016). In their study, Tibboel et al. (2016), demonstrated that the establishment of an action plan via instructions

**Table 6**
Descriptive statistics (mean and standard deviation) for diagnostic trials of Experiment 3.

| Measure | Short CTI | | | | Long CTI | | | |
|---|---|---|---|---|---|---|---|---|
| | Selected | | Unselected | | Selected | | Unselected | |
| | Comp. | Incomp. | Comp. | Incomp. | Comp. | Incomp. | Comp. | Incomp. |
| RTs | 995 (135) | 1034 (145) | 999 (151) | 1009 (134) | 993 (147) | 1015 (142) | 1029 (135) | 1031 (161) |
| ER | 0.01 (0.02) | 0.04 (0.07) | 0.03 (0.05) | 0.03 (0.04) | 0.02 (0.05) | 0.04 (0.07) | 0.01 (0.04) | 0.03 (0.05) |
| IES | 1003 (135) | 1084 (220) | 1035 (172) | 1046 (159) | 1015 (175) | 1060 (191) | 1045 (147) | 1070 (181) |

**Table 7**
ANOVA results for Experiment 3 using RTs.

| | Sum of squares | df | Mean square | F | p | $\eta_p^2$ |
|---|---|---|---|---|---|---|
| CTI | 4796.2 | 1 | 4796.2 | 0.392 | 0.535 | 0.011 |
| Residual | 440,430.4 | 36 | 12,234.2 | | | |
| Selection | 4651.1 | 1 | 4651.1 | 1.379 | 0.248 | 0.037 |
| Residual | 121,421.5 | 36 | 3372.8 | | | |
| Compatibility | 25,413.8 | 1 | 25,413.8 | 7.790 | 0.008 | 0.178 |
| Residual | 117,441.2 | 36 | 3262.3 | | | |
| CTI ∗ Selection | 25,233.9 | 1 | 25,233.9 | 6.532 | 0.015 | 0.154 |
| Residual | 139,068.1 | 36 | 3863.0 | | | |
| CTI ∗ Compatibility | 2586.5 | 1 | 2586.5 | 0.902 | 0.349 | 0.024 |
| Residual | 103,221.3 | 36 | 2867.3 | | | |
| Selection ∗ Compatibility | 11,350.7 | 1 | 11,350.7 | 4.131 | 0.050 | 0.103 |
| Residual | 98,928.6 | 36 | 2748.0 | | | |
| CTI ∗ Selection ∗ Compatibility | 493.3 | 1 | 493.3 | 0.122 | 0.729 | 0.003 |
| Residual | 145,665.2 | 36 | 4046.3 | | | |

*Note.* Type III sum of squares.

**Table 8**
ANOVA results for Experiment 3 using error rates.

|  | Sum of squares | df | Mean square | F | p | $\eta_p^2$ |
|---|---|---|---|---|---|---|
| CTI | 5.274e − 4 | 1 | 5.274e − 4 | 0.146 | 0.704 | 0.004 |
| Residual | 0.130 | 36 | 0.004 |  |  |  |
| Selection | 5.629e − 4 | 1 | 5.629e − 4 | 0.352 | 0.557 | 0.010 |
| Residual | 0.058 | 36 | 0.002 |  |  |  |
| Compatibility | 0.021 | 1 | 0.021 | 9.203 | 0.004 | 0.204 |
| Residual | 0.082 | 36 | 0.002 |  |  |  |
| CTI ∗ Selection | 0.003 | 1 | 0.003 | 2.099 | 0.156 | 0.055 |
| Residual | 0.050 | 36 | 0.001 |  |  |  |
| CTI ∗ Compatibility | 2.036e − 4 | 1 | 2.036e − 4 | 0.080 | 0.779 | 0.002 |
| Residual | 0.091 | 36 | 0.003 |  |  |  |
| Selection ∗ Compatibility | 0.003 | 1 | 0.003 | 2.120 | 0.154 | 0.056 |
| Residual | 0.047 | 36 | 0.001 |  |  |  |
| CTI ∗ Selection ∗ Compatibility | 0.004 | 1 | 0.004 | 1.903 | 0.176 | 0.050 |
| Residual | 0.081 | 36 | 0.002 |  |  |  |

*Note.* Type III sum of squares.

(e.g. if I see DOG, press right) biased attention towards the encoded stimulus (in the previous example, DOG). Importantly, they showed that this bias can take place even if the action plan has never been executed before, as it is the case with novel instructions, but that it disappears if participants' goal is to memorize (instead of implement) the instruction. This shows that linking the representation of a stimulus with the intention to perform an action consolidates it in a prioritized state compared to competing representations in WM. Our findings suggest that this attentional bias to relevant content, in turn, reformats it into a prioritized, action-oriented state in WM.

Furthermore, although not directly related to the main aim of the present study, our results might also shed some light on the debate of whether proceduralization can be understood as merely deeper declarative processing or rather as a differentiated cognitive state. If procedural representations reflect simply deeper declarative representations, then retro-cue benefits should be observed in action-bound processes but also in recall performance. In this regard, our results reveal a specific effect of selection on action-bound processes. This interpretation would be in line with the idea that task implementation entails a unique state in WM that is, first, action-oriented, and second, at least partially distinct from the mere maintenance of task knowledge in WM (Brass et al., 2017). However, a definitive answer to this question goes beyond the goal of the current study, which was optimized to measure the effect of selection on action processes but not on memory recall. It should be mentioned that although there is extensive evidence of retro-cue benefits in recall performance (see Souza & Oberauer, 2016), our findings are not incompatible with these. In such studies, retro-cues signal which items will be probed next in a recall test. This differs from the current study, in which the retro-cue is exclusively informative for inducer task targets and not for recall screens. This suggests that whereas in visual WM studies retro-cues operate on declarative representations, in the current study they prioritize action components of instructions. Therefore, as it has been suggested before, it seems appropriate to conceive retro-cues as inducers of a *behavior-optimized* state for upcoming demands (Myers et al., 2017), independently of the specific defining features of that state. As mentioned above, future studies should explore the specific role of attentional selection in the declarative maintenance of novel task sets.

Last, a question that derives from our findings is whether selection of task sets operates in a zero-sum fashion. Given that prioritization is likely a limited-capacity process (Myers et al., 2017), one possibility is that it comes at a cost of the representations of the unselected mappings. In contrast with this idea, we first observed that IBR effects seem to reflect a benefit for selected compatible responses, rather than a conflict-related cost for incompatible or unselected mappings. Second, in all three experiments, we observed overall optimal memory performance and, importantly, similar recall patterns for selected and unselected mappings. Altogether, and in line with recent findings (Myers

et al., 2018), our results suggest that as long as unselected mappings are still somewhat relevant (as is the case in this study) prioritization benefits can emerge without hindering the declarative representation of the uncued mappings. In this regard, future studies should explore the effects of effectively tagging the uncued mappings as irrelevant. In such a situation, participants could drop this information from WM and a selection effect could potentially also take place in recall performance. Similarly, if IBR effects are understood as a result of shielding relevant representations from interference (Meiran et al., 2015b), it would be interesting to explore whether dropping irrelevant information enhances or reduces IBR effects.

In sum, in the current study, we report the specific effects of attentional selection on the implementation of novel task sets. Across three experiments, we found evidence towards the hypothesis that attentional selection prioritizes previously encoded instructions, reformatting them into a highly-accessible, action-oriented representation. This prioritization entails benefits for execution and does not come at a cost for unselected mappings as long as they are task-relevant. Altogether, our findings support the idea that two differentiated cognitive processes underlie the implementation of novel task sets, one related to maintaining task knowledge, and a second one, action-bound. Critically, they provide, for the first time, evidence that *attention to action* (Allport, 1987; Tibboel et al., 2016) might be the bridging mechanism between such two states.

### Declaration of competing interest

None.

### Acknowledgments

### References

Allport, D. A. (1987). Selection for action: Some behavioral and neurophysiological considerations of attention and action. In H. Heuer, & D. F. Saunders (Eds.). *Perspectives on perception and action* (pp. 395–419). Hilsdale, NJ: Erlbaum.

Bhandari, A., & Duncan, J. (2014). Goal neglect and knowledge chunking in the construction of novel behaviour. *Cognition, 130*(1), 11–30. https://doi.org/10.1016/j.cognition.2013.08.013.

Bourguignon, N. J., Braem, S., Hartstra, E., De Houwer, J., & Brass, M. (2018). Encoding of novel verbal instructions for prospective action in the lateral prefrontal cortex: Evidence from univariate and multivariate functional magnetic resonance imaging analysis. *Journal of Cognitive Neuroscience, 30*(8), 1170–1184. https://doi.org/10.1162/jocn_a_01270.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a

massive storage capacity for object details. *Proceedings of the National Academy of Sciences, 105*(38), 14325–14329. https://doi.org/10.1073/pnas.0803390105.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2013). Real-world objects are not represented as bound units: Independent forgetting of different object details from visual memory. *Journal of Experimental Psychology: General, 142*(3), 791. https://doi.org/10.1037/a0029649.

Braem, S., Liefooghe, B., De Houwer, J., Brass, M., & Abrahamse, E. L. (2017). There are limits to the effects of task instructions: Making the automatic effects of task instructions context-specific takes practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(3), 394–403. https://doi.org/10.1037/xlm0000310.

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. https://doi.org/10.1016/j.neubiorev.2017.02.012.

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) phase ii: 930 new normative photos. *PLoS One, 9*(9), e106953. https://doi.org/10.1371/journal.pone.0106953.

Bundt, C., Bardi, L., Abrahamse, E. L., Brass, M., & Notebaert, W. (2015). It wasn't me! Motor activation from irrelevant spatial information in the absence of a response. *Frontiers in Human Neuroscience, 9*, 539. https://doi.org/10.3389/fnhum.2015.00539.

Cohen-Kdoshay, O., & Meiran, N. (2007). The representation of instructions in working memory leads to autonomous response activation: Evidence from the first trials in the flanker paradigm. *Quarterly Journal of Experimental Psychology, 60*(8), 1140–1154. https://doi.org/10.1080/17470210600896674.

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. https://doi.org/10.3758/s13415-012-0125-7.

Demanet, J., Liefooghe, B., Hartstra, E., Wenke, D., De Houwer, J., & Brass, M. (2016). There is more into 'doing' than 'knowing': The function of the right inferior frontal sulcus is specific for implementing versus memorising verbal instructions. *NeuroImage, 141*, 350–356. https://doi.org/10.1016/j.neuroimage.2016.07.059.

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology, 30*(3), 257–303. https://doi.org/10.1006/cogp.1996.0008.

Everaert, T., Theeuwes, M., Liefooghe, B., & De Houwer, J. (2014). Automatic motor activation by mere instruction. *Cognitive, Affective, & Behavioral Neuroscience, 14*(4), 1300–1309. https://doi.org/10.3758/s13415-014-0294-7.

Fagioli, S., Ferlazzo, F., & Hommel, B. (2007). Controlling attention through action: Observing actions primes action-related stimulus dimensions. *Neuropsychologia, 45*, 3351–3355. https://doi.org/10.1016/j.neuropsychologia.2007.06.012.

Fagioli, S., Hommel, B., & Schubotz, R. I. (2007). Intentional control of attention: Action planning primes action-related stimulus dimensions. *Psychological Research, 71*, 22–29. https://doi.org/10.1007/s00426-005-0033-3.

González-García, C., Arco, J. E., Palenciano, A. F., Ramírez, J., & Ruz, M. (2017). Encoding, preparation and implementation of novel complex verbal instructions. *NeuroImage, 148*, 264–273. https://doi.org/10.1016/j.neuroimage.2017.01.037.

Griffin, G., Holub, A., & Perona, P. (2006). Caltech-256 object category dataset. *Caltech Technical Report*. https://doi.org/10.1021/jp953720e.

Hartstra, E., Kühn, S., Verguts, T., & Brass, M. (2011). The implementation of verbal instructions: An fMRI study. *Human Brain Mapping, 32*(11), 1811–1824. https://doi.org/10.1002/hbm.21152.

JASP Team. (2018). JASP. [Computer Software].

Jeffreys, H. (1998). *The theory of probability*. Oxford: OUP.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General, 139*(3), 558. https://doi.org/10.1037/a0019165.

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science, 8*(4), 355–362. https://doi.org/10.1177/1948550617697177.

Liefooghe, B., & De Houwer, J. (2018). Automatic effects of instructions do not require the intention to execute these instructions. *Journal of Cognitive Psychology, 1*–14. https://doi.org/10.1080/20445911.2017.1365871.

Liefooghe, B., De Houwer, J., & Wenke, D. (2013). Instruction-based response activation depends on task preparation. *Psychonomic Bulletin & Review, 20*(3), 481–487. https://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*(5), 1325–1335. https://doi.org/10.1037/a0028148.

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. https://doi.org/10.3389/fnhum.2012.00104.

Meiran, N., & Pereg, M. (2017). Automatic retrieval of newly instructed cue-task associations seen in task-conflict effects in the first trial after cue-task instructions. *Experimental Psychology, 64*, 37–48. https://doi.org/10.1027/1618-3169/a000349.

Meiran, N., Pereg, M., Kessler, Y., Cole, M. W., & Braver, T. S. (2015a). Reflexive activation of newly instructed stimulus–response rules: Evidence from lateralized readiness potentials in no-go trials. *Cognitive, Affective, & Behavioral Neuroscience, 15*(2), 365–373. https://doi.org/10.3758/s13415-014-0321-8.

Meiran, N., Pereg, M., Kessler, Y., Cole, M. W., & Braver, T. S. (2015b). The power of instructions: Proactive configuration of stimulus–response translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(3), 768–786. https://doi.org/10.1037/xlm0000063.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology.*. https://doi.org/10.20982/tqmp.04.2.p061.

Muhle-Karbe, P. S., Duncan, J., De Baene, W., Mitchell, D. J., & Brass, M. (2017). Neural coding for instruction-based task sets in human frontoparietal and visual cortex. *Cerebral Cortex, 27*(3), 1891–1905. https://doi.org/10.1093/cercor/bhw032.

Myers, N. E., Chekroud, S. R., Stokes, M. G., & Nobre, A. C. (2018). Benefits of flexible prioritization in working memory can arise without costs. *Journal of Experimental Psychology: Human Perception and Performance, 44*(3), 398–411. https://doi.org/10.1037/xhp0000449.

Myers, N. E., Stokes, M. G., & Nobre, A. C. (2017). Prioritizing information during working memory: Beyond sustained internal attention. *Trends in Cognitive Sciences, 21*(6), 449–461. https://doi.org/10.1016/j.tics.2017.03.010.

Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation, 51*, 45–100.

Oberauer, K. (2013). The focus of attention in working memory—From metaphors to mechanisms. *Frontiers in Human Neuroscience, 7*, 673. https://doi.org/10.3389/fnhum.2013.00673.

Palenciano, A. F., González-García, C., Arco, J. E., & Ruz, M. (2018). Transient and sustained control mechanisms supporting novel instructed behavior. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhy273 (bhy273).

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*. https://doi.org/10.1016/j.jneumeth.2006.11.017.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001.

Soto, D., Hodsoll, J., Rotshtein, P., & Humphreys, G. W. (2008). Automatic guidance of attention from working memory. *Trends in Cognitive Sciences, 12*, 248–342. https://doi.org/10.1016/j.tics.2008.05.007.

Souza, A. S., & Oberauer, K. (2016). In search of the focus of attention in working memory: 13 years of the retro-cue effect. *Attention, Perception, & Psychophysics, 78*(7), 1839–1860. https://doi.org/10.3758/s13414-016-1108-5.

Tibboel, H., Liefooghe, B., & De Houwer, J. (2016). Attention to future actions: The influence of instructed S-R versus S-S mappings on attentional control. *Psychological Research, 80*(6), 905–911. https://doi.org/10.1007/s00426-015-0695-4.

Townsend, J., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge: Cambridge: Cambridge University Press.

Wenke, D., Gaschler, R., Nattkemper, D., & Frensch, P. A. (2009). Strategic influences on implementing instructions for future actions. *Psychological Research, 73*(4), 587–601. https://doi.org/10.1007/s00426-009-0239-x.

Whitehead, P. S., & Egner, T. (2018). Frequency of prospective use modulates instructed task-set interference. *Journal of Experimental Psychology: Human Perception and Performance, 44*(12), 1970–1980. https://doi.org/10.1037/xhp0000586.