Journal of Experimental Psychology:
Learning, Memory, and Cognition
2015, Vol. 41, No. 3, 768–786

0278-7393/15/$12.00 http://dx.doi.org/10.1037/xlm0000063

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

Nachshon Meiran, Maayan Pereg, and Yoav Kessler
Ben-Gurion University of the Negev

Michael W. Cole
Washington University and Rutgers University

Todd S. Braver
Washington University

Humans are characterized by an especially highly developed ability to use instructions to prepare toward upcoming events; yet, it is unclear just how powerful instructions can be. Although prior work provides evidence that instructions can be sufficiently powerful to proactively program working memory to execute stimulus–response (S-R) translations, in a reflexlike fashion (intention-based reflexivity [IBR]), the results to date have been equivocal. To overcome this shortcoming, we developed, and tested in 4 studies, a novel paradigm (the *NEXT paradigm*) that isolates IBR effects even prior to first task execution. In each miniblock, participants received S-R mapping instructions for a new task. Prior to implementing this mapping, responses were required to advance through screens during a preparatory (NEXT) phase. When the NEXT response was incompatible with the instructed S-R mapping, interference (IBR effect) was observed. This NEXT compatibility effect and performance in the implementation (GO) trials barely changed when prior practice of a few trials was provided. Finally, a manipulation that encouraged preparation resulted in relatively durable NEXT compatibility effects (indicating durable preparatory efforts) coupled with improved GO performance (indicating the success of these efforts). Together, these findings establish IBR as a marker of instructed proactive control.

Keywords: prepared reflex, instructions, intention, working memory, choice reaction time

Consider a scenario in which the police chase a group of criminals, and one policeman instructs the other regarding the look of the criminals, the direction in which they escaped, and how she or he should react to them. This scenario provides an example for the kind of complex collaborative efforts that require people to instruct one another on the fly. In such instances, performance following the instructions must be highly efficient, despite the lack of opportunity to practice these instructions. Less dramatic situations that require adherence to online instructions are very common, including team sports, when given driving directions, and so forth. In the present work, we asked how efficient can performance be when it is based on instructions alone.

The psychological literature includes thousands of articles showing that human performance can become highly efficient by means of extended practice. Moreover, many of these articles show that as practice continues, task execution becomes so efficient that it is carried out effortlessly and automatically. If performance that is based on instructions alone can be similarly effective, this would imply that humans (perhaps unlike other species; e.g., Baumeister 2005; Herrmann Call, Hernandez-Lloreda, Hare, & Tomasello, 2007) are endowed with a powerful means for collaboration. Indeed, recent empirical (reviewed below) and theoretical work (Bugmann, 2012; Cole, Laurent, & Stocco, 2013; Huang, Hazy, Herd, & O'Reilly, 2013; Ramamoorthy & Verguts, 2012) has started dealing with this long neglected but highly important ability, which is sometimes termed *rapid instructed task learning* (i.e., Cole et al., 2013).

In the present work, we studied the ability to reach high efficiency in relatively simple two-alternative choice tasks in which stimuli were mapped to responses. In describing these tasks, we refer to *stimulus–response* (S-R) associations as a generic term, without committing ourselves to the exact form of the association (e.g., between concrete stimuli and responses, between abstract stimulus categories and responses, and so forth).

We examined the hypothesis that the S-R association becomes immediately operative when given instructions about how to perform a choice task, assuming that sufficient motivation and processing resources are available. By *immediately*

Nachshon Meiran, Maayan Pereg, and Yoav Kessler, Department of Psychology and Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev; Michael W. Cole, Department of Psychology, Washington University, and Center for Molecular and Behavioral Neuroscience, Rutgers University; Todd S. Braver, Department of Psychology, Washington University.

This research was supported by a research grant from the USA–Israel Bi-National Science Foundation, awarded to Nachshon Meiran and Todd S. Braver. We thank Florian Waszak for a stimulating discussion that was instrumental in generating this line of research.

Correspondence concerning this article should be addressed to Nachshon Meiran, Department of Psychology, Ben-Gurion University of the Negev, P.O. Box 653, Beer-Sheva, Israel 84105. E-mail: nmeiran@bgu.ac.il

operative, we mean that, once the association is actively represented in working memory (WM), the response is selected and perhaps even initiated with no additional effort once the stimulus is presented.

The Prepared Reflex (PR)

The hypothesis that we examine is akin to one of the oldest hypotheses in the history of psychology. Specifically, Exner (1879, as cited in Hommel, 2000) described a state of high readiness as one in which, when a stimulus is presented, the prepared response is executed with very little additional involvement of "the will" (or, in modern terms, attention and cognitive control). Woodworth (1938; see also Cattell, 1886; Logan, 1978) termed this state "prepared reflex" (PR). The PR indicates a high-level preparedness toward an upcoming action that is based on an intention to perform it. When the intention to carry out a future action remains active, this readiness state is maintained (at least) until the action is performed. When in this state, the relevant action is triggered autonomously by a relevant stimulus in a ballistic, reflexlike manner, even when the action is unwarranted. The PR idea has been influential (sometimes indirectly) in diverse areas of psychological science, including planning and motivation (e.g., Gollwitzer, 1999), human performance (e.g., Logan, 1978), and the concept of mind-set (Gibson, 1941; Langer, 2000; Luchins, 1942). We (Meiran, Cole, & Braver, 2012) have recently described an action mode, related specifically to instructed S-R rules and called it "intention-based reflexivity" (henceforth, IBR). We chose the term *reflexivity* to distinguish this mode from "automaticity" that has been described as resulting from extensive practice (see more below). Unlike PR, which may also involve long-term memory (LTM), IBR places a clear emphasis on the fact that behavior results from preparing to execute an action on the basis of instructions alone and does not stem from LTM-based knowledge that has been acquired by prior task executions. In that regard, Braver and colleagues (Braver, 2012; Braver, Gray, & Burgess, 2007; Braver, Paxton, Locke, & Barch, 2009) distinguished between proactive and reactive control modes. These two modes were so far primarily distinguished by brain activation dynamics. Meiran et al. have argued that, perhaps paradoxically, proactive control, which emphasizes *future events*, is associated with loss of online control seen in the *immediate events* that precede the future events toward which one is prepared. For example, a police person who anticipates a criminal might accidentally shoot an unlucky civilian who happened to be where the criminal was expected. The present work provides a behavioral marker of IBR, which is therefore also a marker of proactive control.

Before reviewing the literature on IBR, it is important to distinguish it from a related phenomenon, the commission errors seen in prospective memory tasks. In these tasks, participants perform an ongoing task (such as lexical decision) and are told to additionally execute the prospective memory task (e.g., press a special key) on rare occasions when specific target stimuli appear. Recent work indicates that participants erroneously execute the prospective memory task (commission errors) after it had been declared as irrelevant (e.g., Walser, Fischer, & Goschke, 2012). Importantly, Bugg and Scullin (2013) recently showed that commission errors are most frequent when the prospective memory task has been canceled before having been executed. This finding, similar to IBR, suggests that the intention to execute a task in the future causes reflexivity. The core difference between commission errors and IBR is that the former pertain to a task that is no longer required, whereas the latter pertain to a pending task that is about to be executed.

Prerequisites

We suggest that, despite its high relevance, previous demonstrations of IBR may be open to some alternative explanations. To support this argument, we set a list of prerequisites that need to be met in order to convincingly rule out these alternative accounts. The prerequisites that we set are (a) reflexive responding; (b) no involvement of LTM of prior (overt) task executions; (c) demonstration that reflexive processing involves S-R translation, namely, that it leads to response choice; (d) separate (and independent) measurements of IBR and task performance, which enables examination of (e): demonstration that IBR contributes to task preparation.

The state of action preparedness described as IBR resembles in many ways a state of highly skilled performance, as described by contemporary theorists of skill-based automaticity (see Moors & De Houwer, 2006, for a recent review). However, to maintain the distinction between IBR and skill-based automaticity, we refer to the core prerequisite of IBR as *reflexivity* (rather than *automaticity*). Nevertheless, reflexivity of a process involves two key features typically attributed to skill-based automaticity: efficiency and autonomy. Efficiency means being quick and error free (e.g., Anderson, 1982; Rosenbloom & Newell, 1986). Autonomy (or uncontrollability; Bargh 1992) means being run to completion without guidance or monitoring. Theorists, including Bargh (1992) and Tzelgov (1997), have suggested autonomy as the minimal diagnostic criterion of automaticity. Consequently, we also adopt *autonomy as the first prerequisite for diagnosing IBR.*

The goal of distinguishing IBR from skill-based automaticity motivates the second prerequisite for IBR, discussed next. Specifically, skills, as described in contemporary literature, are acquired through repeated overt task execution, which results in LTM modification either through a change in processing efficiency (Anderson, 1982; Rosenbloom & Newell, 1986; Shiffrin & Schneider, 1977) or through the accumulation of episodic traces (Logan, 1988, 1992).

Obviously, preparation for action is not limited to newly instructed tasks because one could be highly prepared to execute a skilled task. In fact, much of the early evidence suggestive of IBR comes from experiments in which the task has been repeatedly executed (see Hommel, 2000, for a review), such as the flanker-noise paradigm (Eriksen, 1995) and the backward compatibility effect (Hommel, 1998). In all these cases, task rules are shown to operate autonomously, but the repetitive task execution makes it *possible* that performance is driven by the retrieval of episodic traces of prior task execution (see, e.g., Kiesel, Wendt, & Peters, 2007; Meiran & Kessler 2008, for evidence coming from task switching). Thus, to show that IBR is not just a reflection of skill-based automaticity, our *second prerequisite is that unequivocal evidence for IBR must be obtained when episodic retrieval of prior (overt) task executions is reasonably ruled out*. Our approach in dealing with that prerequisite is to focus on novel tasks that had never been *overtly* executed beforehand. Of course, processing the instructions in preparation for task execution may involve covert task execution. Actually, such covert task execution may be the means to encode the new instructions into WM. We deal with this issue in Experiment 3.

The third prerequisite is showing that autonomous processing involves S-R translation that leads all the way until (at least) response activation. By *response activation*, we mean the choice of the *instructed response* possibly including its motor programming. This prerequisite follows from the definition of IBR relating it to instructed S-R rules. Consequently, what should be shown to become autonomous is the translation of stimulus information into response information. This prerequisite rules out prior demonstrations that attentional direction becomes reflexive as a result of attentional setting (Folk, Remington, & Johnston, 1992; see additional review in Cohen-Kdoshay & Meiran, 2007). We suggest that these demonstrations cannot be taken as evidence for IBR (except perhaps in cases in which the required "response" is to direct attention). In other words, the evidence should show that the presentation of the stimulus (following the instructions, of course) has led to the retrieval of response information.

The fourth and fifth prerequisites are related to one another. Specifically, *the fifth prerequisite is that IBR contributes to successful task execution*. Although all of the prior experiments testing IBR were based on the instruction to prepare to execute a given task, this in itself does not constitute a strong enough demonstration that IBR *contributes* to successful task preparation. Alternatively, one could argue that the instruction resulted in some passive memory representation, which did not guide task execution but still gave rise to IBR (see, e.g., Liefooghe, Wenke, & De Houwer, 2012, Experiment 2). One way to show the required link is to demonstrate that IBR predicts successful task execution. It seems to follow that, in order to meet the fifth prerequisite, one needs to meet *the fourth prerequisite that IBR and task performance should be separately measured* (henceforth, "*separate measurement*").

Review of Existing Methods

A few groups of researchers (including ours) have continuously tried to refine a paradigm that would meet the prerequisites listed above by using choice reaction time (RT) tasks. In this brief review, we deal only with Prerequisites 1–4 (see Table 1), given the fact that only two published articles had so far dealt with Prerequisite 5 (see Experiment 4). Actually, if a study meets Prerequisite 4, this means that the paradigm that was adopted made it possible to also address Prerequisite 5.

In describing these prior findings on IBR effects, we adopt Liefooghe et al.'s (2012) terminology distinguishing between an *inducer task* and a *diagnostic task*. The inducer task is a novel task for which instructions were provided; the task that participants are presumably ready to execute for the first time. The diagnostic task is another task, usually a task that has been repeatedly executed in the course of the experiment, which is executed while the inducer task is still pending and is used to detect reflexivity. The first three paradigms used this inducer-diagnostic design in which the sequence of critical events included (a) instructions for the inducer task, (b) performance of the diagnostic task, and (c) performance of the inducer task.

In one approach (e.g., Kunde, Kiesel, & Hoffmann, 2003; see also Kunde, Elsner, & Kiesel, 2007, for a review), the instructions map *categories* to responses (e.g., "odd numbers go with the left key") instead of mapping *specific stimuli* to responses (e.g., "the number 3 goes with the left key"). This made it possible to study performance when a stimulus such as the digit *3* appeared in the experiment for the first time. In this case, performance is unlikely to be guided by LTM traces in which a specific stimulus (e.g., *3*) is linked to a response (e.g., "press the left key"). However, given the fact that a single task is used in the entire experiment, nearly all the results consist of trials in which the task (rather than the specific response to the digit *3*) has been executed beforehand. This implies that LTM traces linking *categories* (e.g., "odd number") to responses (Meiran & Kessler, 2008; Pashler & Baylis, 1991) could have been formed beforehand.

In De Houwer, Beckers, Vandorp, and Custers's (2005, Experiment 2) study, the diagnostic task required using specific response utterances ("bee" or "boo") to indicate the color classification of presented stimuli. Colored stimuli appeared to the left or right, but this was an irrelevant dimension. The inducer task was a right/left location judgment task, to be performed when the stimulus was colorless, in which location was also indicated with the same two response utterances. Unbeknownst to the participants, the inducer task was just instructed but was never actually performed. The results indicated quicker compatible than incompatible color responses. Namely, if a colored stimulus appeared on the right, for

Table 1
Existing Techniques and Prerequisites

Variable	Prerequisite/criterion 1 Autonomy	 2 LTM	 3 Response	 4 Sep. measurement
Kunde et al. (2003)	✓	—	✓	—
De Houwer et al. (2005)	✓	—	✓	Possible
Wenke et al. (2007, 2009)	✓	✓	—	✓
Liefooghe et al. (2012, 2013)	✓	Possible	✓	✓
Cohen-Kdoshay & Meiran (2007, 2009)	✓	✓	—	—
NEXT	✓✓	✓	✓	✓

Note. LTM = ruling out long-term memory involvement; response = showing reflexive response activation; sep. measurement = separate measurement of intention-based reflexivity and task performance; a dash indicates that the prerequisite is not met; two check marks indicate that the criterion is exceptionally well met. See the main text for details.

example, the reaction to the color of this stimulus was quicker if the required response-utterance was the same as that required for the right position than if it were the response-utterance associated with the left position. Thus, the results indicate autonomous processing of the inducer task. Unfortunately, although the inducer task was never *explicitly* required, it might have been erroneously (and latently) executed in compatible trials, in which the response is identical to that required in the inducer task. Such latent execution could have led to the formation of LTM traces and to skill-based automaticity. This possibility is especially worrisome given theoretical claims (Logan, 1988) implying that the formation of even a single LTM trace *might* lead to LTM-based response retrieval.

A similar criticism applies to two recent studies by Liefooghe et al. (2012; Liefooghe, De Houwer, & Wenke, 2013). In these studies, the inducer task was novel in each block and linked specific attributes of letter stimuli to right/left keypress responses. While being prepared to execute the inducer task, participants were asked to execute the diagnostic task (font classification) for a few trials, also using left/right keypresses. The results indicated quicker font classification in trials in which the inducer task's response was the same (compatible) than when it was different (incompatible). However, the researchers averaged performance across all the trials of the diagnostic task, leaving open the possibility that the inducer task was latently executed (and formed LTM traces) in the congruent trials. This possibility can be ruled out by analyzing just the first trial of the diagnostic trial in each block, something that remains to be done.

In Wenke and colleagues' (Wenke, Gaschler, & Natttkemper, 2007; Wenke, Gaschler, Natttkemper, & Frensch, 2009) studies, the inducer and the diagnostic task were executed only once in each miniblock, thus ruling out potential LTM involvement. The inducer task mapped two letter stimuli to right/left keypresses. In the diagnostic task, the two letters were presented in different font sizes (or in different colors), and participants were required to indicate the right-versus-left position of the larger letter (or its color). Incongruent trials were those in which the letter position in the pair was incongruent with how the letters were mapped to the right and the left positions in the instruction. Unfortunately, this compatibility effect could reflect the (mis)match between the letter positions in the diagnostic task and the letter positions in the instructions, rather than, or in addition to, the responses themselves. This interpretation implies that what has become autonomous is not response activation but stimulus processing, thus failing to meet the third prerequisite.

Cohen-Kdoshay and Meiran (2007, 2009) did *not* use the inducer-diagnostic design (and thus used a task that fails to meet Prerequisite 4 and by extension also Prerequisite 5). Their focus was on compatibility effects in the flanker paradigm (Eriksen, 1995), which were induced purely by novel task instructions. Thus, unlike the usual flanker paradigm, in which there is a single set of instructions that is first practiced and then executed throughout the entire experiment, Cohen-Kdoshay and Meiran introduced a new set of instructions in every block. This made it possible to accumulate enough first trials that immediately follow the S-R instructions (namely, enough first trials from each block).

In detail, in the flanker paradigm, participants respond to a central target that is flanked by noise stimuli. Participants were given a new set of category response mappings in every miniblock (e.g., respond left if stimulus is a letter from the beginning of the alphabet, right if a letter from the end of the alphabet). The core finding was the "first trial flanker compatibility effect," indexing IBR. Thus, the response to an incompatible flanker stimulus (e.g., "WBW") was slower relative to a compatible one (e.g., "BAB") even on the first trial after receiving instructions, thus ruling out any involvement of LTM retrieval. A serious shortcoming of these studies is that the first-trial flanker compatibility effect could reflect automatic influences on stimulus encoding rather than on response selection/activation (thus failing to meet Prerequisite 3). Specifically, it is conceivable that, in compatible stimuli such as "BAB," the flankers ("B") have primed target (*A*) processing because of the common semantic category ("beginning of the alphabet"). In other words, the first-trial flanker compatibility effect may not reflect reflexive *response activation*, but rather semantically primed stimulus processing. This is especially true given the fact that the stimuli were presented in the instruction phase as linked to this semantic category and had thus primed the association between the stimuli and their semantic category. In fact, the semantic priming phenomenon (see, e.g., Neely, 1991, for a review), which presumably indexes facilitated stimulus processing, was originally demonstrated with very similar stimuli (e.g., Meyer, Schvaneveldt, & Ruddy, 1975).

The Present Study

To summarize, we argue that previous demonstrations of IBR may be open to alternative accounts. Thus, the main goal of the present study was to provide a demonstration of IBR that would meet all the five prerequisites that we listed. To this end, we used a new paradigm, termed the *NEXT paradigm*. Other than meeting the prerequisites, the NEXT paradigm makes it possible to track the unfolding of IBR at the level of individual trials, a feature that has proved to be highly informative (see especially Experiment 4).

The NEXT paradigm (see Figure 1) consists of miniblocks. Each miniblock begins with instructions for the inducer task, a novel two-alternative forced-choice task in which two new stimuli (which were never used in the experiment beforehand) are mapped to the right/left response keys. After the instructions, the stimuli are presented in two phases: (a) In the NEXT phase, which serves

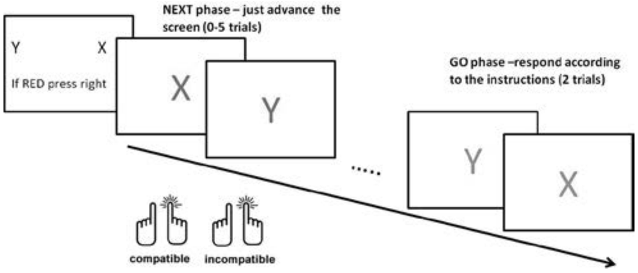


Figure 1. The NEXT paradigm: Participants (a) received instruction regarding the GO phase (first screen), (b) advance the screen during the NEXT phase (when the targets appear in RED), and (c) execute the instructions in the GO phase (when the targets appear in GREEN). NEXT phase screen-advance responses can be compatible (the first NEXT trial with the target "X") or incompatible (the second NEXT trial with the target "Y") with the instructed GO phase response. See the online article for a color version of this figure.

as the diagnostic task, the stimuli appear in red color, indicating that responding according to the instructions should be withheld. At an unexpected point in time, (b) the GO phase begins, in which the instructed mapping has to be applied. The beginning of the GO phase is indicated when stimuli appear in green.

Participants used one of the keys (the same key was used throughout the experiment) to advance the screen during the NEXT phase. RT in these screen advancement responses was the primary dependent variable. Critically, the response key used to indicate NEXT responses was the same as one of the keys used in the inducer task. This enabled differentiation of compatible and incompatible NEXT responses. Compatible responses are those in which the S-R instructions indicated the same key as that used to advance the screen. Incompatible responses are those in which the S-R instructions indicated the opposite key. Although the NEXT compatibility effect in the first NEXT trial provides the least equivocal evidence for IBR, the results from Experiment 3 suggest that the NEXT compatibility effect seen in the advanced NEXT trials can also be used.

The NEXT compatibility effect meets all the aforementioned prerequisites to demonstrate IBR. Specifically, it meets Prerequisite 1 (reflexivity) by showing that the instruction information is processed autonomously. It meets Prerequisite 2 (ruling out episodic LTM retrieval of prior task executions) because autonomous processing is seen before the instruction was ever (overtly) executed, and because it is seen even in the first trial after the instruction. It meets Prerequisite 3 (involvement of response activation) because what differentiates between compatible and incompatible trials is the S-R rule. Because the NEXT paradigm adopts the inducer-diagnostic design (especially Liefooghe et al., 2012), it made it possible to separately assess IBR (seen in NEXT responses) and task execution (seen in GO responses), and it thus meets Prerequisite 4. In Experiment 4, we capitalize on this feature to show that the IBR is linked to task preparation so that the paradigm also meets Prerequisite 5.

When presenting the results of the first two experiments, we focused on the NEXT phase. The analyses of the GO responses in these experiments are reported in the Appendix. The correspondence between the NEXT phase and the GO phase was examined in Experiment 4.

Experiment 1

To ensure that participants maintain high readiness to execute the GO trials throughout the NEXT phase, the transition between the phases took place at an unexpected point in time. Furthermore, readiness had to be high because there were only two opportunities to execute the GO task before the miniblock ended and a new one (involving a new task) began.

To ensure that the GO phase appeared at an unexpected point in time, we varied the length of the NEXT phase. First, there were miniblocks with no NEXT trials to ensure that participants would be ready to execute the GO task immediately after the instruction. Second, we were concerned with the fact that equally probable NEXT phase lengths enable temporal predictability. Specifically, when using rectangular length distribution, the *conditional* probability that the GO phase is imminent increases with the progression of the NEXT phase.[1] We partly controlled for this problem by monotonically decreasing the probability with increasing NEXT phase length. Specifically, 10%, 30%, 20%, 20%, 10%, and 10% of the miniblocks had NEXT phase length of zero to five trials, respectively.

Method

Participants. Twenty Ben-Gurion University of the Negev students participated in the experiment in return for course credit (18 women, mean age = 22.85, *SD* = 0.87). All the participants reported having normal or corrected-to-normal vision, including intact color vision, and not having diagnosed attention deficits.

Materials and procedure. The experiment was run individually. Participants were presented with the stimuli on a 17-in. monitor controlled by a desktop computer, with software written in E-Prime 2.0 (Psychology Software Tools, 2010). There were 220 stimuli, made of 26 English letters, 10 digits, 24 Hebrew letters,[2] 20 symbols (e.g., arithmetic symbols), and 140 pictures (e.g., shapes and different objects). The letters, symbols, and most of the pictures came from Microsoft PowerPoint symbols pool, and the rest of the pictures were sketches drawn from free Internet image search bases. The size of the stimuli was 3 × 3 cm; digits and letters appeared in a Calibri font. The two new stimuli that were chosen in each choice task came from the same stimulus group (e.g., two digits, two pictures). Otherwise, the stimuli were chosen pseudorandomly without replacement. Each stimulus was used only once in the course of the experiment.

The paradigm consisted of 110 two-choice tasks. Each choice task involved two stimuli that were arbitrarily mapped to a right and left key (*L* and *A* on a QWERTY keyboard, respectively). Each miniblock consisted of an instruction screen for the new choice task. This screen was presented until the participants pressed the spacebar, but not sooner than 3 s. It was followed by a NEXT phase of variable number of trials (see below), then a GO phase that consisted on only two trials, and finally a feedback screen reporting accuracy and RT in the GO phase.

In the instruction screen, two stimuli were presented in *white* color, one on the right and the other on the left (each stimulus center was placed 15.5 cm from the center of the screen). This indicated that the stimulus on the right was mapped to the right response key (*L*), whereas the stimulus on the left was mapped to the left response key (*A*). The participants were required to place their fingers on the response keys and be ready to execute the GO task. In order to maximize their motivation, the participants were told that the two participants exhibiting the best (GO) performance will get bonus credit points or additional payment.

The NEXT phase preceded the GO phase. The phases were made visually discriminable by means of the color in which the stimulus appeared. If the stimulus was presented in *red* color, this indicated a NEXT trial, requiring a NEXT response. If the stimulus was presented in *green* color, this indicated a GO trial, which

---

[1] For example, with equally probable NEXT phase lengths (zero through five), the conditional probability that there would not be any NEXT trials (NEXT−length = 0) would be 1/6. However, the probability that the NEXT phase will end after one trial would be 1/5; the probability the NEXT phase will end after two trials, given that it did not end after one trial, would be [1/4], and the like.

[2] There are only 22 letters in the Hebrew Alphabet, but some of the letters have a different shape when coming at the end of the word, a fact that enabled us to slightly extend the number of stimuli.

required highly accurate and quick responding according to the instructions (see Figure 1). In the NEXT phase, participants needed to avoid performing the two-choice GO task and were asked to simply press a given key to advance the screen (the NEXT key; this key remained fixed throughout the entire experiment, with half the participants using the *A*/left and half the *L*/right). Critically, however, the NEXT key overlapped with one of the GO responses, although this overlap was incidental from the participant's perspective. On each NEXT or GO trial, a 500-ms fixation preceded the target stimulus. This target was then presented until the participant responded, after which a black screen was shown for an 800-ms intertrial interval.

When the NEXT phase ended, the GO phase immediately started, and there were only two GO trials, after which the next miniblock began. As noted, some NEXT responses overlapped with the (planned) GO responses (compatible), and some were different from the (planned) GO responses (incompatible).

Before the experiment proper, there was a brief familiarization phase that was included to ensure that participants get adjusted to the structure of the experiment. In this familiarization phase, participants were given three miniblocks, each involving a novel set of S-R instructions. To expose the participants to the variability of the NEXT phase length, each one of the miniblocks in the familiarization phase had a different NEXT length (zero, one, and five NEXT trials). This brief familiarization stage was followed by the experiment proper that included 11 identical task blocks, each consisting of 10 miniblocks (110 miniblocks in total). In order to further ensure high alertness, participants were asked to get up from the chair, walk a little, and return to the chair between the blocks.

Data analysis. The first miniblock in the session as well as GO trials with an error were omitted from all analyses. Trials with an RT shorter than 100 ms or longer than 3,000 ms were also not analyzed.

Results and Discussion

NEXT phase. Given the fact that NEXT responses did not involve a choice, there were no errors and only RTs were analyzed. These results were submitted to a two-way analysis of variance (ANOVA), with compatibility (compatible-incompatible) and NEXT trial (one through five) as within-subjects independent variables. The main effect of NEXT trial, $F(4, 76) = 21.75$, $p < .001$, $MSE = 4{,}079.85$, $\eta_p^2 = .53$, was significant (indicating slowest first NEXT trial). More importantly, the main effect of compatibility was also significant, $F(1, 19) = 13.99$, $p = .001$, $MSE = 1{,}506.73$, $\eta_p^2 = .42$, indicating the predicted NEXT compatibility effect. The two-way interaction was *not* significant, $F(4, 76) = 1.25$, $p = .297$, $MSE = 994.45$, $\eta_p^2 = .06$. As seen in Figure 2, the NEXT compatibility effect was numerically largest in the first NEXT trial (36 in the first NEXT vs. an average of 17 ms in the following NEXT trials), and it was significant already in the first trial, $F(1, 19) = 15.38$, $p < .001$, $MSE = 828.69$, $\eta_p^2 = .45$, which provides the cleanest demonstration of IBR, because this is the very first time a key was pressed after the instructions. Subsequent NEXT trials could have *potentially* been influenced by prior NEXT responses.

Although we have ruled out any involvement of skill acquisition, we wanted to see whether the NEXT compatibility effect increases as participants gain skill with the paradigm (rather than with a particular choice task). Such an increase would indicate that at least part of the compatibility effect is skill based. To address this point, we divided the results of the entire experiment into roughly equivalent parts, corresponding to the initial, middle, and last part of the experiment (each third containing 35–36 miniblocks). NEXT RTs were analyzed in a two-way ANOVA with the within-subjects variables progress (one through three) and compatibility. We omitted the NEXT trial number (one through five) variable because this variable did not interact with compatibility. The compatibility main effect, $F(1, 19) = 11.64$, $p = .003$, $MSE = 1{,}040.41$, $\eta_p^2 = .38$, and the progress main effect, $F(2, 38) = 70.9$, $p < .001$, $MSE = 936.4$, $\eta_p^2 = .79$, were significant. Nonetheless, the interaction was *not* significant, $F(2, 38) = 0.82$, $p = .449$, thus further ruling out experimentwide practice effects.

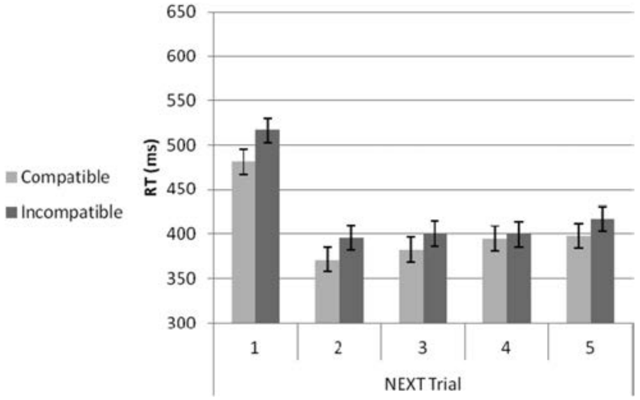


Figure 2. Experiment 1. NEXT RT (in milliseconds) as a function of NEXT trial and compatibility. Error bars represent within-subject confidence intervals (Hollands & Jarmasz, 2010; Jarmasz, & Hollands, 2009). RT = response time.

To summarize, the key finding in this experiment is the predicted NEXT compatibility effect. This effect was present already in the first NEXT response, where it was numerically the largest.

Experiment 2

In Experiment 2, we addressed a troubling alternative explanation. Specifically, the NEXT compatibility effect could reflect the erroneous yet intended (but covert) execution of the GO task in *compatible* NEXT trials. Note that the notion of autonomous processing implies that a process is being carried out in spite of the instructions to execute another task (or not executing any task, for that matter). If, however, participants have erroneously encoded the stimulus color in the NEXT phase as green (indicating GO) and have thus erroneously committed the GO task, this implies that, for them, processing the *GO* task was required and, in that sense, was not autonomous. As noted by Bargh (1992) and Tzelgov (1997), for processing to be regarded as autonomous, it should take place in spite of the instruction to carry out another process.

Importantly, the aforementioned alternative account holds only for *compatible* NEXT trials. Thus, it can be ruled out by showing that the NEXT compatibility effect is predominantly (or also) caused by the *incompatible* trials, for which the account does not apply.

In Experiment 2, we tackled this issue by adding a neutral condition in which the stimulus that was presented in the NEXT phase was not connected to any response. This neutral condition made it possible to separate the NEXT compatibility effect into two components: compatibility-related facilitation, caused by the *compatible* trials (neutral RT minus compatible RT) and incompatibility-related interference, caused by the *incompatible* trials (incompatible RT minus neutral RT).

The use of neutral conditions was criticized before (Jonides & Mack, 1984), especially because it is very difficult to create conditions that are truly neutral. We therefore made a serious attempt to ensure that this neutral stimulus is equivalent to the (in)compatible stimuli in every other respect. First, the stimuli that were used as neutral stimuli were the same stimuli (averaged across participants) as those used in the (in)compatible conditions given their random selection. Nonetheless, the stimuli in the (in)compatible condition could have been more familiar given the fact that they were introduced in the instructions. To equate the stimuli in this regard, we presented the neutral stimulus in the instruction phase just like the (in)compatible stimuli.

To summarize, IBR is most clearly indicated by incompatibility-related interference, because (a) such interference indicates that the stimulus has activated a response that has not yet been executed and (b) this could not be regarded as intentional (yet covert) erroneous execution of the GO instructions, manifested in a NEXT facilitation effect. We therefore predicted that there would be a significant incompatibility-related interference component of the NEXT compatibility effect.

Method

Participants. Forty-five individuals (26 women, mean age = 23.86, *SD* = 1.97), similar in attributes to participants from the previous study, took part in the experiment in return for course credit or for monetary compensation (25 NIS, ~$7 U. S.). Three participants were excluded from the analyses due to unusually high error proportion (PE) in the GO phase, indicating that they did not perform the GO task as instructed (PE = .18–.25 as compared with .00–.13 in the remaining participants). The increase in the number of participants, compared with Experiment 1, reflects the additional statistical power required to further separate the NEXT compatibility effect into (smaller) components.

Materials and procedure. The procedure was the same as in Experiment 1, aside from the inclusion of a neutral condition in the NEXT phase, in which a third stimulus was introduced during the instructions (after providing relevant S-R mappings). This neutral stimulus was not linked to any GO trial response, and participants were informed that this stimulus would only appear in the NEXT phase (in *red*). The neutral stimulus belonged to the same stimulus category as the two targets (e.g., all three were digits). The purpose of introducing this stimulus in the instruction phase was to match preminiblock familiarity for all three stimuli used in the subsequent miniblock. Because we used the same pool of pictures as before, this lead to a reduction in the number of miniblocks to 70 (seven blocks).

Results

NEXT phase. RTs were analyzed in a two-way ANOVA with the within-subjects independent variables compatibility (compatible-neutral-incompatible) and NEXT trial (one through five). The main effect of NEXT trial, $F(4, 140) = 26.44$, $p < .001$, $MSE = 3{,}535.01$, $\eta_p^2 = .43$, as well as the main effect of compatibility, $F(2, 70) = 7.90$, $p < .001$, $MSE = 1{,}382.51$, $\eta_p^2 = .18$, were significant. Mean NEXT RT was 399, 387, and 385 ms in incompatible, neutral, and compatible conditions, respectively, indicating incompatibility-related slowing of 12 ms and compatibility-related facilitation of 2 ms. The two-way interaction was not significant, $F(8, 280) < 1.00$, $p = .512$, indicating, as in Experiment 1, that the compatibility effect did not significantly change during the NEXT phase.

The compatibility main effect was first probed by means of two independent comparisons. The first comparison indicated a significant difference between the incompatible condition, on the one hand, and the compatible and neutral conditions, on the other hand, $F(1, 35) = 15.01$, $p < .001$, $MSE = 1{,}451.53$, $\eta_p^2 = .30$. The second comparison showed a nonsignificant difference between the compatible condition and the neutral condition, $F(1, 35) = 0.05$, $p = .822$.

Because we were especially interested in incompatibility-related slowing, we also applied a more standard set of nonorthogonal contrasts. The first contrast (between incompatible and neutral trials) indicated a significant incompatibility-related slowing, $F(1, 35) = 10.30$, $p = .003$, $MSE = 1485.65$, $\eta_p^2 = .23$. The second contrast (between neutral and compatible trials) has already been reported to be nonsignificant in the previous paragraph.

We conducted an ANOVA on NEXT RT, with compatibility and progress (each third containing 23–24 miniblocks) as independent variables (see Experiment 1). The compatibility main effect was significant, $F(2, 82) = 15.13$, $p < .001$, $MSE = 815.24$, $\eta_p^2 = .27$, as was the main effect of progress, $F(2, 82) = 73.68$, $p < .001$, $MSE = 2{,}3632.85$, $\eta_p^2 = .64$. Unlike in Experiment 1, the interaction was also significant, $F(4, 164) = 2.56$, $p = .040$, $MSE = 751.31$, $\eta_p^2 = .05$, indicating a decrease in the NEXT compatibility effect with session progress (see Figure 3). Given that this was a decrease rather than an increase, the result does not suggest that the NEXT compatibility effect reflects increased LTM-based automaticity.

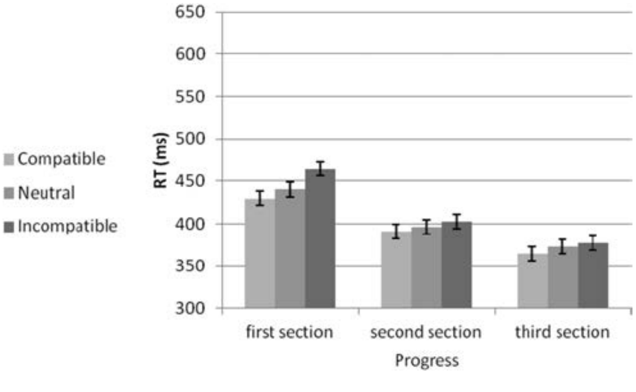


Figure 3. Experiment 2. NEXT RT (in milliseconds) as a function of compatibility and the progress in the session. Error bars represent within-subject confidence intervals (Hollands & Jarmasz, 2010; Jarmasz, & Hollands, 2009). RT = response time.

Discussion

IBR is most clearly indicated by incompatibility-related interference, because (a) such interference indicates that the stimulus has activated a response that has not yet been executed and (b) this could *not* be regarded as intentional (yet covert) erroneous execution of the GO instructions. Here, we demonstrated that the NEXT compatibility effect predominantly reflects incompatibility-related interference. Unlike in Experiment 1, the NEXT compatibility effect diminished during the course of the session.

Experiment 3

What do participants do when they prepare themselves covertly to execute the instructions? One possibility is that they rehearse the instructions, perhaps even verbally (see, e.g., van't Wout, Lavric, & Monsell, 2013, for evidence). Another possibility is that they execute the task in imagery. Regardless of the exact nature of this covert process, we wanted to compare its efficiency with *overt* short practice. Our rationale was that executing the task overtly for a few times should mimic, in a sense, the covert imagined execution. We have thus allowed participants to overtly execute the newly instructed task in half of the miniblocks and compared this condition with a condition that involved only covert preparation. In our comparison, we focused on both GO performance and reflexivity, as indexed in the NEXT compatibility effect.

Logan and Klapp (1991; see also Zbrodoff, 1999), who were interested in the preconditions of skill-based automaticity, studied performance in an alphabet arithmetic task in which participants verify the correctness of arithmetic equations in which letters replace digits. These authors showed that overt rehearsal of facts regarding this task (e.g., "B + 2=?" for which the answer could be "D") for about 15 min was as efficient as several sessions of practice in alphabet arithmetic in producing automaticity. Although very relevant to the present issue, Logan and Klapp's findings still involve overt (rather than covert) rehearsal. Moreover, their fact-retrieval practice was around 100 times longer than the few seconds it took participants to prepare to execute the instructions in the present experiments. As such, Logan and Klapp's study does not address our question. Just recall our policemen story in which the instructions must be immediately and efficiently applied. One characteristic of these situations is that they do not provide the luxury of prior practice, not even one that lasts a few minutes.

In the present experiment, in half of the miniblocks, participants executed the new instructions eight times before the NEXT phase began (prior practice condition) and went directly to the NEXT phase in the remaining miniblocks (no prior practice condition). The number of overt task executions that we chose (eight) was in the ballpark of the potential number of covert time executions that was permitted in the study phase, lasting a few seconds. We were interested in examining the influence of prior practice (with vs. without) on the NEXT compatibility effect and on GO performance. We reasoned that if covert rehearsal is as efficient as eight overt task executions, the NEXT compatibility effect and GO performance should be comparable in the two conditions.

Method

Participants. Twenty-seven individuals (18 women, mean age = 23.92, *SD* = 2.35), similar in attributes to participants from the previous experiments, took part in the experiment in return for monetary compensation (45 NIS, ~$12 U. S.) or course credit.

Materials and procedure. The procedure was the same as in Experiment 1 with the exception that half of the miniblocks (55) began with a short practice phase prior to the NEXT phase while there was no prior practice in the remaining miniblocks (55). Participants terminated the presentation of the instruction screen by pressing the spacebar, but no sooner than after 3 s had elapsed. In the miniblocks with practice, after the instructions, a screen appeared that indicated the beginning of the practice phase ("now let us practice"). The presentation of this screen was terminated by pressing the spacebar. The short practice phase included eight trials (four of each stimulus), in which the stimuli were colored in *white*, and a 400-ms beep tone was played if an error was made. After this practice, a screen announced the beginning of the NEXT/GO phase ("now let us begin the task"). This screen was presented immediately following the instructions in those miniblocks in which there was no prior practice. The presentation of this screen was terminated when participants pressed the spacebar.

After running two participants, we included in the E-Prime program a measurement of the time it took participants to terminate the instructions and the time it took them to terminate the presentation of the screen announcing the beginning of the NEXT/GO phase.

Results

Instruction study time. This information was available on 25 of the participants, who took 2,052 ms to terminate the instruction screen beyond the enforced minimum of 3 s. Thus, the total time it took participants to study the instructions was 5,052 ms on average. At this point, the participants did not know whether they would be practicing the instructions or move directly to the NEXT phase. This information became available immediately after.

One concern is that participants could have taken advantage of the fact that starting the NEXT phase was self-paced to compensate for the lack of prior practice. This hypothesis predicts that participants would take longer to start the NEXT phase if they did not practice the instructions beforehand. In fact, the time taken to terminate the screen announcing the NEXT phase was 1,106 ms without prior overt practice and 1,646 ms after overt practice ($p =$ .017). This result certainly rules out the aforementioned possibility given that the trend was in an opposite-to-expected direction.

To summarize, participants took about 5 s to rehearse the instructions, and there was no evidence that practice led to greater readiness to start the NEXT phase.

NEXT phase. RTs were analyzed in a three-way ANOVA with the within-subjects independent variables compatibility, NEXT trial (1–5), and prior practice (practiced vs. no practiced). The main effect of NEXT trial, $F(4, 92) = 30.51$, $p < .001$, $MSE = 10{,}862.58$, $\eta_p^2 = .57$, and the main effect of compatibility, $F(1, 23) = 17.62$, $p < .001$, $MSE = 11{,}871.09$, $\eta_p^2 = .43$, were significant. These main effects indicate a NEXT compatibility effect and the fact that the first NEXT trial was relatively slower. The two-way interaction between trial and compatibility was also significant, $F(4, 92) = 3.60$, $p = .009$, $MSE = 3{,}568.56$, $\eta_p^2 = .13$, indicating a decrease in the compatibility effect with the progress in the NEXT phase. The two-way interaction between trial and prior practice was significant as well, $F(4, 92) = 3.58$, $p = .009$, $MSE = 3{,}633.01$, $\eta_p^2 = .13$. The two-way interaction between compatibility and prior practice was not significant, $F(1, 23) = 1.99$, $p = .171$, $MSE = 5{,}549.30$, $\eta_p^2 = .08$. This result shows that the numeric trend for a larger NEXT compatibility effect in the practiced condition (51 ms) than without prior practice (32 ms) was nonsignificant.

This trend was also seen in the first NEXT trial and is where IBR is least equivocal, at least when there is no prior practice. In that trial, the NEXT compatibility effect was numerically larger with practice (81 ms) than without practice (66 ms), although this difference still did not approach significance, $F(1, 23) < 1.00$, $p = .503$.

The three-way interaction between prior practice, NEXT trial, and compatibility also did not reach significance, $F(4, 92) < 1.00$, $p = .503$ (see Figure 4).

GO phase. Both RT and PE were analyzed in three-way ANOVAs, with the within-subjects independent variables prior practice and NEXT length (zero through five), which indicates the number of NEXT trials preceding the GO trials, and GO trial (first vs. second). In none of these analyses was there a hint for an interaction involving prior practice. Nonetheless, the main effect for prior practice was significant in RT, $F(1, 26) = 8.10$, $p = .008$, $MSE = 2{,}206.09$, $\eta_p^2 = .24$, indicating quicker responses in the practiced condition (432 ms) relative to without prior practice (442 ms). In RTs, the GO trial main effect, $F(1, 26) = 107.12$, $p < .001$, $MSE = 17{,}958.35$, $\eta_p^2 = .80$, was significant, indicating, as in the previous experiments (Appendix), slower first than second GO trials. Additionally, there was a significant main effect of NEXT length, $F(5, 130) = 7.67$, $p < .001$, $MSE = 3{,}547.03$, $\eta_p^2 = .23$, which was qualified by a significant two-way interaction between NEXT length and GO trial, $F(5, 130) = 10.45$, $p < .001$, $MSE = 2{,}379.49$, $\eta_p^2 = .29$. This interaction indicates that the GO trial effect was larger in the absence of NEXT trials that preceded it (see Figure 5).

In PE, the only significant effect was the GO trial main effect, $F(1, 26) = 51.72$, $p < .001$, $MSE = 0.01$, $\eta_p^2 = .66$. None of the other interactions reached significance (see Figure 5).

Discussion

The present experiment shows that prior limited GO practice led to only a modest and nonsignificant increase in the NEXT compatibility effect in the first NEXT trial, but improved GO performance in general. This modest influence of task execution on the NEXT compatibility effect is corroborated by participants' informal reports that executing the eight practice trials distracted them and made the experimental task more difficult. In a sense, our finding echoes Logan and Klapp's (1991) result, but carries it one step further. Specifically, these authors showed that covert fact retrieval rehearsal can replace extended overt practice. We show that (limited) overt task execution barely influences IBR.

Experiment 4

In this experiment, we examined the hypothesis that the IBR effect appearing during the NEXT phase should be dependent on the degree of preparation for the GO phase. Specifically, we reasoned that if the NEXT compatibility effect reflects task readiness, then it should also predict GO performance. Previous studies have already shown correspondence between IBR (measured in the diagnostic task, the NEXT phase in this case) and task performance (measured in the inducer task, the GO phase in this case). Wenke et al. (2009) found that the IBR was smaller and no longer statistically significant when the requirement to execute the inducer task was frequently aborted. This manipulation presumably discourages preparation toward the inducer task. Liefooghe et al. (2012, Experiment 2) showed that evidence for IBR was eliminated when participants just had to memorize the instruction but were not asked to execute it. Liefooghe et al. (2013) further showed that evidence for IBR was eliminated when participants did not have to memorize the instruction for the inducer task (because it was presented again, just before the inducer task, Experiment 1). These authors additionally showed (Experiment 2) that IBR was more robust when response speed was emphasized for the inducer task. Unfortunately, all of the aforementioned studies suffer from shortcomings. In Wenke et al.'s (2009) paradigm, it is unclear whether the effect used to index IBR truly reflects response activation. In Liefooghe et al.'s (2012, 2013) paradigm, the authors averaged across all the trials of the diagnostic task and therefore did not

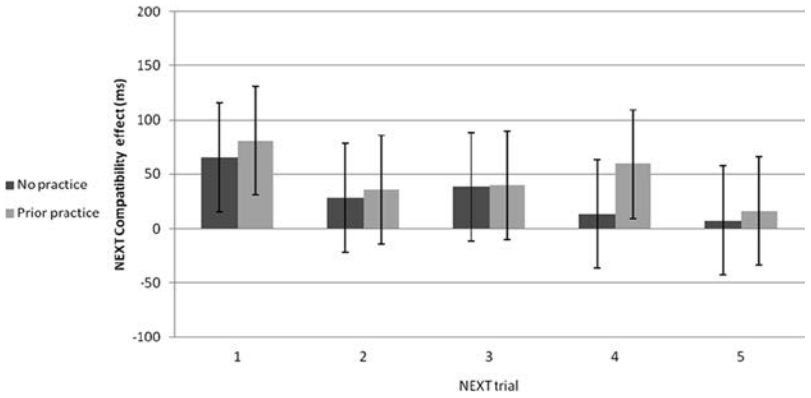


Figure 4. Experiment 3. NEXT compatibility effect (in milliseconds) as a function of prior practice and NEXT trial. Error bars represent within-subject confidence intervals (Hollands & Jarmasz, 2010; Jarmasz, & Hollands, 2009).

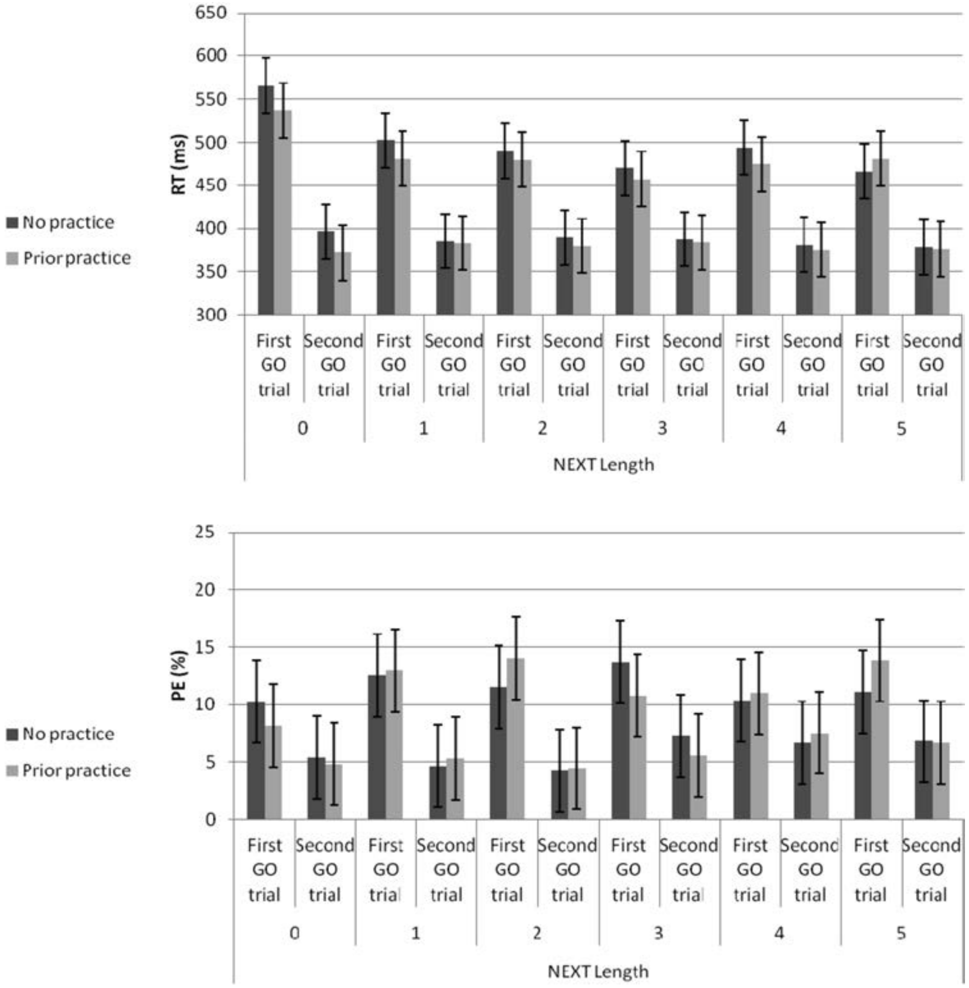


Figure 5. Experiment 3. GO phase RT (top panel) and PE (bottom panel) as a function of NEXT length, GO trial, and prior practice. Error bars represent within-subject confidence intervals (Hollands & Jarmasz, 2010; Jarmasz, & Hollands, 2009). RT = response time; PE = proportion of error.

examine how IBR is influenced by the passage of time/trials, a feature that has proved very important in the present experiment.

Examining how IBR develops in the course of the NEXT phase can tell us whether participants maintained their readiness throughout the phase (indicating proactive control). An important issue in this regard is whether we can assess the development of IBR during the NEXT phase. Arguably, the first NEXT trial is the only trial in which prior (full-blown) task execution could not contribute to the NEXT compatibility effect, and advanced NEXT trials may be contaminated by prior covert GO task execution (see Experiment 2). Nonetheless, the results of Experiment 3 suggest that prior (overt) task execution contributes only minimally to the NEXT compatibility effect. Moreover, this contribution was roughly equal across all the trials of the NEXT phase. Had there been covert execution of the GO task during the NEXT phase, it should have influenced the NEXT compatibility effects in advanced NEXT trials and should have reduced the slight difference between the prior practice conditions of Experiment 3. The fact that this did not happen further suggests that covert GO task execution does not influence the NEXT compatibility effect in advanced NEXT trials.

Given the above considerations, we conclude that the NEXT compatibility effect in the advanced NEXT trials can be taken as a reasonably valid index of IBR. This conclusion enabled us to track the development of preparation via the NEXT compatibility effect. For example, we could see whether preparation is sustained in the course of the NEXT run as reflected in a sustained NEXT compatibility effect.

Another noteworthy limitation of Liefooghe et al.'s (2012, 2013) and Wenke et al.'s (2009) studies with regard to our focus is in answering an important question of whether the IBR indicates the *degree of preparation* for an upcoming task. One reason why having a *graded* index of preparation relates to Braver's (2012) claim that proactive-versus-reactive control is more an issue of relative emphasis. We therefore reasoned that an all-or-none index is unlikely to detect changes in emphasis, and a *graded* index would be more suitable. Specifically, in some of the aforementioned experiments (Wenke et al., 2009, Experiment 1; Liefooghe et al., 2012, Experiment 2; Liefooghe et al., 2013, Experiment 1), the authors introduced a condition that strongly discouraged preparation; accordingly, the evidence for IBR was eliminated. Thus, these studies revealed conditions in which the IBR appeared to be present or absent, rather than conditions in which the IBR varied in magnitude. Only one experiment to date has shown evidence that the IBR effect may be graded. Specifically, Liefooghe et al. (2013, Experiment 2) showed larger IBR effects when participants were given a response deadline in the inducer task. This was coupled by shorter RTs in the inducer task (an effect of 74 ms), a result that indicates gradual rather than an all-or-none effect. Unfortunately, this RT pattern was accompanied by a nonsignificant, yet nontrivial opposite trend in PE (an effect of .05 in proportion units). Thus, the results may reflect a shift in speed–accuracy trade-off rather than a true effect on processing efficiency. This remains a viable option given the typical negatively accelerating speed-accuracy functions when PE is very low; shifts in speed-accuracy trade-off result in relatively large RT difference coupled by small PE differences (Wickelgren, 1977).

We also wanted to link between IBR and GO performance and asked ourselves which aspect of the *GO* performance is most suitable to reflect task preparation? We reasoned that the very first GO trial is a trial in which performance is entirely based on preparation. More advanced trials may reflect training (that took place in the first GO trial), or performance-based task setting. Regarding the latter, Rogers and Monsell (1995) suggested a "stimulus cued completion" hypothesis according to which seeing the task's stimulus is required to complete task setting. Although they attributed it to seeing the stimulus, in their experiments it was impossible to discriminate between seeing the stimulus and executing the task. Luria, Meiran, and Dekel-Cohen (2006), who studied tasks that required multiple responses, showed that the responsible factor is actual task execution.

In any event, the conclusion is that the difference between the first GO trial and the remaining GO trials (in the prior experiments, there was just one additional second GO trial) provides an index of the benefits (or costs) associated with preparation (or lack of it). We have labeled this difference *GO trial effect*. The prediction was therefore that states associated with better preparedness would reflect in larger NEXT compatibility effects (presumably reflecting preparatory effort) and smaller GO trial effects (reflecting the success of these efforts). As reported in the Appendix, the results so far are unclear. Briefly, the NEXT compatibility effect remained stable throughout the NEXT phase. Likewise, the GO Trial effect was not influenced by the length of the preceding NEXT phase. However, in Experiment 2, the NEXT compatibility effect diminished in size during the course of the experiment. Contrary to the prediction, the GO Trial effect also diminished in size during the course of the experiment. Nonetheless, this evidence is limited because it relies either on the acceptance of the null hypothesis (no effect of NEXT trial and NEXT length) or on an unexpected finding that was seen only in one experiment (the reduction in the effects during the course of the experiment, found only in Experiment 2). Thus, the goal of this experiment was to provide a more clear-cut test of the prediction that states associated with high preparedness are characterized by large NEXT compatibility effects and small GO trial effects.

We therefore introduced two (related) manipulations that were hypothesized to influence the motivation to *proactively* configure S-R translation rather than rely on a "wait-and-see" approach, which is primarily reactive (Braver, 2012), because of relying on performance monitoring and memory retrieval. We were specifically interested in motivational manipulations that would shift the emphasis toward proactive control and away from reactive control. Therefore, manipulations related to overall success, for example, were unsuitable for our purposes. The first manipulation was the length of the GO phase. The rationale was that GO performance can improve as a result of task execution when the GO phase is long. In such cases, a wait-and-see approach may be suitable because GO performance can improve by practice. However, when the GO phase is very short, performance in the instructed task must be successful from the outset, given the lack of practice opportunity. We began the experiment by familiarizing participants with short and long GO phases, and we also instructed them concerning the implications of this difference. In the experiment proper, there was a cue that was given immediately after the task instruction, indicating whether the upcoming GO phase will be short or long.

The success of the first manipulation depends on participants' ability to flexibly change their preparatory strategy between miniblocks. We were unsure, however, whether participants had this flexibility. For that reason, we included a second manipulation in which we varied (between groups) the proportion of long/short GO phases. We reasoned that if participants cannot change their preparatory strategy on a miniblock basis, they might be able to do so at the level of the entire experiment.

In detail, we predicted that high motivation to prepare (a short GO phase or a high proportion of short GO phases) would reflect in larger IBR (i.e., NEXT compatibility) effects and in smaller GO trial effects.

Method

Participants. Forty-four individuals (28 women, mean age = 23.98, *SD* = 2.38), similar in attributes to participants from the previous studies, took part in the experiment in return for course credit or for monetary compensation (40 NIS, ~$11 U. S.).

Materials and procedure. The procedure was based on Experiment 1. We introduced two manipulations related to the incentive to prepare for GO trials. For participants randomly assigned to a "mostly long" condition, only 25% of the miniblocks had two GO trials, and 75% had 10 GO trials. For participants in the "mostly short" condition, 75% of the miniblocks had two GO trials, and 25% of the miniblocks had 10 GO trials. In order to ensure that GO length primarily influenced the motivation to use the instructions rather than to encode them, the cue indicating the length of the GO phase was given only after the instructions for the miniblock were given. To ensure that the effect was not tied to any specific cue, three different sets of cues indicated the GO phase length, and cue type was randomly chosen in each miniblock. These cues were either "2" (vs. "10"), the Hebrew equivalents of "short" (vs. "long"), or "few" (vs. "many"). Cue type was included as an additional within-subjects variable in preliminary ANOVAs, but, given the fact that it was not involved in any significant effects, it was dropped from the report.

Participants were informed during the initial experimental instructions regarding the two possible lengths of the GO phase and were also informed about their specific group assignment. The participants were encouraged to try to improve their GO performance across the GO phase when this phase was long and were notified that this strategy was not possible for short GO phases. A 400-ms beep tone was played when an error occurred in the GO phase.

Results

NEXT phase. RTs were analyzed in a four-way ANOVA with the within-subjects independent variables compatibility (compatible-incompatible), NEXT trial (one through five), and GO length (short-long). The fourth (between-subjects) independent variable was proportion (mostly long-mostly short GO phases). Because the selection of trial conditions within the experiment was random, some of the conditions did not occur (for instance, a compatible trial in the fifth NEXT trial in a miniblock with a long GO phase, when short GO phases were more likely), thus creating ANOVA design cells with missing data. We therefore conducted the analysis twice, once on the results of participants without missing data (23 participants) and once on all the participants, this time dividing NEXT trial into two conditions instead of five conditions (one through two vs. three, four, five), thereby avoiding cells with missing data. Note that the partial sample may be viewed as a randomly selected subsample because the inclusion in this sample did not depend on performance. It only depended on the random selection of trials by the computer program. Given the fact that the two analyses yielded similar results, we report only the analysis conducted on the full sample.

There were no significant effects involving proportion in this analysis. There was a significant main effect of GO length, $F(1, 42) = 9.43$, $p = .004$, $MSE = 1{,}436.50$, $\eta_p^2 = .18$; NEXT trial, $F(1, 42) = 31.84$, $p < .001$, $MSE = 6{,}310.10$, $\eta_p^2 = .43$; and compatibility, $F(1, 42) = 23.43$, $p < .001$, $MSE = 3{,}567.77$, $\eta_p^2 = .36$. The interaction between these three variables reached significance, $F(1, 42) = 15.38$, $p < .001$, $MSE = 831.85$, $\eta_p^2 = .27$, indicating a reduction in the NEXT compatibility effect with NEXT phase progression, but only when GO length was long (see Figure 6). Importantly, this interaction had the same trend in the analysis of the partial sample that did not have missing data, $F(4, 84) = 2.51$, $p = .047$, $MSE = 2{,}139.81$, $\eta_p^2 = .11$ (see Figure 6). Finally, the three subordinate two-way interactions also reached significance—between GO length and NEXT trial, $F(1, 42) = 21.70$, $p < .001$, $MSE = 851.20$, $\eta_p^2 = .34$; between GO length and compatibility, $F(1, 42) = 4.10$, $p = .049$, $MSE = 987.98$, $\eta_p^2 = .09$; and between NEXT trial and compatibility, $F(1, 42) = 6.01$, $p = .018$, $MSE = 1{,}497.97$, $\eta_p^2 = .12$.

GO phase. The analyses focused on the comparison between short and long GO phases, and thus included only the first two GO trials from the long GO phase (because the short GO phase had only two trials). Both RTs and PEs were analyzed in four-way ANOVAs, with the within-subjects independent variables NEXT length (zero through five), GO trial (1–2), and GO length, and the between-subjects variable proportion. In RTs, no effect involving proportion was found to be significant. A significant main effect was found for NEXT length, $F(5, 190) = 24.44$, $p < .001$, $MSE = 3{,}504.30$, $\eta_p^2 = .39$; GO trial, $F(1, 38) = 185.80$, $p < .001$, $MSE = 16{,}529.48$, $\eta_p^2 = .83$; and GO length, $F(1, 38) = 4.76$, $p = .035$, $MSE = 6{,}204.63$, $\eta_p^2 = .11$. Significant two-way interactions were found between NEXT length and GO trial, $F(5, 190) = 16.93$, $p < .001$, $MSE = 3{,}485.63$, $\eta_p^2 = .31$, and between GO length and GO trial, $F(1, 38) = 10.92$, $p = .002$, $MSE = 2{,}672.05$, $\eta_p^2 = .22$ (see Figure 7). The first interaction indicates that the GO trial effect was largest when there were no NEXT trials. The second interaction is the critical one, which indicates, as predicted, that when there was a short GO phase, the GO trial effect was reduced.

At this stage, we remind the reader that the NEXT compatibility effect diminished in the course of the NEXT phase when participants knew that the upcoming GO phase would be long. Thus, the difference in NEXT performance became apparent in the last (third through fifth) NEXT trials. Does the reduced NEXT compatibility effect reflect incomplete preparation for the upcoming GO phase? If so, then following these NEXT phase trials, there should also be an increased GO trial effect. To address this question, we ran an ANOVA that just focused on GO trial performance following a NEXT phase that contained three to five trials. The independent variables in this ANOVA were NEXT length (three to five), Go trial, Go length, and proportion. As predicted, a significant two-way interaction between GO length and GO trial was observed, $F(1, 40) = 5.75$, $p = .021$, $MSE = 2{,}427.86$, $\eta_p^2 = .14$. This result indicates that, indeed, the smaller NEXT compatibility effect seen in the third to fifth NEXT trials (when the anticipated GO phase was long) was associated with larger GO trial effects (109 ms), when compared with the matched condition in which anticipated GO length was short (GO trial effect = 88 ms; see Figure 8).

The analysis of PE effects yielded comparable results. A main GO trial effect was observed, $F(1, 40) = 70.78$, $p < .001$, $MSE = 0.02$, $\eta_p^2 = .64$, along with a significant three-way interaction between GO length, NEXT length, and GO trial, $F(5, 200) = 5.70$, $p < .001$, $MSE = 0.01$, $\eta_p^2 = .12$. This interaction indicates that the

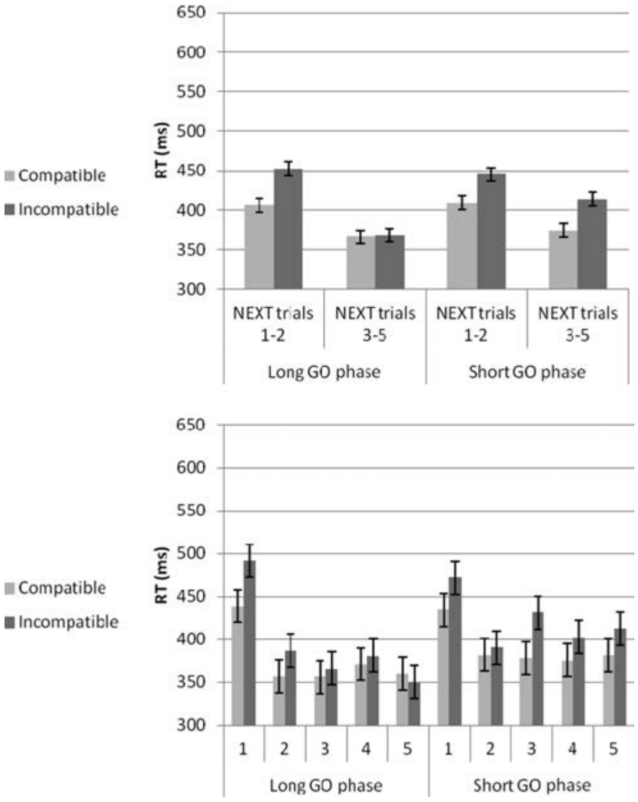


Figure 6. Experiment 4. NEXT RT (in milliseconds) as a function of NEXT trial, compatibility, and GO length for the entire sample (top panel) and the partial sample ($n = 23$, bottom panel). Error bars represent within-subject confidence intervals (Hollands & Jarmasz, 2010; Jarmasz, & Hollands, 2009). RT = response time.

GO trial–PE effect was more pronounced in miniblocks with a long GO phase, especially when no NEXT trials preceded the GO phase (see Figure 7). Finally, a significant two-way interaction was found between GO length and GO trial, $F(1, 40) = 5.74$, $p = .021$, $MSE = 0.013$, $\eta_p^2 = .12$, indicating that short GO phases had a smaller GO trial–PE effect. However, we analyzed GO trials following three to five NEXT trials (in which the GO length affected NEXT performance), both the interaction between GO length and GO trial, $F(1, 41) < 1.00$, $p = .453$, and the GO length main effect, $F(1, 41) < 1.00$, $p = .625$, were nonsignificant. Thus, the durable NEXT compatibility effect influenced the RT–GO trial effect, but did not influence its PE counterpart.

One potential concern regarding the GO trial effect is that, rather than serving as an index of task preparation, it primarily reflects effects of response repetition and congruency on the first GO trial. In particular, for the first GO trial, *response repetition* refers to whether this trial required the same response as the preceding NEXT trial(s). Note that, in this case, the response repetition effect is also a task-rule congruency effect (see Meiran & Kessler, 2008). This is due to the fact that the repeated response condition was also congruent in the sense that both the NEXT rule and the GO rule indicated this response as the correct response. In any event, this effect was significant neither in RT nor in PE, although the raw trend indicated quicker and more accurate repeated responses (500 vs. 484 ms, PE = .12, .09, $p = .12$ in both cases). The second repetition effect is in the second GO trial, and it refers to whether the response was the same as in the first GO trial. Although the RT effect approached significance (398 vs. 387 ms, $p = .069$), the PE effect was far from significant (PE = .044 vs. .042, $p = .770$).

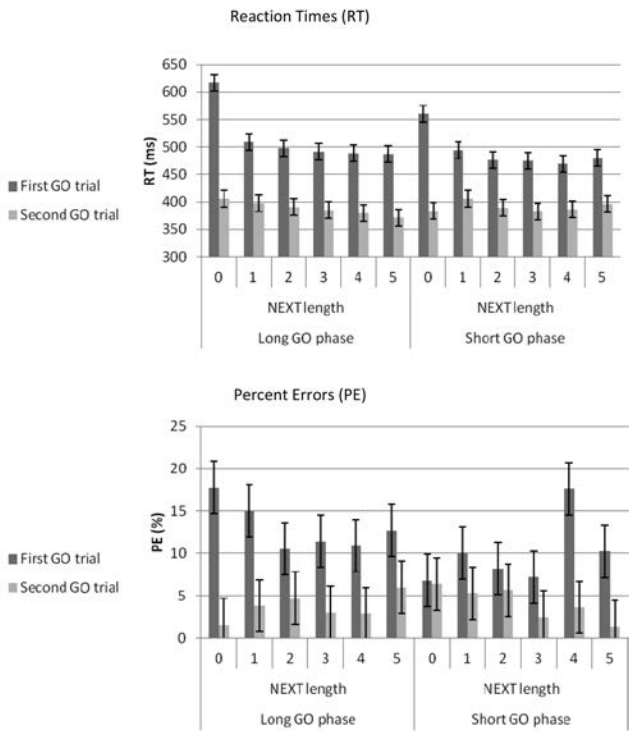


Figure 7. Experiment 4. GO performance as a function of GO length, Go trial, and the length of preceding the NEXT phase. Error bars represent within-subject confidence intervals (Hollands & Jarmasz, 2010; Jarmasz, & Hollands, 2009).

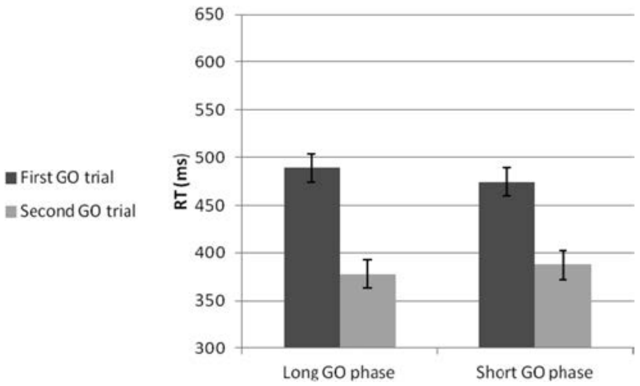


Figure 8. Experiment 4. GO trial RT (in milliseconds) after three to five NEXT trials as a function of GO trial and GO length. Error bars represent within-subject confidence intervals (Hollands & Jarmasz, 2010; Jarmasz, & Hollands, 2009). RT = response time.

Discussion

In this experiment, we manipulated the emphasis on proactive control by varying both GO length and the proportion of short GO lengths (across participants). We have also capitalized on the results of Experiment 3, which suggest that the NEXT compatibility effect seen in advanced NEXT trials is a trustworthy index of IBR. This is so despite the *potential* of being influenced by previous covert GO task execution. We predicted that stronger and more complete task setting (seen in larger or more durable NEXT compatibility effects) would be followed by small GO trial effects. The results generally support these predictions. When the GO phase was short, participants could not rely on learning the instructed task based on their success and failure, and thus had to rely heavily on proactive control. This was seen in relatively durable NEXT compatibility effects (replicating Experiments 1 and 2) coupled with small GO trial effects. In contrast, when the GO phase was long, the NEXT compatibility effect quickly diminished during the NEXT phase, a result that we interpret as evidence for failing to fully maintain task-related associations in WM. This was followed by amplified GO trial effects, indicating that the GO trial effect results, at least in part, from incomplete proactive configuration of the instructed S-R mapping.

The results thus show that participants have flexible control over their preparatory mode, and can change it quickly, on the level of a miniblock, which implies that they did not have to adjust it at the level of the entire experiment (seen in the general lack of proportion effects). We tentatively suggest that the NEXT compatibility effect results from the retrieval of instruction information from WM. Immediately following the instruction, this information was held in WM. When prewarned that the upcoming GO length would be short, participants maintained S-R translation information in WM throughout the NEXT phase, which is why the NEXT compatibility effect remained stable throughout the NEXT phase. However, when the GO length was long, lesser (or no) effort was invested in maintaining S-R translation information in WM, resulting in the quick shrinkage of the NEXT compatibility effect during the NEXT phase. As a result, when the GO phase began, initial GO performance was poor.[3]

General Discussion

In the present work, we were interested in demonstrating IBR under conditions that overcome limitations of previous works. Overcoming these limitations was important in order to show that instructions alone can lead to highly efficient performance, so efficient that it becomes reflexive. We set up five prerequisites that we considered as essential to rule out all the alternative accounts that we could think of and showed than none of the published works has met all of them, including works coming from our own group. Across four experiments with our newly designed NEXT paradigm, we showed evidence for IBR that met all the prerequisites. We additionally report several new results that have not been previously reported, mainly because our paradigm allowed us to examine trial-by-trial changes in IBR and task performance.

The NEXT compatibility effect was observed in all experiments. In Experiment 2, we addressed the possibility that participants covertly execute the GO task in the compatible trials of the NEXT phase, rendering results from these trials equivocal. In that experiment, we showed that the NEXT compatibility effect was predominantly due to incompatibility-related interference rather than to compatibility-related facilitation. In Experiment 3, we observed that participants took about 5 s to encode the instructions, suggesting the engagement of a covert preparatory process; however, additional (limited) overt practice did not produce any further significant benefits to performance. Lastly, in Experiment 4, we addressed the correspondence between preparation (indexed by the NEXT compatibility effect) and task performance. As an index of task performance, we used the GO trial effect (the difference between the first and second GO trials). This measure was used because the first GO trial involves performance that is purely instruction based, whereas performance in the remaining GO trials was expected to be more efficient due to prior experiences of actual task execution. Our findings indicate that, indeed, states associated with stronger preparation were characterized by better sustained NEXT compatibility effects and also by smaller GO trial effects, both indicating better preparation.

The results of Experiment 3 provide good news regarding existing paradigms to study IBR, especially De Hower et al.'s (2005) and Liefooghe et al.'s (2012, 2013) in showing that a few incidences of overt task execution probably do not modify the presence of reflexivity. This conclusion is in line with the literature on skill-based automaticity (Anderson, 1982; Logan, 1988; Rosenbloom & Newell, 1986) indicating that automaticity requires many task executions. In that sense, it is not surprising that the few task executions that were permitted did not make much of a difference.

Perhaps the most intriguing aspect about IBR is what characterized the preparatory processes that produce it. The results of Experiment 3, showing that preparation took approximately 5 s, suggest that merely encoding the instructions is insufficient to generate IBR. The few seconds it took to prepare also suggest that a mental simulation of actual task execution (assuming that such simulation is as potent as actual task execution) is also unlikely to have produced automaticity as it is discussed in the skill acquisition literature. Specifically, even if participants have simulated, say, two choices per second (an amazing rate!), they could have simulated up to 10 choices within the 5-s period. Our results suggest that this amount of practice is unlikely to have been sufficient to build a skill as described in the skill acquisition literature.

Below we detail a speculative account of what could have taken place during this period. Our speculation is mainly based on two (related) lines of theorizing. One that has already been described is the distinction between proactive and reactive control (Braver, 2012). The other is the notion that WM can be in either one of two modes: an updatable mode and a maintenance mode (e.g., Frank, Loughry, & O'Reilly, 2001; O'Reilly & Frank, 2006), and the notion that switching between these modes takes considerable time (~500–750 ms according to Kessler & Oberauer, 2014).

WM provides the ability to maintain information in an intact and undistracted manner. To explain how such maintenance is implemented, and how it is coordinated with the requirement to update the contents of WM from time to time, several theorists suggested a gating system that controls the input from perceptual representations to WM (e.g., Braver & Cohen, 2000; Frank et al., 2001; O'Reilly, Braver, & Cohen, 1999) or control the output from WM to the action systems (Chatham, Frank, & Badre, 2014). Building on this notion, we suggest that maintenance and updating are not only carried out by different functional states of WM but also can be regarded as two modes of information processing. Specifically, an updating mode is characterized by relatively nonselective intake of information, whereas maintenance mode is characterized by a relatively rigid adherence to the stored representations (if input is gated) or the action mode (if output from WM is gated).

We suggest that situations differ in the need for updating and that participants balance between the maintenance and updating modes of operation according to the situation. Practice-based learning does not rely on WM, and hence can be carried out while WM is engaged in maintaining other task-irrelevant information. By contrast, goal-directed (i.e., newly instructed) behavior relies on maintaining the relevant task in WM (Miller & Cohen, 2001) and preventing new input from interfering with WM contents. Alternatively, it may rely on preventing newly updated WM representations from interfering in the course of already launched motor plans through the gating of WM output.

Although goal-directed behavior is generally assumed to involve active WM representation, they also rely—although to a much lesser extent—on LTM-based associations that were learned through practice (e.g., Cole et al., 2013; Waszak, Hommel, & Allport, 2003). By contrast, newly instructed task performance and the related IBR phenomenon exclusively depends on the WM-based representation of the goal. For this reason, the need to maintain the task representation following the instructions is even stronger than in other goal-directed situations that

---

[3] We note that, although the GO length effects were based on prewarning, we cannot tell to what degree the behavioral change was driven by conscious strategic processing. Actually, it is quite likely that participants have quickly learned to adjust their performance without any clear conscious sense of what exactly they were doing. In any event, the results show that participants have at least some control over their preparatory efforts.

are highly practiced, such as in task-switching paradigms. Furthermore, the demand to execute the task efficiently starting from the first trial, and the lack of opportunity to practice the task result in an especially high-readiness state, manifested by making the goal representation highly activated and highly shielded from interference. However, as much as WM updating supports cognitive flexibility, needing to shield newly formed plans results in excessive rigidity reflected in IBR (Meiran et al., 2012).

To appreciate this point, consider an ideal performer who would be able to execute the NEXT phase while being completely uninfluenced by the preparedness to execute the following GO phase. Such an ideal performer can flexibly shift between the NEXT phase and the GO phase, or, more generally, is able to immediately update WM with the task information, without paying the price of distractibility by degrading the maintained S-R associations.

When an opportunity to update the plan by trial and error is given (as in the long GO phase in Experiment 4), participants adopt a processing mode in which updating the plan is allowed. Consequently, two (related) things happen to them. On the one hand, they quickly adjust to the requirements of the NEXT phase, which are to ignore the shapes of the stimuli. As a result, the interference from the shapes is quickly abolished, as seen in the quickly diminishing NEXT compatibility effect in that condition. On the other hand, allowing new stimuli to influence WM contents results in a less optimally prepared GO phase that was seen in relatively poor first-GO trial performance when there were 10 GO trials. As suggested by Meiran and colleagues (2012), the balance between updating the goal and maintaining the instructions leads to paradoxical effects. A proactive processing strategy relies on robust maintenance and hence leads to rigidity, whereas reactive processing, typically regarded as involving less control, relies more heavily on the environmental input and hence leads to flexibility.

Recent modeling works (especially Huang et al., 2013; Ramamoorthy & Verguts, 2012) describe learning through instruction by two routes (these routes differ somewhat in the two models). One route involves the prefrontal cortex and is described as slow, but flexible. The other route is quick, but its learning requires extensive repetitions. This quick route gives rise to automaticity, as usually described, and thus needs considerable practice. Had it been that IBR was caused by the automatic route, its existence would be challenging to these theories that assume that automaticity does not develop within a matter of seconds. However, we suggest that IBR reflects the operation of the slow and flexible prefrontal route and is thus qualitatively different from automaticity. Unlike automaticity that may be indifferent to the current goal (e.g., Moors & de Houwer, 2006), IBR reflects the shielding of the instructions against distraction (see Dreisbach, 2012, for a related point) in order to ensure their maintenance. Such shielding is required given the fragile nature of the associated prefrontal representations (Cole et al. 2013).

References

Allport, D. A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In M. Moscovitch & C. Umiltà (Eds.), *Attention and performance XVIII: Conscious and nonconscious information processing* (pp. 421–452). Cambridge, MA: MIT Press.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89,* 369–406. doi:10.1037/0033-295X.89.4.369

Bargh, J. A. (1992). The ecology of automaticity: Toward establishing the conditions needed to produce automatic processing effects. *American Journal of Psychology, 105,* 181–199. doi:10.2307/1423027

Baumeister, R. F. (2005). *The cultural animal: Human nature, meaning, and social life*. New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195167030.001.0001

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences, 16,* 106–113. doi:10.1016/j.tics.2011.12.010

Braver, T. S., & Cohen, J. D. (2000). On the control of control: The role of dopamine in regulating prefrontal function and working memory. In S. Monsell & J. Driver (Eds.), *Attention and performance XVIII* (pp. 713–737). Cambridge, MA: MIT Press.

Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 76–106). New York, NY: Oxford University Press.

Braver, T. S., Paxton, J. L., Locke, H. S., & Barch, D. M. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences, 106,* 7351–7356. doi:10.1073/pnas.0808187106

Bugg, J. M., & Scullin, M. K. (2013). Controlling intentions: The surprising ease of stopping after going relative to stopping after never having gone. *Psychological Science, 24,* 2463–2471. doi:10.1177/0956797613494850

Bugmann, G. (2012). Modeling fast stimulus-response association learning along the occipito-parieto-frontal pathway following rule instructions. *Brain Research, 1434,* 73–89. doi:10.1016/j.brainres.2011.09.028

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind, 11,* 63–65. doi:10.1093/mind/os-XI.41.63

Chatham, C. H., Frank, M. J., & Badre, D. (2014). Corticostriatal output gating during selection from working memory. *Neuron, 81,* 930–942. doi:10.1016/j.neuron.2014.01.002

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,* 1140–1154.

Cohen-Kdoshay, O., & Meiran, N. (2009). The representation of instructions operates like a prepared reflex: Flanker compatibility effects found in first trial following S-R instructions. *Experimental Psychology, 56,* 128–133. doi:10.1027/1618-3169.56.2.128

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

De Houwer, J., Beckers, T., Vandorp, S., & Custers, R. (2005). Further evidence for the role of mode-independent short-term associations in spatial Simon effects. *Perception & Psychophysics, 67,* 659–666. doi:10.3758/BF03193522

Dreisbach, G. (2012). Mechanisms of cognitive control: The functional role of task rules. *Current Directions in Psychological Science, 21,* 227–231. doi:10.1177/0963721412449830

Eriksen, C. W. (1995). The flanker task and response competition: A useful tool for investigating a variety of cognitive problems. *Visual Cognition, 2,* 101–118. doi:10.1080/13506289508401726

Exner, S. (1879). Physiologie der Grosshirnrinde. In L. Hermann (Ed.), *Handbuch Der Physiologie* [Handbook of psychology] (Vol. 2, pp. 189–350). Leipzig, Germany: Vogel.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of*

Experimental Psychology: Human Perception and Performance, 18, 1030–1044. doi:10.1037/0096-1523.18.4.1030

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, & Behavioral Neuroscience, 1,* 137–160. doi:10.3758/CABN.1.2.137

Gibson, J. J. (1941). A critical review of the concept of set in contemporary experimental psychology. *Psychological Bulletin, 38,* 781–817. doi:10.1037/h0055307

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist, 54,* 493–503. doi:10.1037/0003-066X.54.7.493

Herrmann, E., Call, J., Hernandez-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science, 317,* 1360–1366. doi:10.1126/science.1146282

Hollands, J. G., & Jarmasz, J. (2010). Revisiting confidence intervals for repeated measures designs. *Psychonomic Bulletin & Review, 17,* 135–138. doi:10.3758/PBR.17.1.135

Hommel, B. (1998). Automatic stimulus–response translation in dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1368–1384. doi:10.1037/0096-1523.24.5.1368

Hommel, B. (2000). The prepared reflex: Automaticity and control in stimulus response translation. In S. Monsell & J. Driver (Eds.), *Attention and performance, 18: Control of cognitive processes* (pp. 247–273). Cambridge, MA: MIT Press.

Huang, T. R., Hazy, T. E., Herd, S. A., & O'Reilly, R. C. (2013). Assembling old tricks for new tasks: A neural model of instructional learning and control. *Journal of Cognitive Neuroscience, 25,* 843–851. doi:10.1162/jocn_a_00365

Jarmasz, J., & Hollands, J. G. (2009). Confidence intervals in repeated-measures designs: The number of observations principle. *Canadian Journal of Experimental Psychology, 63,* 124–138. doi:10.1037/a0014164

Jonides, J., & Mack, R. (1984). On the cost and benefit of cost and benefit. *Psychological Bulletin, 96,* 29–44. doi:10.1037/0033-2909.96.1.29

Kessler, Y., & Oberauer, K. (2014). Working memory updating latency reflects the cost of switching between maintenance and updating modes of operation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 738–754. doi:10.1037/a0035545

Kiesel, A., Wendt, M., & Peters, A. (2007). Task switching: On the origins of response congruency effects. *Psychological Research, 71,* 117–125. doi:10.1007/s00426-005-0004-8

Kunde, W., Elsner, K., & Kiesel, A. (2007). No anticipation–no action: The role of anticipation in action and perception. *Cognitive Processes, 8,* 71–78. doi:10.1007/s10339-007-0162-2

Kunde, W., Kiesel, A., & Hoffmann, J. (2003). Conscious control over the content of unconscious cognition. *Cognition, 88,* 223–242. doi:10.1016/S0010-0277(03)00023-4

Langer, E. (2000). Mindful learning. *Current Directions in Psychological Science, 9,* 220–223. doi:10.1111/1467-8721.00099

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

Logan, G. D. (1978). Attention in character classification tasks: Evidence for the automaticity of component stages. *Journal of Experimental Psychology: General, 107,* 32–63. doi:10.1037/0096-3445.107.1.32

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527. doi:10.1037/0033-295X.95.4.492

Logan, G. D. (1992). Attention and preattention in theories of automaticity. *American Journal of Psychology, 105,* 317–339. doi:10.2307/1423031

Logan, G. D., & Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 179–195. doi:10.1037/0278-7393.17.2.179

Luchins, A. S. (1942). Mechanization in problem solving: The effect of *Einstellung. Psychological Monographs, 54,* i–95.

Luria, R., Meiran, N., & Dekel-Cohen, C. (2006). Stimulus cued completion of reconfiguration and retroactive adjustment as causes for the residual switching cost in multi-step tasks. *European Journal of Cognitive Psychology, 18,* 652–668. doi:10.1080/09541440500423293

Meiran, N., & Cohen-Kdoshay, O. (2012). Working memory load but not multitasking eliminates the prepared reflex: Further evidence from the adapted flanker paradigm. *Acta Psychologica, 139,* 309–313. doi:10.1016/j.actpsy.2011.12.008

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. doi:10.3389/fnhum.2012.00104

Meiran, N., & Kessler, Y. (2008). The task rule congruency effect in task switching reflects activated long-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 137–157. doi:10.1037/0096-1523.34.1.137

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1975). Loci of contextual effects in visual word recognition. In P. M. A. Rabbitt (Ed.), *Attention and performance V* (pp. 98–118). New York, NY: Academic Press.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24,* 167–202. doi:10.1146/annurev.neuro.24.1.167

Moors, A., & De Houwer, J. (2006). Automaticity: A conceptual and theoretical analysis. *Psychological Bulletin, 132,* 297–326. doi:10.1037/0033-2909.132.2.297

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.

O'Reilly, R. C., Braver, T. S., & Cohen, J. D. (1999). A biologically-based computational model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). Cambridge, United Kingdom: Cambridge University Press.

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation, 18,* 283–328. doi:10.1162/089976606775093909

Pashler, H., & Baylis, G. (1991). Procedural learning: I. Locus of practice effects in speeded choice tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 20–32. doi:10.1037/0278-7393.17.1.20

Psychology Software Tools. (2010). *E-Prime 2.0*. Sharpsburg, PA: Author.

Ramamoorthy, A., & Verguts, T. (2012). Word and deed: A computational model of instruction following. *Brain Research, 1439,* 54–65. doi:10.1016/j.brainres.2011.12.025

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124,* 207–231. doi:10.1037/0096-3445.124.2.207

Rosenbloom, P. S., & Newell, A. (1986). The chunking of goal hierarchies: A generalized model of practice. In R. S. Michaliski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (Vol. 2, pp. 247–288). Los Altos, CA: Morgan Kaufmann.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84,* 127–190. doi:10.1037/0033-295X.84.2.127

Tzelgov, J. (1997). Automatic but conscious: That is how we act most of the time. In R. Wyer (Ed.), *Advances in social cognition* (Vol. X, pp. 217–230). Mahwah, NJ: Lawrence Erlbaum Associates.

van't Wout, F., Lavric, A., & Monsell, S. (2013). Are stimulus–response rules represented phonologically for task-set preparation and maintenance? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1538–1551. doi:10.1037/a0031672

Walser, M., Fischer, R., & Goschke, T. (2012). The failure of deactivating intentions: Aftereffects of completed intentions in the repeated prospective memory cue paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1030–1044. doi:10.1037/a0027000

Waszak, F., Hommel, B., & Allport, A. (2003). Task-switching and long-term priming: Role of episodic stimulus-task bindings in task-shift costs. *Cognitive Psychology, 46,* 361–413. doi:10.1016/S0010-0285(02)00520-0

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

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

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41,* 67–85. doi:10.1016/0001-6918(77)90012-9

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

Zbrodoff, N. J. (1999). Effects of counting in alphabet arithmetic: Opportunistic stopping and priming of intermediate steps. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 299–317. doi:10.1037/0278-7393.25.2.299

Appendix

Analysis of GO Performance

Experiments 1 and 2

Both response time (RT) and proportion of errors (PE) were analyzed in two-way analyses of variance (ANOVAs), with the within-subjects independent variables NEXT length (zero through six), which indicates the number of NEXT trials preceding the GO trials, and GO trial (first vs. second) (see Figure A1). In RTs, both the NEXT length main effect, $F(5, 95) = 8.22$, $p < .001$, $MSE =$

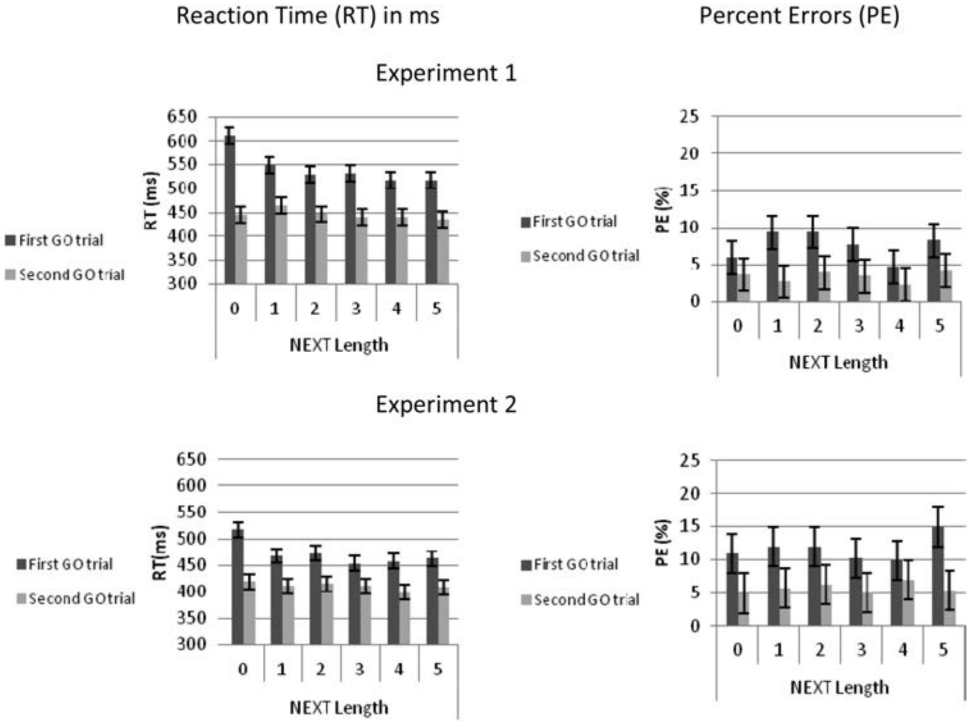
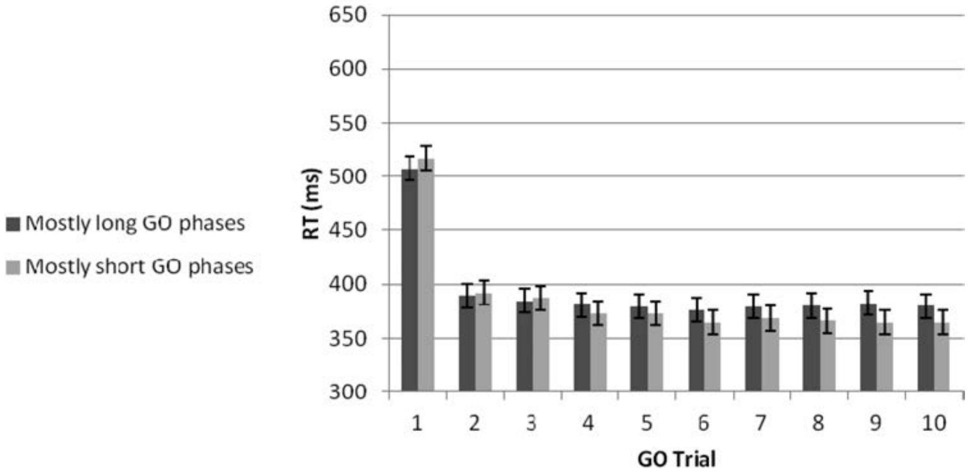


Figure A1. Experiments 1 and 2. GO performance as a function of GO trial and the length of the NEXT phase, which preceded it in RTs (in milliseconds, left panels) and percent errors (right panels). Error bars represent within-subject confidence intervals (Hollands & Jarmasz, 2010; Jarmasz, & Hollands, 2009).

(*Appendix continues*)

1,910.68, $\eta_p^2 = .30$ (Experiment 1); $F(5, 205) = 8.56$, $p < .001$, $MSE = 1{,}997.95$, $\eta_p^2 = .17$ (Experiment 2), and the GO trial main effect, $F(1, 19) = 56.06$, $p < .001$, $MSE = 10{,}091.48$, $\eta_p^2 = .75$ (Experiment 1); $F(1, 41) = 60.41$, $p < .001$, $MSE = 7{,}848.79$, $\eta_p^2 = .59$ (Experiment 2), were significant. The two-way interaction was also significant, $F(5, 95) = 8.26$, $p < .001$, $MSE = 1{,}427.26$, $\eta_p^2 = .30$ (Experiment 1); $F(5, 205) = 3.59$, $p = .004$, $MSE = 2{,}106.64$, $\eta_p^2 = .08$ (Experiment 2), indicating that the GO trial effect depended on the length of the preceding NEXT phase. However, this interaction was caused by the discrepant condition of zero NEXT trials. When considering just NEXT Lengths 1–5, the interaction vanished, $F(4, 76) < 1.00$, $p = .940$ (Experiment 1); $F(4, 164) < 1.00$, $p = .748$ (Experiment 2). In Experiment 2, in which the NEXT compatibility effect diminished with progress, the GO trial effect also diminished with progress, but the relevant two-way interaction between progress and GO trial was only marginally significant, $F(2, 82) = 2.55$, $p = .084$, $MSE = 769.71$, $\eta_p^2 = .06$. However, the (linear) component of that interaction was significant, $F(1, 41) = 4.53$, $p = .039$, $MSE = 678.77$, $\eta_p^2 = .09$, reflecting a reduction in the GO trial effect with progress (the

Reaction Time (RT) in ms



Percent Errors (PE)

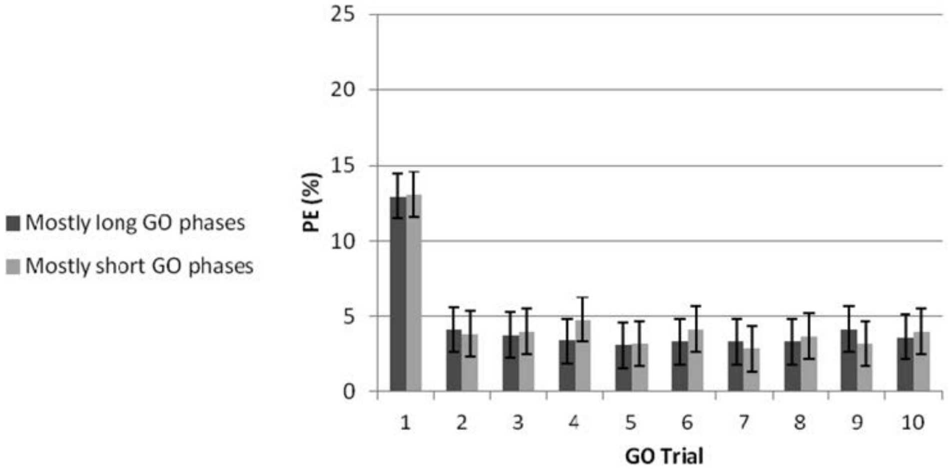


Figure A2. Experiment 4. GO performance in long GO phases as a function of proportion and GO trial. Error bars represent within-subject confidence intervals (Hollands & Jarmasz, 2010; Jarmasz, & Hollands, 2009).

(*Appendix continues*)

effect was 70, 54, and 51 ms in the first through third part of the session, respectively).

The PE analysis of Experiments 1–2 indicated only a significant GO trial main effect, $F(1, 19) = 19.32$, $p < .001$, $MSE = 0.01$, $\eta_p^2 = .50$ (Experiment 1); $F(1, 41) = 42.54$, $p < .001$, $MSE = 0.01$, $\eta_p^2 = .51$ (Experiment 2).

Another finding is the GO trial effect, which was largest when there were no NEXT trials. Two factors may be responsible for this finding. One is unexpectedness, because there were only 10% such trials. Another potential factor is related to the task from which participants switched when beginning the GO phase. When there were NEXT trials, the switch was from the NEXT phase. However, when there were no NEXT trials, the switch was from the instruction phase. Moreover, we know that switch costs may be enlarged when one switches from a difficult task (e.g., Allport , Styles, & Hsieh, 1994), and encoding the instruction was probably more demanding than executing a NEXT response.

In the next analyses, we examined response repetition effects, partly in order to show that there is a genuine GO trial effect. We analyzed these repetition effects in each GO response (first, second) separately. This was done because the response-repetition effects in thefirst GO responses were more likely to be influenced by task-rule congruency effects (Meiran & Kessler, 2008). Specifically, when the response repeats from the NEXT phase to the GO phase, the first GO response is identical with the NEXT response, meaning that it could reflect the erroneous (and undetectable) execution of a NEXT response. This is less likely to happen in the second GO response, because participants had the time to realize that the stimuli are now appearing in *green*.

Second GO trial. The RT results indicate a significant response-repetition effect in the second GO trial in Experiments 1 and 2 (463 vs. 436 ms), $F(1, 19) = 4.83$, $p = .041$, $MSE = 1{,}482.76$, $\eta_p^2 = .20$ (Experiment 1); (427 vs. 398 ms), $F(1, 41) = 11.25$, $p = .002$, $MSE = 1{,}592.62$, $\eta_p^2 = .21$ (Experiment 2). The PE results indicate a significant effect in Experiments 1 (.04 vs. .02), $F(1, 19) = 5.57$, $p = .029$, $MSE = 0.0006$, $\eta_p^2 = .22$, but not in Experiment 2 (.07 vs. .05, $p = .064$).

Given the facilitatory effects of response repetition, we wanted to see whether the GO trial effect remains significant even when we compare the first GO response with the second GO response that does not involve response repletion. This contrast proved significant in RT in all three experiments (542 vs. 463 ms), $F(1, 19) = 38.77$, $p < .001$, $MSE = 1{,}596.42$, $\eta_p^2 = .67$ (Experiment 1); (472 vs. 427 ms), $F(1, 41) = 38.51$, $p < .001$, $MSE = 1{,}059.54$, $\eta_p^2 = .48$ (Experiment 2), and in PE in both experiments (.08 vs. .04), $F(1, 19) = 26.11$, $p < .001$, $MSE = 0.0005$, $\eta_p^2 = .58$ (Experiment 1); (.11 vs. .07), $F(1, 41) = 23.61$, $p < .001$, $MSE = 0.0020$, $\eta_p^2 = .36$ (Experiment 2).

First GO trial. These analyses indicate opposite nonsignificant trends in RT (548 vs. 526 ms, for switched and repeated responses, respectively, $p = .28$, Experiment 1; 466 vs. 473 ms, $p = .57$, Experiment 2). In Experiment 1, there was a just significant decrease in PE due to response repetition (.11 vs. .06, for switched and repeated responses, respectively), $F(1, 19) = 4.40$, $p = .049$, $MSE = 0.006$, $\eta_p^2 = .19$, but not in Experiment 2 (.12 in both switched and repeated responses, $p = .930$).

Experiment 4

Additional analyses to those reported in the main body of the article were conducted on the long GO phase (see Figure A2). The main motivation to run these analyses was to establish the second GO trial as a suitable baseline (reflecting performance after nearly complete setting). The long GO phases made it possible to test whether this was indeed the case. Nearly complete setting would therefore reflect in minor differences between the second and additional (i.e., third – 10th) GO trials. The pattern of results (see Figure A2) mostly confirmed these predictions in showing that the largest improvement in GO performance was seen after the first GO trial. We tested the significance of these trends in the following analyses, which also included the between-subjects variable proportion (mostly long vs. mostly short GO phases).

RT. The main effect of GO trial was significant, $F(9, 378) = 146.14$, $p < .001$, $MSE = 562.39$, $\eta_p^2 = .78$, indicating a significant decrease in RT as a function of the progress in the GO phase, especially between the first and the remaining GO trials. The interaction between GO trial and proportion was marginally significant, $F(9, 378) = 1.90$, $p = .051$, $MSE = 562.39$, $\eta_p^2 = .04$. When the first GO trial was omitted from the analysis, the GO trial main effect remained significant, $F(8, 336) = 11.55$, $p < .001$, $MSE = 179.76$, $\eta_p^2 = .21$. Moreover, there was a significant difference in the linear trend between the proportion groups, $F(1, 42) = 5.38$, $p = .025$, $MSE = 1{,}607.65$, $\eta_p^2 = .11$, indicating a response speedup during the GO phase, which was more pronounced in the group who performed mostly short GO phases.

PE. The main effect of GO trial was significant, $F(9, 378) = 28.80$, $p < .001$, $MSE = 0.04$, $\eta_p^2 = .41$ (see Figure A2). When the first GO trial was omitted from the analysis, the main effect of GO trial was no longer significant, $F(8, 336) < 1$, $p = .749$.

Received March 17, 2014
Revision received July 16, 2014
Accepted July 22, 2014 ■