



Automatic effects of instructions: a tale of two paradigms

Inbar Amir¹ · Liran Peleg¹ · Nachshon Meiran¹

Received: 8 June 2021 / Accepted: 9 September 2021 / Published online: 28 September 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

When examining rapid instructed task learning behaviorally, one out of two paradigms is usually used, the Inducer-Diagnostic (I-D) and the NEXT paradigm. Even though both paradigms are supposed to examine the same phenomenon of Automatic Effect of Instructions (AEI), there are some meaningful differences between them, notably in the size of the AEI. In the current work, we examined, in two pre-registered studies, the potential reasons for these differences in AEI size. Study 1 examined the influence of the data-analytic approach by comparing two existing relatively large data-sets, one from each paradigm (Braem et al., in *Mem Cogn* 47:1582–1591, 2019; Meiran et al., in *Neuropsychologia* 90:180–189, 2016). Study 2 focused on the influence of instruction type (concrete, as in NEXT, and abstract, as in I-D) and choice complexity of the task in which AEI-interference is assessed. We did that while using variants of the NEXT paradigm, some with modifications that approximated it to the I-D paradigm. Results from Study 1 indicate that the data-analytic approach partially explains the differences between the paradigms in terms of AEI size. Still, the paradigms remained different with respect to individual differences and with respect to AEI size in the first step following the instructions. Results from Study 2 indicate that Instruction type and the choice complexity in the phase in which AEI is assessed do not influence AEI size, or at least not in the expected direction. Theoretical and study-design implications are discussed.

Introduction

Imagine that you are cooking a new dish. You are familiar with all the needed ingredients, you are even experienced with the relevant cooking techniques, yet you have never cooked this particular dish beforehand. Still, a quick glance at the recipe might be sufficient for you to execute the procedure accurately. Though at first sight, the ability to translate instructions into action may seem trivial, a second thought reveals that in truth, it is remarkable: We are able to immediately and efficiently translate semantic-verbal-declarative input such as "pour half a cup of water into the sifted flour" into an accurate and efficient procedural representation and action. This remarkable ability, referred to as Rapid Instructed Task Learning (RITL; Cole et al., 2013), is evident in a wide array of everyday tasks, including instruction-based road navigation, building your new Ikea furniture, and more.

Two main behavioral markers indicate that implementing new instructions is highly efficient and autonomous: (1) High accuracy rate from first execution (around 90%, for few examples, see Cole, 2009; Pereg & Meiran, 2019; Theeuwes et al., 2015). This marker indicates that the instructions had been learned (to a substantial degree) without any prior practice. (2) Automatic Effects of Instructions (AEI). Meiran et al. (2017) suggested that AEI, which is measured by a congruency effect (see below), represents the unintentional activation of the quickly proceduralized new instructions leading to interference in the execution of a different, familiar task, even though the newly learned instructions have never been executed yet. Given that autonomous processing is taken as a core marker of skill-based automaticity (Moors, 2016), the fact that it takes place immediately upon instructions may be regarded as no less than amazing!

When examining AEI behaviorally, one of the two paradigms is usually used, (1) Inducer-Diagnostic paradigm [I-D, first presented in Liefoghe et al. (2012); see Fig. 1A], and (2) NEXT paradigm [first presented by Meiran et al. (2015); see Fig. 1B]. Both paradigms involve simple, newly instructed task rules in each experimental mini-block, thus creating many first encounters with stimulus–response (S–R) rule sets. Both paradigms are designed similarly. In

✉ Inbar Amir
inbarami@post.bgu.ac.il

¹ Department of Psychology and Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, Beer-Sheva, Israel

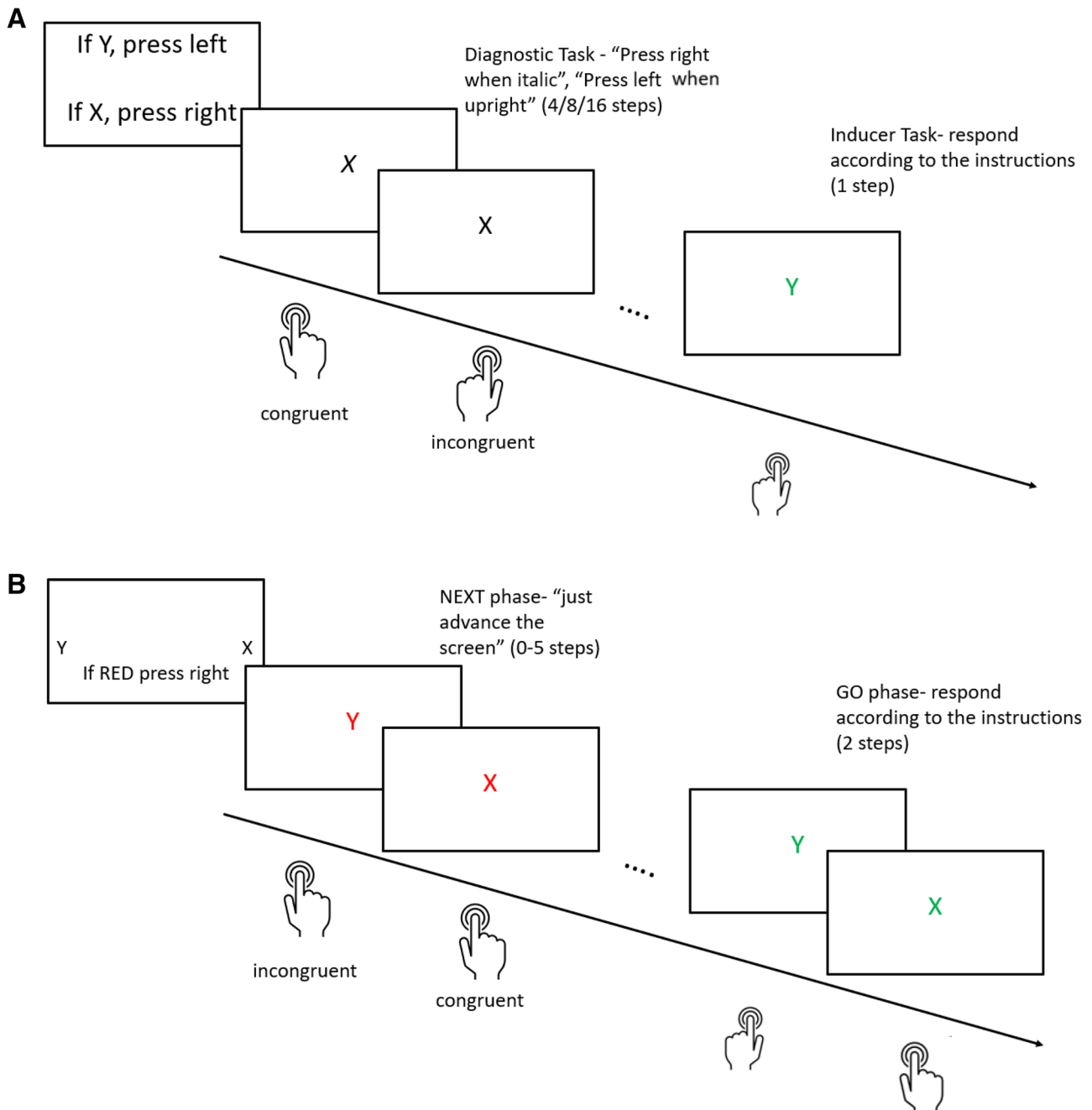


Fig. 1 **A** Inducer-Diagnostic paradigm (Liefoghe et al., 2012). **B** NEXT paradigm (Meiran et al., 2015)

accordance, they comprise multiple mini-blocks. Each mini-block contains 3 phases:

1. Instruction phase, where the new instructions are presented;
2. Interference phase (termed "diagnostic" and "NEXT", respectively), in which participants execute a constant task, different from the newly instructed task that had been introduced in the instruction phase, but one involv-

ing the same stimuli and responses. This design creates congruent steps (the correct reaction in the Interference task is equivalent to the reaction learned in the Instruction phase) and incongruent steps (the correct reaction in the task executed during the Interference task is opposite to that in the newly instructed task). The performance difference between congruent and incongruent steps provides a measure of AEI, i.e., the unintentional activation of the rules that were instructed in the instruction phase;

Table 1 Analytic and design aspects that differ between the I-D and the NEXT paradigms

Factor	Paradigm	
	Inducer-Diagnostic paradigm	NEXT paradigm
Log(RT) transformation	✓ (in early studies)	–
Inclusion of the first block when calculating AEI	–	✓
Inclusion of the first mini-block when calculating AEI	–	✓
Inclusion of the first step when calculating AEI	–	✓
Length of Interference phase	48\16, equally probable	0–5, with an approximate geometric distribution
Calculation of AEI	Congruency effect pooled across steps	Congruency as a function of step, with a focus on the first step
Instruction's presentation method	Verbal-declarative ("if N, press left")	Spatial-procedural (location of stimuli indicates the reaction for it)
The complexity of the Interference phase (number of possible responses)	Two choice task. Response depends on the orientation of the stimulus (italic vs. upright)	Fixed key

3. Execution phase (termed "inducer" and "GO", in I-D and NEXT, respectively), where participants need to execute, for the first time, the newly instructed task, the one that has been introduced in the Instruction phase.

Even though these two paradigms measure congruency effects as an expression of AEI, the respective AEI is markedly different in size. When examining AEI in the I-D paradigm, it is usually about 20 ms (Everaert et al., 2014; Liefoghe et al., 2012, 2013; Theeuwes et al., 2015). In contrast, in the NEXT paradigm, AEI ranges from about 40 to 100 ms (Meiran, et al., 2015, 2016; Pereg & Meiran, 2019). Do these differences tell us something about the conditions that promote the automaticity of instructions? If so, what are those conditions, and how are they expressed in each of the paradigms? To answer these questions, it is essential to systematically compare the two paradigms and find the factors that could cause the difference in AEI size between the two paradigms.

Differences between the paradigms

To address this issue, we first review the paradigm differences that we found (summarized in Table 1). By doing so, we could reveal some potentially relevant factors that may be responsible for the discrepancy in AEI size.

Instruction phase

The instructions in the Inducer-Diagnostic paradigm are presented in a verbal-declarative way ("if N, press left"), i.e., abstractly (see Fig. 1A). In contrast, the instructions in the NEXT paradigm are more concrete in nature: They are

indicated by the spatial location of each stimulus (i.e., one stimulus is presented on the left side of the screen and the second on the right side of the screen, see Fig. 1B). This difference hints at the possibility that the instructions' concrete/abstract nature promotes a concrete/abstract mental representation. Notably, there are theories suggesting that such differences in mental representation may influence resistance to interference and thus may also influence AEI size.

One such theory is the "task buffer" theory, developed by Cole et al. (2017). The theory suggests that the task buffer is a brain mechanism that represents the new instructions in a manner that is removed from immediate embodied representations (i.e., abstract), allowing for the maintenance of task representation without interference with ongoing task performance. According to this idea, the task buffer holds the instructions during the Interference phase, which allows remembering the instructions without executing them. Cole et al. linked the task buffer to the anterior Prefrontal Cortex (aPFC), a region that represents relatively abstract information (O'Reilly, 2010). Accordingly, it may be assumed that employing abstract instructions (as in the I-D paradigm) would facilitate the formation of relatively abstract brain representations (i.e., in the task buffer) as compared to when using relatively concrete instructions (as in the NEXT paradigm). In turn, the relatively abstract instructions would enable efficient blockage of distracting information during the Interference phase, thus leading to smaller AEI.

The notion that working memory (WM) involves multiple representation formats that differ in the degree of their abstractness agrees with Dreisbach's (2012) theory that distinguishes between concrete *S–R* rules and abstract task rules. According to her, even though the (concrete) *S–R* rules are executed faster (i.e., are more accessible), the performance which they guide is susceptible to interference from

irrelevant information, as compared to using abstract task rules. By Dreisbach's logic, the relatively concrete instructions in the NEXT paradigm may facilitate concrete representation, which in turn would cause greater interference in the NEXT phase and hence, a larger AEI. Note that both Cole et al.'s (2017) and Dreisbach's (2012) theories assume the existence of multiple formats for WM representation. This notion plays a key feature in other theories, such as Oberauer et al.'s (2013) theory assuming declarative and procedural WM. It also receives empirical support, such as from Formica et al. (2020), who showed evidence for declarative and procedural load's unique influence.

Contrary to the aforementioned account, some studies did not find an influence of manipulations that are expected to promote abstract representation on performance. For example, in an effort to determine the degree of abstractness of the representation, Liefoghe et al., (2012; Experiment 1) examined the influence of the degree of overlap in response between Interference and Execution on AEI size. Specifically, this could be the same response, a different response but one having the same right-left status, or a different response, being a key-press during Execution and a spoken response during Interference. Results show a non-significant influence of this manipulation on AEI. This finding may imply that instructions are represented abstractly by default, i.e., even when they are presented in a relatively concrete manner. Along a similar line, another study in which the NEXT paradigm was used indicates that the level of abstractness of the stimuli does not influence AEI (Longman et al., 2019; especially Experiment 1b). In this experiment, participants performed the NEXT task with multiple groups of stimuli that differed in their abstractness level (for example, a picture of a fish as opposed to the word "fish"). After task completion, participants had to rate the level of abstractness of each stimulus that appeared during the task. A non-significant influence of the level of abstractness of the stimuli on AEI size was found. Nevertheless, it is important to note that none of these studies compared task instructions that are very concrete, as in NEXT, to relatively abstract instructions, as in typical I-D studies. Such a comparison between the two instruction presentation methods could shed further light on the influence of the abstractness level on AEI.

Interference phase

One difference between the paradigms is the number of steps in the Interference phase. Most of the studies that used the NEXT paradigm involved 0–5 steps, with the number of steps determined from a quasi-exponential distribution to minimize changes in temporal expectancy with task progression. This arrangement results in most mini-blocks having very few (1–2) Interference steps. In

contrast, the I-D paradigm contains far more steps (4\8\16) with equal probability for these numbers. One may assume that the additional practice afforded in the I-D paradigm due to having many more Interference trials enables participants to overcome interference, resulting in a smaller AEI. Contrary to this hypothesis, Meiran et al. (2015) showed that although AEI size decreases from the first step, it does not change after it. Thus, although we examined the influence of this variable because it represents a difference between paradigms, there was no reason to expect that it would influence AEI size.

Another difference is the number of alternative responses in the Interference phase task (or task complexity). Specifically, the Interference phase in the I-D paradigm requires remembering and choosing the correct response from a set of two rules (if italic press right; if upright press left). Differently, in the NEXT paradigm, participants are not required to make any choice but are instead asked to "simply advance the screen using a constant NEXT response" (for example, see Pereg & Meiran, 2019). One may suggest that this dissimilarity causes different WM loads between the two paradigms, with a higher WM load in I-D than in NEXT.

It is still unclear whether the number of task rules held in WM during the Interference phase influences AEI size. Specifically, Pereg and Meiran (2019), who used the NEXT paradigm, showed that while novel rule implementation was impaired by increasing the number of alternatives in the newly instructed task (i.e., Execution phase), this manipulation did not influence AEI size. Notable, Pereg and Meiran (2019) studied the influence of the number of novel rules. Still, the differences between the two paradigms concern the number of familiar rules since the same Interference phase rules apply for the entire experiment. Given Meiran and Cohen-Kadosh's (2012) results showing a lack of an influence of the number of familiar rules on AEI, this makes it even less likely that this factor is responsible for the AEI size differences between paradigms.

A related line of investigation concerns individual differences. Specifically, individual differences in WM capacity (complex-span; Unsworth et al., 2005) were not significantly correlated with AEI size (Meiran et al., 2016). These results further argue against the possible influence of WM load on AEI size. We note that this last result has been challenged in a recent study by Braem et al. (2019) that employed the I-D paradigm and showed a negative correlation between AEI size and task execution efficiency. Obviously, one major difference between Meiran et al. (2016) and Braem et al. (2019) is the paradigm, making it difficult to judge whether the conclusion is general or paradigm specific.

Length of the execution phase

In most studies employing the NEXT paradigm, the Execution phase included two steps. In contrast, in most studies that employed the I-D paradigm, the Execution phase included one step. Can this difference between one and two steps influence AEI size? It initially may seem that because the Execution phase comes after the Interference phase, it is unlikely to influence what happens (e.g., AEI) during the Interference phase. However, one experiment suggests that such potential for an influence may exist. Specifically, one experiment studied (and found) an influence of the length of the Execution phase (two vs. ten steps) on AEI size (Meiran et al., 2015, Experiment 4). This experiment indicated a reduction in AEI size with NEXT phase progression, but only when GO length was long (i.e., ten steps). The authors attributed the results to the greater need to prepare in advance in the two-step condition. When the instructed task is executed many times (ten), participants can invest less effort in preparation and can rely instead on gradual learning that is afforded when the Execution phase is relatively long. Nonetheless, we find it unlikely that this reasoning would also apply to a difference between one vs. two steps (the difference between the I-D and NEXT) since two steps hardly enable gradual learning.

Analytic differences

Aside from the methodological differences between the two paradigms, there is another perhaps much more mundane explanation for the discrepancy in AEI size. According to it, this discrepancy between I-D and NEXT is (partly) due to the different analytic approaches that are typically employed. That such possibility exists was supported by a recent work that uncovered substantial variability in behavioral results across analysis teams who analyzed the same data-sets (e.g., Silberzahn et al., 2018). When examining reports concerning the two paradigms, we noticed the following differences:

Data exclusion

In reports of the NEXT paradigm, the analysis includes all the blocks and mini-blocks, except the very few training mini-blocks. In contrast, the typical analysis in reports from the I-D paradigm excludes the first block and sometimes the first Interference steps in each mini-block. These differences represent different research foci and considerations. Specifically, the NEXT team sees the first Interference step as the cleanest indication of AEI (Cole et al., 2017; Meiran et al., 2015; Pereg & Meiran, 2019). This notion assumes that Interference steps, especially congruent steps, could create LTM traces and hence represent instance-retrieval-based

automaticity (Logan, 1988) rather than representing an influence of instructions alone.

In contrast, the I-D team seems to consider the first step as being confounded by task-switching (e.g., Liefoghe et al., 2013). This seemingly small analytic difference may have important implications since the largest AEI is found in the first Interference step (for example, in Meiran et al., 2015). Moreover, the relative weight of the first step on the overall AEI is large when there are only a few Interference steps (as in NEXT). Accordingly, eliminating the first step from the AEI calculation in the I-D paradigm could be the major cause for the smaller AEI size compared to NEXT.

Log-transformation

Unlike in NEXT, where reaction times (RTs) are raw, in early I-D papers, the analysis employed log-transformed RTs as commonly done (see Ratcliff, 1993) partly in order to make the RT-distribution nearly symmetric and thus make the variance independent of the mean (Ratcliff & Murdock, 1976). However, more recent I-D papers do not use this method anymore (for example, see Braem et al., 2019 and Tibboel et al., 2016), and yet, AEI size in those papers remains the same ballpark as in the previous reports. Therefore, it is unlikely that log transformation is the reason for the AEI size differences between the two paradigms.

To conclude, although there are several analytic differences, the only difference that is likely to have a marked effect is the inclusion/exclusion of the first Interference step.

The present study

In two pre-registered studies, we examined the possible reasons for the aforementioned differences in AEI size between the two paradigms (see Table 1). In the first study, we examined the influence of the analytic approach on the different AEI size between the NEXT and the I-D paradigms. To do so, the current study used two existing data-sets from Meiran et al. (2016) and Braem et al. (2019). The data-sets were used to exploratorily examine the influence of most of the factors mentioned in the Introduction. By applying the same analytic method to both data-sets, it was possible to compare, for the first time, the size of the AEI in a manner that is not confounded by the analytic method as well as to assess the unique influence of each analytic factor on AEI size. Since this study was purely exploratory, we did not pre-register any hypotheses regarding each factor's specific influence on the AEI size. However, in retrospect, it seems that we should have hypothesized that the inclusion/exclusion of the first Interference step would have an influence. Although Study 1 may seem mundane, it is not so because reaching general conclusions, which are the basis for theorizing, make it necessary to compare conditions/paradigms

using uniform analytic approaches. This step thus helps to avoid theory development that relies on results that merely represent analytic choices rather than any real phenomena.

Study 2 focused on the influence of two experimental factors whose influence could not be isolated when comparing the existing data-sets: (1) instruction type and (2) choice complexity in the Interference task. We decided to focus on these two factors, given that both of them are (a) likely to show an affect and (b) may lead to theoretically interesting conclusions. Specifically, whether the instruction format encourages forming a specific mental representation format is an important issue for WM theorizing and research on instructions. The same also holds true for studying whether loading WM with procedural information (higher load in I-D than NEXT) influences automaticity. Study 2 thus employed the NEXT paradigm with adjustments that made it more similar to the I-D paradigm. Based on Cole et al.'s (2017) and Dreisbach's (2012) theories, we predicted that AEI would be smaller when the instruction format encourages abstract compared to concrete representation. However, it is important to note that some conflicting results (Liefoghe et al., 2012; Longman et al., 2019) cast doubt on the validity of our hypothesis. In addition, and in line with previous NEXT results (Pereg & Meiran, 2019) that indicate lack of influence of WM load on AEI size, we hypothesized that the choice complexity (and resultant WM load) in the Interference task would not influence AEI size. As for the previous hypothesis, some results from the I-D paradigm [i.e., the individual differences correlation reported by Braem et al., (2019)] suggest an opposite prediction.

Study 1

The aim of this study was to examine the influence of six data-analytic factors¹ that could potentially explain the discrepancy between the two paradigms in terms of AEI size.

¹ One technical difference was examined on an existing database (Meiran et al., 2016, $N=175$). Specifically, we wanted to reject the claim that the differences between the two paradigms are due to stimulus type (i.e., a significant portion of the stimuli presented in the NEXT paradigms are pictures, but there were no picture targets in most of the studies employing the I-D paradigm). RTs were analyzed in a two-way ANOVA with a within-subjects independent variables congruency (congruent -incongruent) and stimulus type (symbols, letters, digits, and pictures). Mauchly's test indicated that the assumption of sphericity had been violated for the interaction between stimulus type and congruency ($\chi^2(4)=0.38$, $p<0.001$). Therefore, the degrees of freedom were corrected using Greenhouse–Geisser estimates of sphericity ($\epsilon=0.67$). As predicted, the interaction was not significant, $F(2.68, 410.18)=0.20$, $p=0.880$, $\eta_p^2=0.0003$, and permitted acceptance of H_0 , $BF_{10}=0.002$. A focused contrast examining the AEI of pictures as compared to symbols, letters, and digits, pooled (which used the pooled error term) showed a non-significant difference as well with $t(612)=0.05$, $p=0.962$, with $BF_{10}=0.020$, indicating support for H_0 that stimulus type did not influence AEI.

These factors are: (1) log-transformed RT. (2) Inclusion\exclusion of the first block. (3) Inclusion\exclusion of the first mini-block in each experimental block. (4) Inclusion\exclusion of the first step of each mini-block. (5) The number of interfering steps included in the calculation of the AEI; and (6) examining AEI only in the first step.

We re-analyzed data from Meiran et al. (2016) and Braem et al. (2019). These data-sets were chosen because they employ the most typical paradigm format and are relatively large scale in terms of their N . (Pre-registration, R codes, and data-sets are available at <https://osf.io/ctvj2/>).

Methods

Participants

Participants were 175 Ben-Gurion University (Israel) students (Meiran et al., 2016) and 182 Ghent University (Belgium) students (Braem et al., 2019).

Data analysis

Data cleaning We used the R studio software (R Core Team, 2014) for pre-processing and analysis. Participants were removed because accuracy rates were below chance (50%) in either Interference or Execution, or below 2 SDs below the sample average, separately computed for each data-set. Nine participants were removed from the Meiran et al.'s (2016) data-set and fourteen from the Braem et al.'s (2019) data-set. After eliminations, the Meiran et al.'s sample included 166 participants, and the Braem et al.'s sample included 168 participants.

Additionally, Interference steps with an error were omitted from the RT analyses (0% from the Meiran et al. data-set and 7.47% from Braem et al. data-set). Steps with RT quicker than 200 ms were omitted from the analysis (0.20% of the remaining steps from the Meiran et al.'s data-set and 0% of the remaining steps from Braem et al.'s data-set) as well as steps in which RT exceeded 2SDs from the participants' mean RT as computed per participant, per task, per condition (4.17% of the remaining steps from Meiran et al.'s data-set and 4.97% of the remaining steps from Braem et al.'s data-set). In total, 4.37% of the trials were omitted from Meiran et al.'s data-set and 12.08% from Braem et al.'s data-set. All these pre-processing steps were executed as planned and are described in the pre-registration. After all eliminations, Meiran et al.'s data-set contained a mean of 60.8 observations per participant in the congruent condition (minimum of 43 observations) and a mean of 60.7 observations per participant in the incongruent condition (minimum of 46 observations). Braem et al.'s data-set contained a mean of 86.0 observations per participant in the congruent condition (minimum of 67 observations) and a

mean of 82.8 observations per participant in the incongruent condition (minimum of 31 observations).

Analytic procedure We conducted an exploratory comparison in which we examined the influence of the six analytic factors on AEI size. Given our emphasis, we analyzed only RT and only from the Interference phase while focusing on the descriptive statistics.

We created seven comparisons (the first one was the baseline). We started by calculating AEI in the same manner as typically employed when using the I-D paradigm (see Table 1). We then changed one analytic factor in each subsequent comparison until the analytic approach became similar to that in reports employing the NEXT paradigm. In the last comparison, we examined the AEI size separately for each interference step. By doing so, we could determine if and which of these six factors contribute to the discrepancy between paradigms in AEI size. In addition, we examined the correlation between AEI and performance in the Execution phase. This comparison was executed due to the conflicting results in the papers from which the data-sets were taken (see "Interference phase" subsection in the Introduction for details). When computing Bayes Factors, we used BANOVA with the default Cauchy prior of 0.707.

Results

Pre-registered analyses

Compering AEI size as a function of analytic factor Although the focus was on the descriptive results (see Table 2), it is important to mention that all the AEIs reached our pre-set threshold for NHT and BF significance ($p < 0.001$, $BF > 3$; see Table S1 in Supplementary Materials). This ensures that we are dealing with real effects rather than statistical errors. One exception is the interaction between congruency and step in the I-D data-set, which reached the NHT threshold, but its BF results were inconclusive ($BF_{10} = 0.921$).

Typical calculation in I-D reports As seen in Table 2, employing this analytic procedure caused a considerable closing of the gap between the two paradigms in terms of AEI size, with 16.4 ms in the I-D paradigm and 30.6 ms in the NEXT paradigm (a decrease of almost 20 ms as compared to the effect size when employing the original analytic procedure). Although Braem et al. (2019) did not apply log transformation on RT, many reports of experiments using the I-D paradigm did. We, therefore, wanted to examine the influence of this analytic aspect on the results. Indeed, when applying $\log(RT)$ transformation, we found an even smaller difference between the two paradigms with AEI size of 14.6 ms in the I-D paradigm and 22.4 ms in the NEXT paradigm (i.e., decrease of 1.8 ms and 8.2 ms, respectively).

The first block, first mini-block, and Steps 5–16 Adding the first block and the first mini-block (as done in most of the reports on the NEXT paradigm) did not make much of a difference, with AEI size being 15.6–16.4 ms in the I-D paradigm and 30.6–31.8 ms in the NEXT paradigm (i.e., a decrease of up to 0.8 ms and 1.2 ms, respectively, as compared to the first calculation). A similar picture emerged when only Steps 2–5 were analyzed (and thus the number of steps in each mini-block was equal for both paradigms), with AEI size of 14.4 ms in the I-D paradigm and 32.0 ms in the NEXT paradigm (i.e., decrease of 1.4 ms and increase of 0.2 ms, respectively). Hence, it is safe to say that these three factors are not responsible for the differences in AEI.

The first step Next, we included data from the first step (as typically done in NEXT reports). While the inclusion of the first step barely influenced AEI size in the I-D paradigm (18.4 ms, an increase of about 4 ms), this inclusion increased AEI size in the NEXT paradigm substantially, by nearly 20 ms, to 50 ms.

Applying typical NEXT analysis We employed a two-way design including congruency and step (as typically calculated in reports of the NEXT paradigm). By doing so, we could examine AEI separately for each step, and specifically, focus on the first step. The results, presented in Table 2, show that the largest difference between the two paradigms is found in the first step, indicating AEI of 37 ms in the I-D paradigm and 72.4 ms in the NEXT paradigm.

Correlation between AEI and execution performance In Meiran et al. (2016), there was a positive correlation between AEI and performance in the Execution phase (AEI being a signature of poor performance), while in Braem et al. (2019), the correlation was negative (AEI being indicative of good performance). To shed light on the potential causes for this discrepancy, we examined the aforementioned correlation in Meiran et al.'s data-set according to Braem et al.'s (2019) analytic approach. In line with Meiran et al.'s (2016) original analyses, we found a positive correlation between AEI and performance in the first Execution step in RT, $r = 0.610$, $p < 0.001$ ($n = 166$). When examining the correlation concerning the dependent variable percentage of errors (PE), the correlation was insignificant, $r = 0.030$, $p = 0.720$ ($n = 166$).

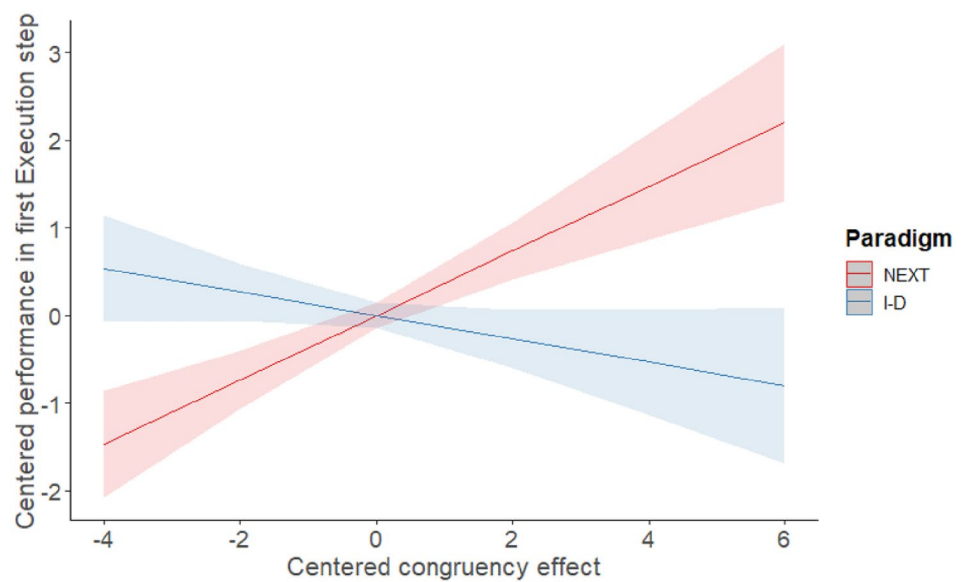
To verify that these differences between the paradigms (and samples) are statistically robust, we conducted an additional analysis that was mistakenly not pre-registered. In this analysis, a multiple regression model was calculated to predict RT in the first Execution step based on the independent variables AEI size, paradigm, and the interaction between paradigm and AEI (see Fig. 2). Due to the very different RTs, both AEI and performance in the first Execution step were centered (sample mean subtracted from individual participants' values). A significant regression equation was

Table 2 Descriptive statistics of congruency effect in each paradigm for each comparison step

Comparison	Factor	Paradigm									
		Inducer-Diagnostic paradigm				NEXT paradigm					
		Congruency									
		Congruent		Incongruent		AEI		Congruent		Incongruent	
<i>M</i>	<i>M</i>	<i>M</i>	SE	<i>M</i>	SE	<i>M</i>	<i>M</i>	<i>M</i>	SE		
1	✗	Log(RT) transformation	643	659	16.4	2.9	397	428	30.6	2.8	
	✗	Inclusion of first block									
	✗	Inclusion of first mini-block									
	✗	Inclusion of first step									
	✓	Inclusion of steps 5–16									
	✗	Congruency X step interaction									
2	✓	Log(RT) transformation	628	643	14.6	2.6	388	410	22.4	2.8	
	✗	Inclusion of first block									
	✗	Inclusion of first mini-block									
	✗	Inclusion of first step									
	✓	Inclusion of steps 5–16									
	✗	Congruency X step interaction									
3	✗	Log(RT) transformation	650	666	16.4	2.6	397	428	30.6	3.4	
	✓	Inclusion of first block									
	✗	Inclusion of first mini-block									
	✗	Inclusion of first step									
	✓	Inclusion of steps 5–16									
	✗	Congruency X step interaction									
4	✗	Log(RT) transformation	651	666	15.6	2.5	404	435	31.8	3.6	
	✓	Inclusion of first block									
	✓	Inclusion of first mini-block									
	✗	Inclusion of first step									
	✓	Inclusion of steps 5–16									
	✗	Congruency X step interaction									
5	✗	Log(RT) transformation	644	658	14.4	3.2	404	436	32.0	3.8	
	✓	Inclusion of first block									
	✓	Inclusion of first mini-block									
	✗	Inclusion of first step									
	✗	Inclusion of steps 5–16									
	✗	Congruency X step interaction									
6	✗	Log(RT) transformation	675	693	18.4	3.5	441	491	50.0	5.3	
	✓	Inclusion of first block									
	✓	Inclusion of first mini-block									
	✓	Inclusion of first step									
	✗	Inclusion of steps 5–16									
	✗	Congruency X step interaction									
7	✗	Log(RT) transformation	step								
	✓	Inclusion of first block	1	786	823	36.9	6.8	494	566	72.4	7.6
	✓	Inclusion of first mini-block	2	636	653	17.8	5.1	402	441	38.7	5.0
	✓	Inclusion of first step	3	656	663	6.2	5.8	404	431	27.6	5.1
	✗	Inclusion of steps 5–16	4	639	656	16.5	5.2	406	427	21.0	7.1
	✓	Congruency X step interaction									

M represents mean RT; SE represents standard error. Factor column presents the inclusion/exclusion of each of the 6 factors in the current comparison. For instance, comparison 1 includes Steps 5–16 in calculating AEI but excludes all other factors

Fig. 2 Correlation between congruency effect and performance in the first Execution step in both I-D and NEXT paradigms



found, $F(3, 330) = 9.07$, $p < 0.001$, $BF_{10} = 733.210$ with an $R^2 = 0.08$. Only the interaction between paradigm and AEI significantly predicted performance in the first Execution step, $b = 0.50$, $t(330) = 4.74$, $p < 0.001$, $BF_{10} > 1000$. These results strengthen our conclusions regarding a real difference between paradigms (or participants' samples) in terms of their individual differences.

Additional (non-pre-registered) analyses

Because the following analyses were not pre-registered, we decided not to include their introduction in the general Introduction to avoid a false impression that these issues were considered before the study was conducted.

Controlling for RT range using Vincentizing One of the most noticeable differences between the two paradigms is the RT range. While in the NEXT paradigm, the RT range is ~350–600 ms, in the I-D paradigm, the RT range is ~550–1000 ms. It has been previously suggested that the differences are due to WM load. According to this idea, the difference between paradigm in the Interference task involves four different task rules (two for Interference and two for Execution) in I-D and only two (Execution) rules in NEXT (Pereg & Meiran, 2019). Given these differences, one could argue that the differences between the two paradigms arise because slow Interference steps are relatively insensitive to incongruency. We decided to examine the AEI of the two paradigms in the same RT range to address this issue. To do so, we used Vincentizing (Ratcliff, 1979), which divides the RTs by percentiles for each participant in each condition and then averages across participants. Using this method, we could examine AEI size when considering only RTs that

are roughly equivalent across paradigms. By doing so, we controlled the possible differential sensitivity for incongruency across the RT range.

As we can see in Table 3, even when examining the same RT range (i.e., comparing AEI in the 15th RT percentile in Braem et al., 2019, 22 ms, with the 85th percentile in Meiran et al., 2016, 63 ms), the difference between paradigms remains substantial. Hence, we can conclude that there is no hint that AEI size is sensitive to the RT range in which it is being examined.

Another interesting result is the differences in AEI size when examining the 15th RT percentile representing the quickest, best-prepared responses (e.g., De Jong, 2000). In the current case, we can see that the paradigm differences in AEI size nearly disappear. These results hint at a possibility that one difference between the paradigms may be related to what happens in the slowest responses. We will return to this surprising finding in the “General discussion”.

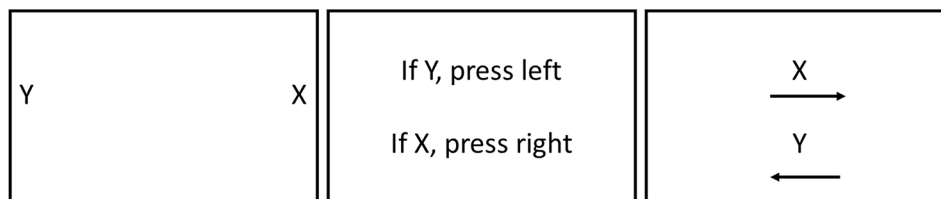
Discussion

In the current study, we compared two relatively large datasets, one from each paradigm, both employing the most typical paradigm format (Braem et al., 2019 for I-D paradigm and Meiran et al., 2016 for NEXT paradigm). We found that some aspects of the analytic approach, mostly whether the first Interference step is included in the analysis, influence AEI size. The other analytic differences we considered, such as the inclusion of the first block and first mini-block and the exclusion of Steps 5–16, did not cause any meaningful differences in AEI size. We have also observed that the log transformation influences AEI size, but mainly in NEXT, where it has never been used beforehand. Moreover, this

Table 3 Vincentizing analysis. Mean RT-quantile (in ms) as a function of Step, and Congruency

Quantile	Step	Paradigm					
		I-D paradigm			NEXT paradigm		
		Congruency					
		Congruent	Incongruent	AEI	Congruent	Incongruent	AEI
0.15	1	643	665	22	396	421	25
0.85	1	934	990	56	603	709	106
0.15	2-end	507	514	7	330	332	2
0.85	2-end	791	813	22	480	543	63

Fig. 3 The presentation of instructions in each condition of Instructions type. Left panel: concrete instructions. Middle panel: abstract instructions. Right panel: spatial controlled



issue seems to have minimal implication since the log transformation is no longer being used.

The influence of including the first step is in line with previous results showing that this step produces the largest AEI (e.g., Meiran et al., 2015). As reviewed above, the consideration of including the first step in the analysis reflects the different theoretical emphases of the I-D team and the NEXT team, and the current results cannot inform this debate. Nonetheless, it is clear that any attempt to draw general conclusions must consider this difference. For future studies, we recommend that, at minimum, researchers should provide the required information (e.g., AEI size in the first step) that is needed for other teams to draw conclusions.

Despite the aforementioned influence of the analytic approach, some evidence emphasizes that this is not the entire story. First, the influence of the analytic approach was asymmetrical. Specifically, applying the I-D analytic approach (in which the first Interference step is omitted) to results from the NEXT paradigm has shrunk AEI by nearly 39% (and almost 45% when applying log transformation). In contrast, applying the NEXT analytic approach to results from the I-D paradigm (i.e., keeping the first Interference step) enlarged AEI by only 12%. This could imply that the I-D paradigm is less sensitive to this aspect of the analytic approach, at least regarding AEI.

Importantly (for Study 2), the correlational analyses and Vincentizing further show that some substantial differences between the paradigms remain even after equating the analytic procedures. Therefore, further investigation of the influence of additional factors not examined in the current study is needed. Study 2 explored two potential factors that could further explain the differences in AEI size between the I-D paradigm and the NEXT paradigm.

Study 2

Study 2 examined the influence of two factors: (1) instruction's presentation mode and (2) choice complexity in the Interference phase. As outlined in the "Introduction", these are two differences that are most promising in terms of actually explaining paradigm differences. Additionally, these two factors are most interesting theoretically, especially concerning the nature of the underlying WM representation and the influence of WM load on AEI size.

To achieve this goal, we used the NEXT paradigm with two core adjustments. The first was the manipulation of instruction presentation mode ("instruction" for short). We compared three instruction types: abstract, concrete, and spatial control (see Fig. 3). The abstract instructions were identical to the instructions that are presented in the I-D paradigm. The concrete instructions were identical to the instructions that are presented in the NEXT paradigm. The spatial-control instructions were added to examine the notion that the critical difference between the preceding instruction types is not the abstraction level but the location of the stimuli on the screen. This third instruction type was planned to be employed in the examination only if we find a difference between the preceding instruction types. Based on Cole et al.'s (2017) and Dreisbach's (2012) theories, we expected smaller AEI size in the abstract-instructions condition compared to the concrete condition. However, it is essential to keep in mind that some results from previous studies reviewed beforehand argue against this prediction (Liefoghe et al., 2012; Longman et al., 2019). Specifically, Liefoghe et al. (2012) found evidence suggesting that abstract

representation is formed by default (though they did not examine instructions that are as concrete as those in NEXT). Longman et al. (2019) similarly found that stimulus representation's required degree of abstractness did not influence AEI size.

The second adjustment was the manipulation of choice complexity in the Interference phase. We compared between two complexity levels: (1) a single fixed key (or low complexity), as in NEXT, and (2) a choice between two alternative responses (or high complexity), as in I-D (see Fig. 4). Despite mixed results in the literature (Pereg & Meiran, 2019 compared to Braem et al., 2019), we tentatively predicted that choice complexity would not influence AEI size. The prediction is mainly based on the findings from a related (yet different) manipulation of WM load in the NEXT paradigm, which was applied to the number of Execution (GO) response alternatives (Pereg & Meiran, 2019) and a similar lack of effect on AEI size of the number of familiar rules (Meiran & Cohen-Kdoshay, 2012), with yet another paradigm.

Due to an error in the required sample size calculation, we ran the experiment twice, first in the lab and second as an online experiment to collect data from additional participants. The second part of the experiment was executed online because of the COVID-19 pandemic, the related lockdowns, and social distancing guidelines. Pre-registrations, data-sets, and analysis code for the current study are available at <https://osf.io/f5awd/registrations>.

Methods

Participants

Thirty Ben-Gurion University of the Negev students participated in the lab part of the experiment (27 women, mean age = 22.7, SD = 1.3). Sixty-nine similar students participated in the online part of the experiment (61 women, mean age = 23, SD = 1.2). In total, 99 participants participated in the study (88 women, mean age = 22.9, SD = 1.3), all in return for course credit. The sample size was determined based on a power analysis using G-Power 3.1.9.4 (Faul et al., 2007) that was set to obtain between-within interaction equivalent to $\eta^2_p = 0.1$ with a power of 0.95 and Alpha of 0.001. We examined two such interactions, one between congruency (2) and instructions (3) (requiring $N = 72$) and one between congruency and choice complexity (requiring $N = 62$). We ran more participants to be on the safe side. All the participants signed an informed consent form, reported normal or corrected-to-normal vision, including intact color vision, and were not diagnosed as suffering from attention deficits.

Design and assignment to groups

The participants were assigned to the six groups according to their serial number (N per group, and their demographic information can be seen in Table S2 in Supplementary Material). The groups were defined by two independent between-participant variables: instructions (3) and choice complexity (2). Congruency and step were manipulated as within participants variables. Congruency had two levels: Congruent and Incongruent. Step had four levels, Step 1 through 4. (Although there were up to five Interference steps, Steps 4 and 5 were combined to ensure a sufficient amount of data).

Materials and procedure

We adapted the NEXT paradigm used by Meiran et al. (2015). The stimuli were randomly drawn from a pool of 226 stimuli, consisting of 26 English letters, 10 digits, 24 Hebrew letters (Hebrew is the language of our participants who also master the English alphabet), 24 symbols (e.g., arithmetic symbols), and 142 pictures (e.g., shapes and different objects). Most stimuli were imported from the Microsoft PowerPoint symbol pool. Some pictures were sketches drawn from accessible Internet image search databases. Stimulus size was 3×3 cm; digits and letters appeared in a Calibri font. The two stimuli chosen in each choice task came from the same stimulus group (e.g., two digits, two pictures) to prevent participants from employing a simple rule that could be repeated such as “digit \rightarrow right, letter \rightarrow left”. Each stimulus was used only once during the experiment.

The paradigm included 80 mini-blocks, divided into four blocks. Each mini-block consisted of a unique two-choice task involving two stimuli arbitrarily mapped to a right and left key (L and A on a QWERTY keyboard). All mini-blocks started with an Instruction screen. This screen was presented until the participant pressed the spacebar. It was followed by an Interference phase. The length of this phase varied between 0 and 5 steps, with the length being randomly selected from a close-to exponential distribution with a 30% chance for one step, 20% chance for 2–3 steps, and a 10% chance for each of 0, 4 or 5 steps. The Execution phase followed the Interference phase and consisted of only two steps. In these steps, the stimuli were randomly chosen (with replacement), and participants had to apply the new instructions. This implies that, in some mini-blocks, only one stimulus was chosen, and it was chosen twice. Finally, a feedback screen was presented, reporting the percentage of errors and mean RT from the Execution phase.

Instructions were manipulated between participants. In the instruction screen, two stimuli were presented in white

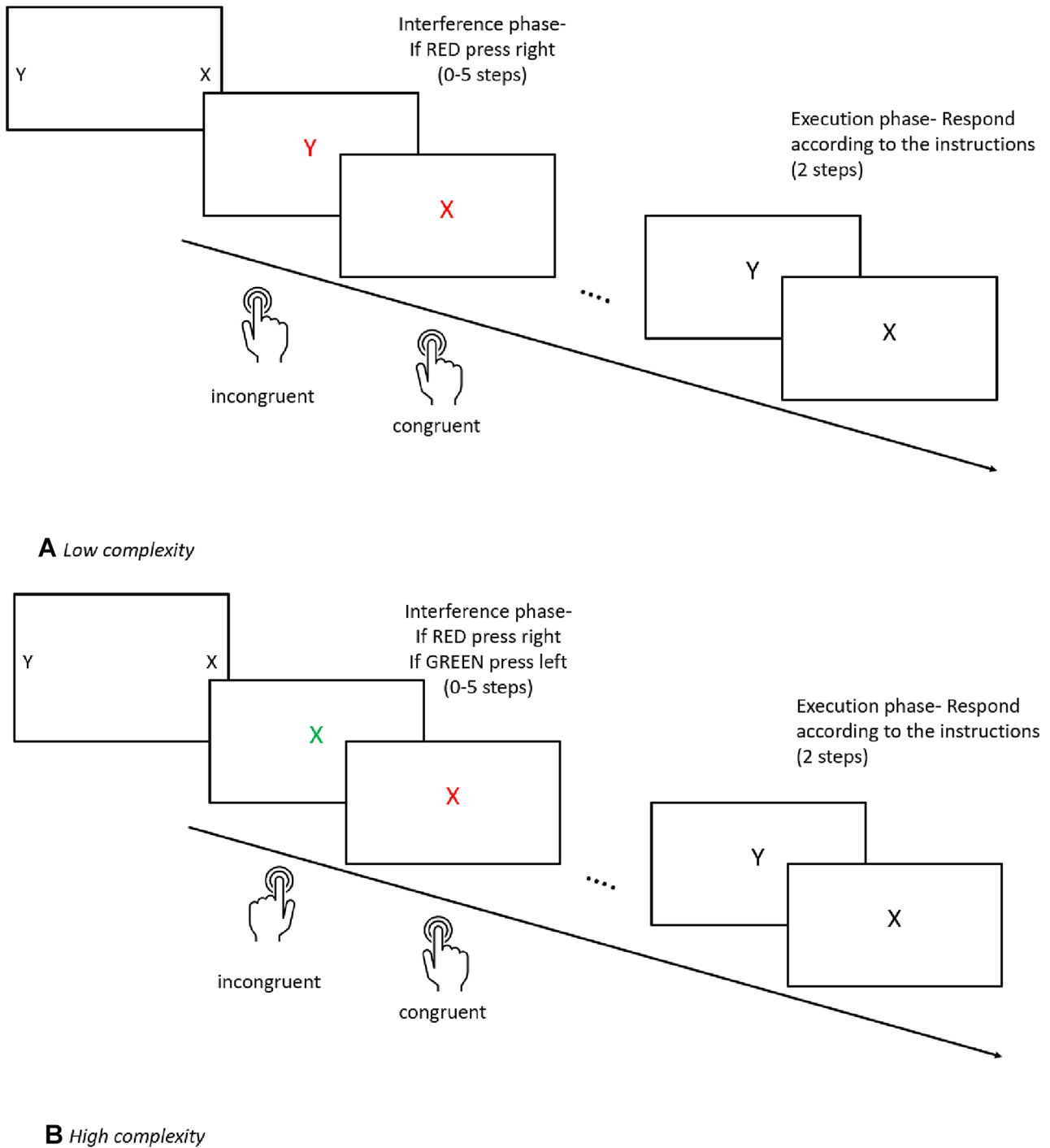


Fig. 4 The presentation of instructions in each condition of *choice* complexity

color. In the condition of the concrete instruction, one stimulus appeared on the right side of the screen and the other on the left (each stimulus was placed 15 cm from the center of the screen). In the condition of the abstract instruction, both stimuli appeared at the center of the screen, one below the other, with verbal directions attached to it (e.g., "If

X press left" appearing above "If Y press right"). In the spatial-control condition, both stimuli appeared at the center of the screen, one below the other, with arrows indicating the appropriate reaction for each stimulus (see Fig. 3). This condition, which employed stimuli presented at the center (like in the abstract instructions) but involved relatively concrete

mode (like in the concrete instructions), was added to enable a better-controlled comparison between the three conditions and to ensure that the differences between concrete and abstract instructions, if existing, are not caused by stimulus location. Participants were required to place their fingers on the response keys and be ready for the Execution task.

Unlike in previous NEXT experiments, in which the green color indicated the Execution phase, we used white color to indicate the transition to the Execution phase. Choice complexity was manipulated (between participants) in the Interference phase, which preceded the Execution phase. In the low complexity condition, the target stimulus in the Interference phase was always presented in red color, indicating to press a fixed key that was introduced at the beginning of the experiment. In the high complexity condition, the stimulus was presented either in red color or in green color. Participants' task in the Interference phase was to indicate the color of the stimulus (e.g., "If green press right, if red press left"). This mapping between colors and responses was introduced at the beginning of the experiment and remained valid throughout the experiment (see Fig. 4). When the Interference phase ended, the Execution phase immediately started. After which, a new mini-block began. To further ensure high alertness, participants had short breaks between blocks.

Testing mode In the lab version, Participants were tested individually, in small lab rooms. The experiment was run on PCs equipped with 19-in. monitors. The software was written in E-Prime 2.0 (Psychology Software Tools, 2010). The same procedure and materials were used in the online part of the experiment except for few changes. First, participants performed the experiment using their own computer with software written in OpenSesame-Web 3.3.6 (Mathôt et al., 2012) and exported to JATOS server (<https://www.jatos.org/>). Second, since the experiment was performed online, after registering the experiment, participants received a zoom (<https://zoom.us/>) link to a video meeting with the experimenter at the time of the experiment. In the zoom meeting, the experimenter explained the experiment and sent a JATOS link to the participant. The zoom meeting lasted until the participant ended the experiment to ensure that he/she did not have any technical problems. However, it is important to mention that during the experiment, the sound and camera of both the participants and the experimenter were turned off and were used only in the case of technical problems.

Dependent variables and data analysis

The focus of Study 2 was on three interaction effects: (1) AEI and instruction, (2) AEI and choice complexity, and (3) AEI, instruction, and choice complexity (Note that the

last interaction was erroneously not pre-registered). All the interactions contained step as an independent variable to enable examining the first step separately. To allow examination of the unique contribution of the interactions, BFs represent the comparison between (1) H0 model containing all main effects and all lower levels interactions (if exist) and (2) H1 model, which also includes the relevant interaction. RT and PE served as dependent variables. In addition to the examination of the Interference phase, we also examined the Execution phase. Respective analyses are reported in Supplementary Materials.

Before performing the core analyses of interest, we wanted to reject the possibility that results are influenced by testing mode (lab/internet). To do so, we examined all the interaction effects involving testing mode using B/ANOVAs.²

Data cleaning procedure Eight participants were removed from the analysis due to below chance accuracy rates or accuracy rate falling 2.5 SD below the average of the entire sample (calculated separately for the Interference and Execution phase; three participants due to poor Execution phase performance, five participants due to poor Interference phase performance); two additional participants were removed from the analysis because their post-experimental debriefing revealed that they did not perform the task as

² When examining the influence of testing mode, 20 interactions were computed (10 for each phase; 5 in RT and 5 in PE). Two out of the 20 interactions had significant *p*-value and/or had BF value that enabled H1 acceptance: (1) The 2-way interaction between testing mode and congruency in PE, $F(1, 75) = 14.06$, $p < 0.001$, $^2_p = 0.056$, $BF_{10} > 1000$ (indicating a larger AEI in lab testing, 5.48% as compared to 0.35% in online testing); (2) The triple interaction among testing mode, congruency and instruction in RT, $F(2, 75) = 2.82$, $p = 0.066$, $^2_p = 0.005$, $BF_{10} = 12.015$. Since the spatial-control condition was added only as an additional reference, we examined if the interaction among testing mode, Congruency, and instruction in RT remains significant when omitting spatial-control and comparing the abstract and concrete conditions. The results show insignificant interaction, with an inconclusive $BF_{10} = 0.417$ (Full details are provided in Table S5). To ensure that any conclusions regarding testing mode do not reflect some speed-accuracy tradeoffs, all the aforementioned ten interactions (five for each phase) were tested once more, now on an integrated speed-accuracy measure, the Linear Integrated Speed–Accuracy Score (LISAS, Vandierendonck, 2017). Only one out of the ten interactions had BF value that enabled H1 acceptance (but showed non-significant *p*-value): The triple interaction among testing mode, congruency and instruction, $F(2, 75) = 2.60$, $p = 0.081$, $^2_p = 0.005$, $BF_{10} = 13.425$. Again, we examined if the interaction among testing mode, Congruency, and instruction was significant when comparing only the abstract and concrete conditions and found it inconclusive, $BF_{10} = 1.072$ (Full details are provided in Table S6 in the Supplementary Materials). Given the Testing mode results, and to be on the safe side, the *B/ANOVA* of the interaction between congruency and instruction in RT will be examined once on the entire sample and once in each testing mode, separately.

Table 4 Study 2- RT B/ANOVA results from Interference phase

Analysis	Effect	RT					
		<i>F</i>	<i>p</i>	<i>df</i>	MSE	BF10	η_p^2
AEI replication	Congruency	31.14	<0.001	(1, 86)	5127.81	> 1000	0.266
	Step	122.48	<0.001	(2.24, 192.30)	7189.73	> 1000	0.587
	Congruency X step	1.69	0.170	(3, 258)	2246.67	0.004	0.019
Influence of instructions on AEI	Instruction	0.02	0.985	(2, 84)	188,706.77	0.078	<0.001
	Instruction X Congruency	0.89	0.413	(2, 84)	5140.46	0.019	0.021
	Instruction X Congruency X step	1.37	0.228	(6, 252)	2227.61	<0.001	0.032
Influence of choice complexity on AEI	Choice complexity	156.29	<0.001	(1, 85)	65,718.10	> 1000	0.648
	Choice complexity X congruency	8.27	0.005	(1, 85)	4728.95	800.562	0.089
	Choice complexity X congruency X step	0.83	0.476	(3, 255)	2250.99	0.008	0.010
Combined influence of choice complexity and instructions on AEI	Choice complexity X instruction	0.12	0.884	(2, 81)	68,627.79	0.069	0.003
	Choice complexity X instruction X congruency	2.48	0.090	(2, 81)	4558.03	7.762	0.058
	Choice complexity X instruction X congruency X step	0.72	0.631	(6, 243)	2245.89	<0.001	0.631

BF10 indicates whether the addition of the interaction would meaningfully improve the fit of the model. Results appear in bold in cases in which the results violated the assumption of sphericity in NHT, and hence were corrected using the Greenhouse–Geisser estimates of sphericity (see Table S7 in Supplementary Material)

instructed; two due to faulty data files. The final sample, after participants' exclusion, included 87 participants.

Steps with RT quicker than 100 ms were omitted from the analysis (0.04% of the steps), as well as steps in which RT exceeded 3 SDs from the mean RT as computed per participant, per task, and per condition (1.76% of the remaining steps). Finally, for Interference phase analyses, mini-blocks in which participant made an error in the first Execution step were excluded (8.23% of remaining Interference steps). These mini-blocks were excluded because the error could indicate that the instructions for the novel task were not properly implemented. In total, 9.82% of the trials were omitted. Finally, for the RT analyses (in both phases), only correct responses were included.

Results

Full inferential results can be seen in Tables 4 (for RT analyses) and 5 (for PE analyses).

Interference phase

AEI replication First, we tested whether the AEI, as presented in Meiran et al. (2015), replicates (see Fig. 5). As expected, participants had significantly quicker responses in congruent steps ($M_{RT}=539$ ms, $SE=16.5$) as compared to incongruent steps ($M_{RT}=569$ ms, $SE=16.5$; $p<0.001$, $BF10>1000$). In addition, participants made less errors in congruent steps ($PE=3.3\%$, $SE=0.7\%$) as compared to incongruent steps ($PE=5.3\%$, $SE=0.7\%$). Although

this result did not reach our pre-set threshold for NHT significance, it clearly reached significance in the BF analysis ($p=0.007$, $BF10>1000$). In addition, participants were significantly slower in the first step³ ($M_{RT}=646$ ms, $SE=17.0$) as compared to the following steps (2–5, respectively; $M_{RT-2}=514$ ms, $M_{RT-3}=531$ ms, $M_{RT-4+5}=526$ ms, $SE=17.0$; $p<0.001$, $BF10>1000$). In contrast to the RT results, participants did *not* make more errors in the first step as compared to the following steps ($PE=4.8\%$, 3.9% , 3.7% , and 4.6% , in Steps 1,2,3,4+5, respectively, $SE=0.7\%$; $p=0.100$, $BF10=0.003$). Similar to some previous results (for example, see Meiran et al., 2015; Experiment 1), the interaction between congruency and step was not significant (and actually allowed H_0 acceptance) both in RT and in PE ($p=0.170$, $BF10=0.004$; $p=0.652$, $BF10<0.001$; respectively).

The influence of instructions on AEI Second, we examined main effect for instruction and the interaction among congruency, step, and instruction (see Fig. 6). Contrary to our predictions, there were no differences among the three levels of instruction both in RT (Abstract = 552 ms, $SE=28.7$; Concrete = 558 ms, $SE=28.9$; Spatial con-

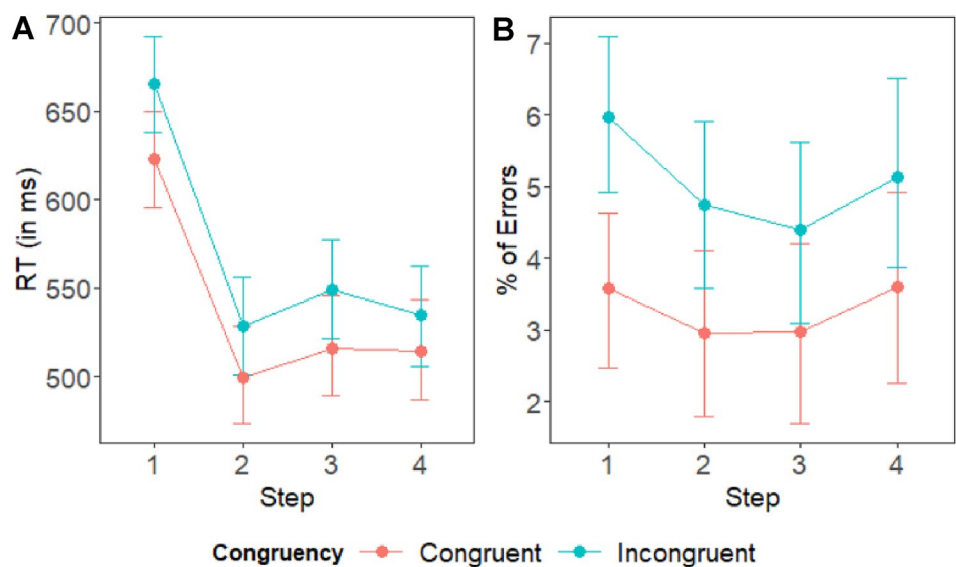
³ In the B/ANOVA, the computation of effects involving step violated the assumption of sphericity both in RT and in PE, which was indicated by Mauchly's test. Therefore, degrees of freedom were corrected using the Greenhouse–Geisser correction. Mauchly's test parameters and the GG estimates values are presented in Table S7 in Supplementary Materials. Results from G–G-corrected effects are presented in Tables 4 and 5 in bold.

Table 5 Study 2-PE B/ANOVA results from Interference phase

Analysis	Effect	PE					
		<i>F</i>	<i>p</i>	<i>df</i>	MSE	BF10	η^2_p
AEI replication	Congruency	7.70	0.007	(1, 86)	0.01	> 1000	0.082
	Step	2.16	0.100	(2.70, 231.99)	0.00	0.003	0.025
	Congruency X step	0.51	0.652	(2.63, 226.05)	0.00	< 0.001	0.006
Influence of instructions on AEI	Instruction	0.72	0.491	(2, 84)	0.02	0.016	0.017
	Instruction X congruency	0.61	0.546	(2, 84)	0.01	0.008	0.014
	Instruction X congruency X step	1.00	0.423	(5.23, 219.48)	0.00	< 0.001	0.023
Influence of choice complexity on AEI	Choice complexity	32.03	< 0.001	(1, 85)	0.02	> 1000	0.274
	Choice complexity X congruency	1.00	0.320	(1, 85)	0.01	0.098	0.012
	Choice complexity X congruency X step	0.19	0.882	(2.63, 223.53)	0.00	0.001	0.002
Combined influence of choice complexity and instructions on AEI	Choice complexity X instruction	2.00	0.142	(2, 81)	0.02	0.075	0.047
	Choice complexity X instruction X congruency	1.32	0.274	(2, 81)	0.01	0.045	0.032
	Choice complexity X instruction X congruency X step	1.64	0.146	(5.25, 212.55)	0.00	0.002	0.039

BF10 indicates whether the addition of the interaction would meaningfully improve the fit of the model. Results appear in bold in cases in which the results violated the assumption of sphericity and hence were corrected using the Greenhouse–Geisser estimates of sphericity (see Table S7 in Supplementary Material)

Fig. 5 Study 2—AEI replication in **A** Reaction Times (RT) and **B** Percentage of Errors (PE). Error bars represent 89% credible intervals (McElreath, 2018)



trol=553 ms, SE=28.1; $p=0.985$, $BF10=0.078$) and in PE (Abstract=4.5%, Concrete=3.3%, Spatial control=5.0%, SE=1.0%; $p=0.491$, $BF10=0.016$). More importantly, and contrary to our predictions, both the 2-way interaction between instruction and congruency and the triple interaction among instruction, congruency and step were non-significant, in RT ($p=0.413$, $BF10=0.019$; $p=0.228$, $BF10<0.001$, respectively) and in PE ($p=0.546$, $BF10=0.008$; $p=0.423$, $BF10<0.001$, respectively) and actually permitted accepting H0. We also conducted BANOVA of the interaction between congruency and instruction in RT separately for each testing

mode. The interaction in both testing modes enabled H0 acceptance ($BF10_{online}=0.001$, $BF10_{lab}=0.004$).

Because our goal was to compare abstract instructions with concrete instructions, we conducted contrast analysis (combined for both testing mode) that compared between the two instruction conditions in terms of congruency and step effects, represented as a 1-*df* contrast. This analysis also showed a non-significant effect in RT and PE, with BF permitting acceptance of H0 ($t(252)=0.21$, $p=0.831$, SE = 80.50, $BF10=0.045$; $t(252)=1.85$, $p=0.066$, SE = 0.08, $BF10=0.006$; respectively).

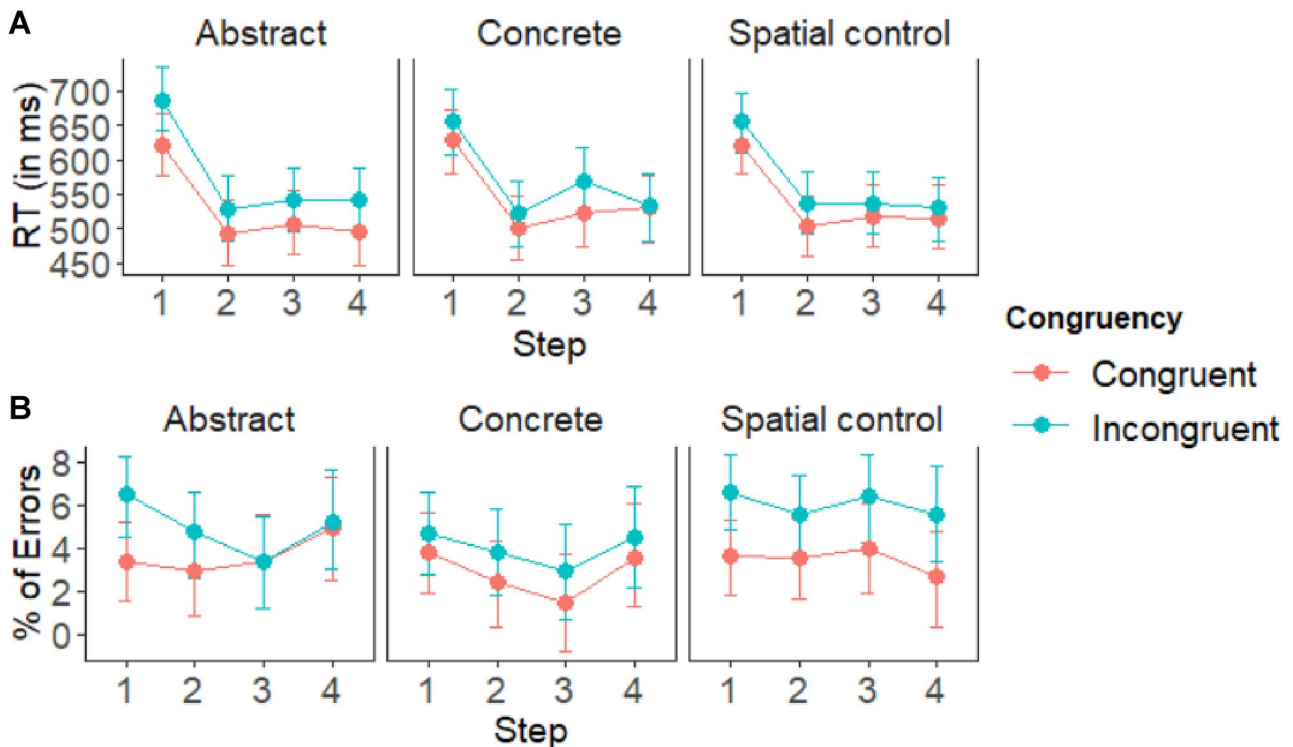


Fig. 6 Study 2—Performance as a function of congruency, step, and instruction in **A** Reaction Times (RT) and **B** Percentage of Errors (PE). Error bars represent 89% credible intervals (McElreath, 2018)

The influence of choice complexity on AEI Third, we examined the main effect of choice complexity, the 2-way interaction between congruency and choice complexity, and the triple interaction among congruency, step, and choice complexity (see Fig. 7). Participants were significantly quicker in the low complexity condition ($M_{RT}=433$ ms, $SE=13.7$) as compared to the high complexity condition ($M_{RT}=676$ ms, $SE=13.7$; $p<0.001$, $BF_{10}>1000$). Participants also made significantly less errors in the low complexity condition ($PE=1.5\%$, $SE=0.7\%$) as compared with the high complexity condition ($PE=7.0\%$, $SE=0.7\%$; $p<0.001$, $BF_{10}>1000$). These results validate the efficiency of our manipulation. As predicted, the triple interaction was non-significant, both in RT ($p=0.476$, $BF_{10}=0.008$) and in PE ($p=0.882$, $BF_{10}=0.001$), with results allowing H0 acceptance. Despite that, and contrary to our predictions, the 2-way interaction between congruency and choice complexity in RT reached significance in the BF analysis (albeit not reaching our pre-set threshold for NHT significance; $p=0.005$, $BF_{10}=800.562$). This 2-way interaction was not significant in PE ($p=0.320$, $BF_{10}=0.098$) and allowed H0 acceptance. The RT results are especially surprising because they present the opposite pattern to that presented in the Introduction. Here, the low complexity condition (similar to

NEXT) had smaller AEI ($M_{\text{difference}}=17$, 89% CI [9, 24]) as compared to the high complexity condition (similar to I-D, $M_{\text{difference}}=46$, 89% CI [39, 54]).

The joint influence of choice complexity and instruction on AEI The 2-way interaction between choice complexity and instruction and the 4-way interaction among choice complexity, instruction, congruency, and step were both non-significant in RT ($p=0.884$, $BF_{10}=0.069$; $p=0.631$, $BF_{10}<0.001$; respectively) and allowed H0 acceptance. The 3-way interaction among choice complexity, instruction, and congruency was non-significant in RT but showed BF value that enabled H1 acceptance ($p=0.090$, $BF_{10}=7.762$; see Fig. 8). All of the interactions were non-significant in PE ($p=0.142$, $BF_{10}=0.075$; $p=0.274$, $BF_{10}=0.045$; $p=0.146$, $BF_{10}=0.002$; respectively) and permitted H0 acceptance.

The results from Study 2 contradict our first hypothesis: instruction did not influence AEI size. As predicted, a 3-way interaction among choice complexity, congruency, and step was not found. Despite that, the 2-way interaction between choice complexity and congruency was found, but its pattern was opposite to that we predicted. Specifically, it indicated a larger AEI in the high complexity condition (resembling the

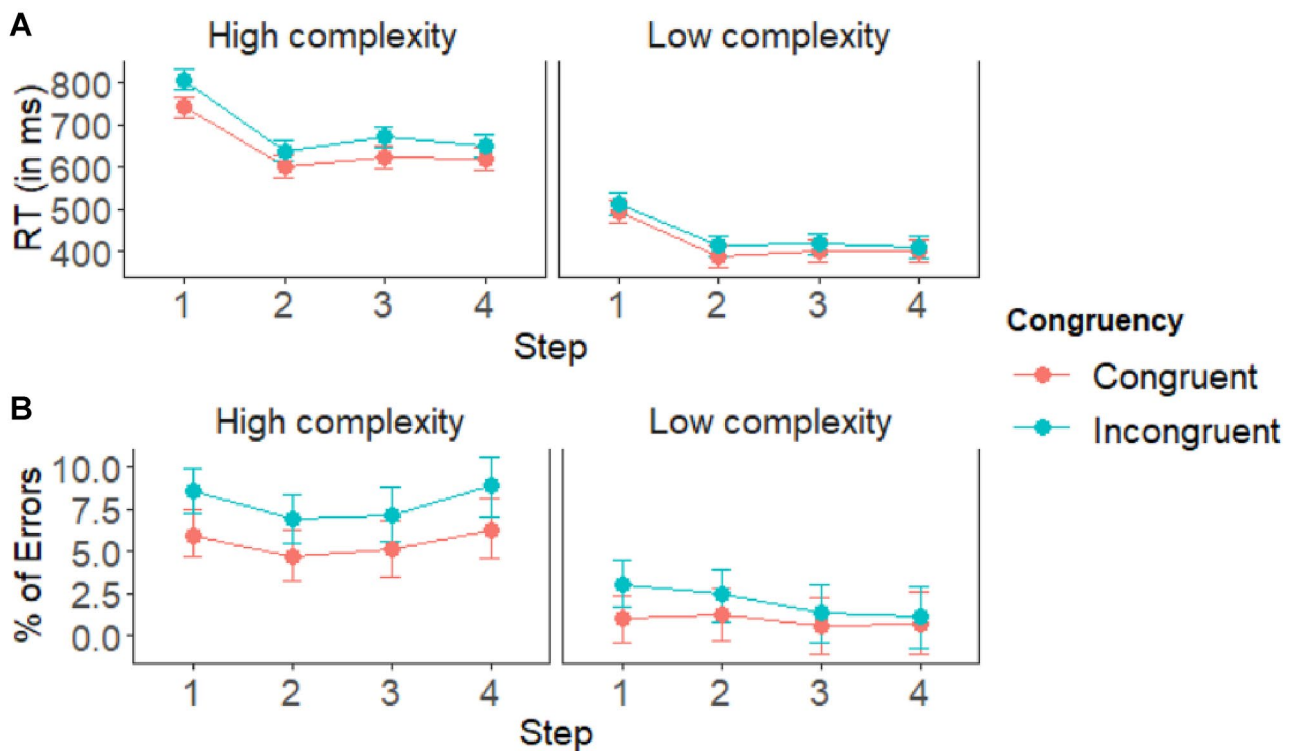


Fig. 7 Study 2—Performance as a function of congruency, step, and choice complexity in **A** Reaction Times (RT) and **B** Percentage of Errors (PE). Error bars represent 89% credible intervals (McElreath, 2018)

I-D paradigm). In addition, the discrepancy in AEI between high/low complexity was more pronounced with abstract than with concrete instructions. This result should be considered carefully for two main reasons: (1) The analysis was not pre-registered, and we did not have any hypothesis regarding it. (2) The results were inconsistent across inferential methods. Hence, a replication of this effect is needed. To summarize, results from the current study indicate that both instruction and choice complexity does *not* cause the expected differences in AEI size between NEXT and I-D.

General discussion

Previous studies indicate that AEI size tends to be considerably smaller in reports using the I-D paradigm (Liefoghe et al., 2012) than those using the NEXT paradigm (Meiran et al., 2015). The present work examined potential factors that may be responsible for this discrepancy. In Study 1, we exploratorily examined the influence of six differences in the analytic methods. Study 2 focused on two additional factors: instruction format and choice complexity in the Interference phase.

Results from Study 1 show rather marked effects of the analytic approach, but some paradigm differences remained

even when equating the analytic approach across paradigms. While the inclusion of the first block and mini-block, and the exclusion of Steps 5–16 did not contribute to paradigm differences, the inclusion of the first Interference step had a marked effect on the results. When including the first step in AEI calculation, the NEXT AEI grew by 56.25%, and in the I-D it grew by 27.78%. As aforementioned, the decision of whether to in/exclude this step reflects differences in theoretical emphases between teams of researchers. While the current examination cannot tell which team is correct in their choice, it indicates that sufficient information must be provided in future studies for any general conclusions to be drawn. Specifically, those who adopt the NEXT tradition should better (additionally) report AEI as computed after averaging Steps 2 and beyond. Those who adopt the I-D tradition should (additionally) report AEI in the first Interference step. Nonetheless, even though the analytic approach influences AEI size and its difference across paradigms, the correlational analyses and Vincentizing indicated some differences between the paradigms extending beyond analysis.

Study 2 tested two additional factors that could explain the differences between the two paradigms and could not have been examined by comparing the two existing data-sets. Contrary to our predictions, Study 2 indicates that instruction format is not a factor that causes differences between

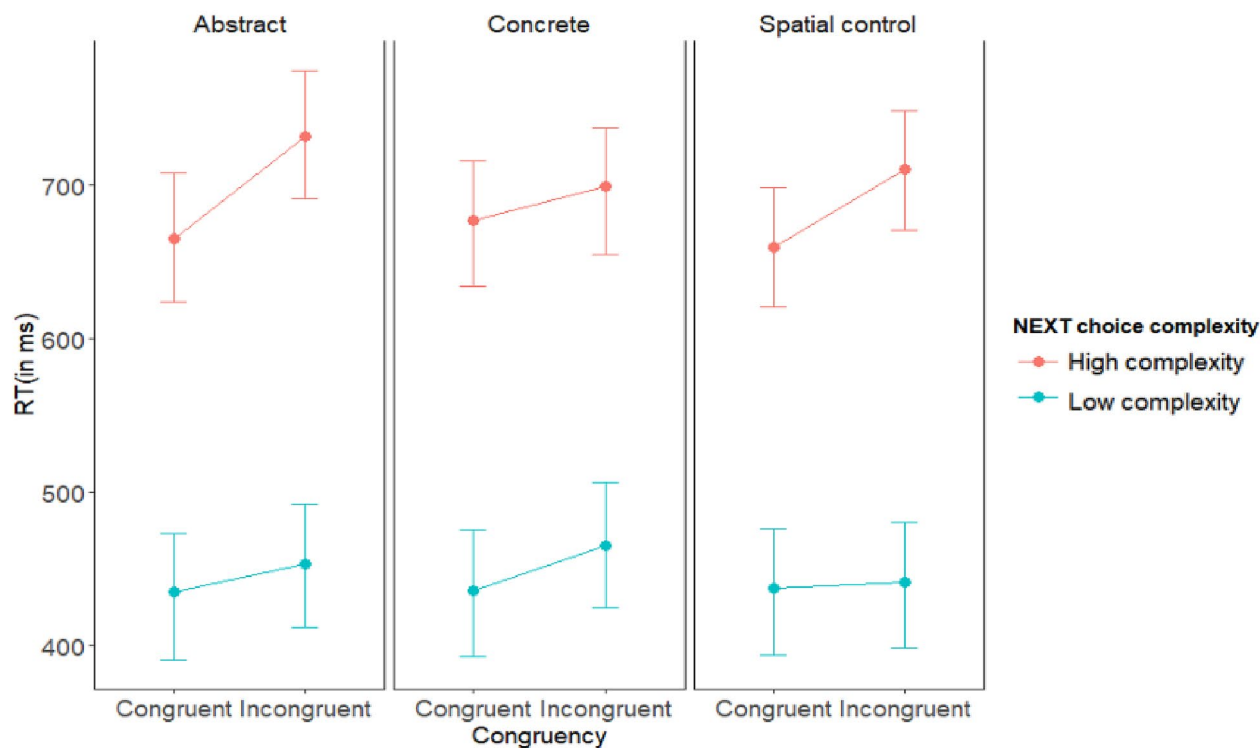


Fig. 8 Study 2—Performance as a function of congruency, choice complexity, and instruction in Reaction Times (RT). Error bars represent 89% credible intervals (McElreath, 2018)

the paradigms in AEI size. This conclusion expands previous results by Liefoghe et al., (2012, Experiment 1) and Longman et al., (2019; especially Experiment 1b) because it demonstrates, for the first time in the NEXT paradigm, that the abstraction level of the instructions does not influence AEI size. This finding accords with the notion that instructions are encoded abstractly by default (see Monsell & Graham, 2021) and extends the conclusions reached by Liefoghe et al., (2012, Experiment 1) regarding abstract representation by showing that this mode of representation does not depend on instruction presentation format.

In addition, the complexity of the Interference task cannot explain the paradigm-related differences in AEI sizes. In fact, our results from Study 2 show that an opposite to predicted pattern seems to have emerged, indicating a larger AEI in conditions that resemble those in the I-D paradigm. The surprising fact that WM load increases are associated with a larger AEI suggests that AEI may be a marker of poor performance, in accordance with Meiran et al.'s (2016) correlational analyses indicating high AEI among poor performers, and contrary to those of Braem et al. (2019) who found the opposite pattern.

After conducting the study, we realized that one procedural difference was not examined and may explain some of our results. This difference concerns certainty

level. Specifically, the Interference phase occurs in each mini-block in the I-D paradigm. In contrast, the Interference phase in the NEXT paradigm is somewhat surprising because it does not always occur (in about 10% of the mini-blocks, the Interference phase is skipped entirely, and the Execution phase immediately follows the instructions). Notably, the surprise is maximal in Step 1, where the largest paradigm discrepancy in AEI is found. This speculation is supported by the fact that the AEI differences between the paradigms almost disappeared when AEI was examined in the best-prepared trials (the 15th RT percentile). Findings from other cognitive tasks might further support this notion. For instance, in a new preprint, Gresch et al. (2021) examined the influence of temporal expectation on WM performance. They found fewer errors when interference could be temporally predicted. Less directly, perhaps, multiple studies found that congruency effects are enlarged in conditions in which incongruent steps are relatively rare (hence, less predictable, see Torres-Quesada et al., 2013 and Bulger et al., 2021). All those studies provide evidence that expectancy improves performance, and most critically, influences the ability to deal with distracting information.

Can our results suggest which paradigm or which analytic procedure is better? We doubt that. We can, however, point to some important considerations. One salient issue

concerns the inclusion of Step 1 in the analysis. Obviously, this step has the largest AEI and also shows the largest paradigm discrepancy. The other analytic differences had a relatively minuscule influence on AEI size. It, thus, seems obvious to us to recommend that future reports report their analyses with and without this 1st step to permit an efficient exchange of information across studies and research teams who employ different paradigms.

In conclusion, the current study shows that the differences in AEI size between the NEXT and I-D paradigms are partly due to the analytic approach. Still, beyond that, there remain some additional differences to which researchers should attend. Perhaps more importantly, the current study highlights the importance of considering the influences of paradigm-specific factors that are often overlooked and have important implications when attempting to draw generalizations.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00426-021-01596-1>.

Funding This research was supported by Research Grant #2015-186 from the US–Israel Binational Science Foundation to Nachshon Meiran, Todd S. Braver, and Michael W. Cole. We would like to thank Baptist Liefoghe for his kind help with providing the I-D data. Novel data are available at <https://osf.io/ctvj2/> (For Study 1), <https://osf.io/f5awd/> (For Study 2).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval In Study 1, two databases from previous studies were analyzed (Braem et al., 2019; Meiran et al., 2016), both approved by their departmental Ethics committees. Study 2 was approved by the departmental ethics committee.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

- Braem, S., Deltomme, B., & Liefoghe, B. (2019). The instruction-based congruency effect predicts task execution efficiency: Evidence from inter- and intra-individual differences. *Memory & Cognition, 47*(8), 1582–1591.
- Bulger, E., Shinn-Cunningham, B. G., & Noyce, A. L. (2021). Distractor probabilities modulate flanker task performance. *Attention, Perception, & Psychophysics, 83*(2), 866–881.
- Cole, M. W. (2009). *The biological basis of rapid instructed task learning* (Doctoral dissertation, University of Pittsburgh, Pennsylvania, USA).
- Cole, M. W., Braver, T. S., & Meiran, N. (2017). The task novelty paradox: Flexible control of inflexible neural pathways during rapid instructed task learning. *Neuroscience & Biobehavioral Reviews, 81*, 4–15.
- 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. <https://doi.org/10.3758/s13415-012-0125-7>
- De Jong, R. (2000). An intention-activation account of the residual switch costs. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 357–376). MIT Press.
- Dreisbach, G. (2012). Mechanisms of cognitive control: The functional role of task rules. *Current Directions in Psychological Science, 21*(4), 227–231.
- Everaert, T., Theeuwes, M., Liefoghe, B., & De Houwer, J. (2014). Automatic motor activation by mere instruction. *Cognitive, Affective, & Behavioral Neuroscience, 14*(4), 1300–1309.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.
- Formica, S., González-García, C., & Brass, M. (2020). The effects of declaratively maintaining and proactively proceduralizing novel stimulus-response mappings. *Cognition, 201*, 104295.
- Gresch, D., Boettcher, S., van Ede, F., & Nobre, A. C. (2021). Shielding working-memory representations from temporally predictable external interference. *PsyArXiv*. <https://doi.org/10.31234/osf.io/wtesm>
- Liefoghe, B., De Houwer, J., & Wenke, D. (2013). Instruction-based response activation depends on task preparation. *Psychonomic Bulletin & Review, 20*(3), 481–487.
- Liefoghe, B., Wenke, D., & De Houwer, J. (2012). Instruction-based task-rule congruency effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(5), 1325–1335.
- Logan, G. D. (1988). Automaticity, resources, and memory: Theoretical controversies and practical implications. *Human Factors, 30*(5), 583–598.
- Longman, C. S., Liefoghe, B., & Verbruggen, F. (2019). How does the (re)presentation of instructions influence their implementation? *Journal of Cognition, 2*(1), 10. <https://doi.org/10.5334/joc.63>
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324. <https://doi.org/10.3758/s13428-011-0168-7>
- McElreath, R. (2018). *Statistical rethinking: A Bayesian course with examples in r and stan*. Chapman; Hall/CRC.
- Meiran, N., & Cohen-Kadosh, O. (2012). Working memory load but not multitasking eliminates the prepared reflex: Further evidence from the adapted flanker paradigm. *Acta Psychologica, 139*(2), 309–313.
- Meiran, N., Liefoghe, B., & De Houwer, J. (2017). Powerful instructions: Automaticity without practice. *Current Directions in Psychological Science, 26*(6), 509–514.
- Meiran, N., Pereg, M., Givon, E., Danieli, G., & Shahar, N. (2016). The role of working memory in rapid instructed task learning and intention-based reflexivity: An individual differences examination. *Neuropsychologia, 90*, 180–189. <https://doi.org/10.1016/j.neuropsychologia.2016.06.037>
- Meiran, N., Pereg, M., Kessler, Y., Cole, M. W., & Braver, T. S. (2015). The power of instructions: Proactive configuration of stimulus-response translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(3), 768–786.
- Monsell, S., & Graham, B. (2021). Role of verbal working memory in rapid procedural acquisition of a choice response task. *Cognition, 214*, 104731.
- Moors, A. (2016). Automaticity: Componential, causal, and mechanistic explanations. *Annual Review of Psychology, 67*, 263–287. <https://doi.org/10.1146/annurev-psych-122414-033550>

- Oberauer, K., Souza, A. S., Druey, M. D., & Gade, M. (2013). Analogous mechanisms of selection and updating in declarative and procedural working memory: Experiments and a computational model. *Cognitive Psychology*, *66*(2), 157–211.
- O'Reilly, R. C. (2010). The what and how of prefrontal cortical organization. *Trends in Neurosciences*, *33*(8), 355–361.
- Pereg, M., & Meiran, N. (2019). Rapid instructed task learning (but not automatic effects of instructions) is influenced by working memory load. *PLoS ONE*, *14*(6), e0217681.
- Psychology Software Tools. (2010). *E-Prime 2.0*. Psychology Software Tools.
- R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org/>. Accessed 14 Apr 2021.
- Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, *86*(3), 446–461.
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510.
- Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review*, *83*(3), 190.
- Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., et al. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, *1*(3), 337–356.
- Theeuwes, M., De Houwer, J., Eder, A., & Liefoghe, B. (2015). Congruency effects on the basis of instructed response-effect contingencies. *Acta Psychologica*, *158*, 43–50.
- Tibboel, H., Liefoghe, B., & De Houwer, J. (2016). Attention to future actions: The influence of instructed SR versus SS mappings on attentional control. *Psychological Research Psychologische Forschung*, *80*(6), 905–911.
- Torres-Quesada, M., Funes, M. J., & Lupiáñez, J. (2013). Dissociating proportion congruent and conflict adaptation effects in a Simon-Stroop procedure. *Acta Psychologica*, *142*(2), 203–210.
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505.
- Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure. *Behavior Research Methods*, *49*(2), 653–673. <https://doi.org/10.3758/s13428-016-0721-5r>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.