



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

Nachshon Meiran^{1*}, Michael W. Cole² and Todd S. Braver²

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

² Department of Psychology, Washington University, St. Louis, MO, USA

Edited by:

Nicola de Pisapia, University of Trento, Italy

Reviewed by:

Juliana Yordanova, Institute of Neurobiology, Bulgaria
Bernhard Hommel, Leiden University, Netherlands

*Correspondence:

Nachshon Meiran, Department of Psychology and Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel.
e-mail: nmeiran@bgu.ac.il

In this review, the authors discuss the seemingly paradoxical loss of control associated with states of high readiness to execute a plan, termed “intention-based reflexivity.” The review suggests that the neuro-cognitive systems involved in the preparation of novel plans are different than those involved in preparation of practiced plans (i.e., those that have been executed beforehand). When the plans are practiced, intention-based reflexivity depends on the prior availability of response codes in long-term memory (LTM). When the plans are novel, reflexivity is observed when the plan is pending and the goal has not yet been achieved. Intention-based reflexivity also depends on the availability of working-memory (WM) limited resources and the motivation to prepare. Reflexivity is probably related to the fact that, unlike reactive control (once a plan is prepared), proactive control tends to be relatively rigid.

Keywords: working-memory, intention, reflexivity, preparation, proactive control, prefrontal cortex

INTRODUCTION

Everyday experience suggests that planning may be useful. For example, when going on a trip abroad, it is usually recommended to book the flights, trains and hotels ahead of the trip. Nonetheless, planning is taxing and advance booking requires precious time and effort. Importantly in the present context, advance booking comes at the cost of reduced flexibility since one may not be able to change the hotel, when finding out upon arrival that a better and cheaper hotel is nearby. This conflict between costs and benefits is even more pronounced when planning takes place in parallel with other demanding activities. For example, toward the end of a talk in a conference, one may plan her or his question, yet this very attempt might prevent the person from truly listening to the end of the talk. These informal observations are supported by laboratory research on simple plans. This research suggests that planning (i.e., preparedness) improves action fluency and accuracy, resistance to interference (Braver et al., 2007; Braver, 2012) and the ability to overcome perseverative tendencies (Meiran and Daichman, 2005; Koch and Allport, 2006). At the same time, planning and holding plans in mind have been shown to consume limited processing resources, and thus interfere with other ongoing activity (e.g., Fagot, 1994; Meiran, 2000; Braver et al., 2003; Smith, 2003). Moreover, the fact that planning improves resistance to interference may actually prevent us from processing highly relevant information (e.g., Goschke and Dreisbach, 2008).

In the present work, we discuss an additional drawback of planning, the seemingly paradoxical loss of flexible online action control when the action plan is still pending. We describe this loss of control as “intention-based reflexivity,” or simply “reflexivity.” By “reflexivity” we mean that a cognitive process is triggered even when it is *not* required or intended *at the given moment* (see

also how Bargh and Gollwitzer, 1994; Tzelgov, 1997; Bargh et al., 2001, characterize “automaticity”). Although it may seem self contradictory to refer to the unintended aspects of intended acts, there is no contradiction here. This is because the unintended (possibly partial) execution of the plan may take place prematurely, i.e., before the intended execution, when the plan is still pending¹. Intention-based reflexivity can potentially have quite dramatic real life consequences such as when a policeman is aiming a gun in anticipation for an attack and accidentally shoots at a civilian who innocently passes by. In the lab, intention-based reflexivity can be studied with simple plans in which a given stimulus or a stimulus-category is linked to a particular planned response. An example for such a plan is to press the right key if the letter is from the beginning of the alphabet and press the left key if the letter is from the end of the alphabet.

NOVEL PLANS, PRACTICED PLANS, AND WORKING-MEMORY

Since plans must be stored and represented in memory, it is critical to consider the likely memory system that is involved. Based on considerations that are detailed below, we distinguish between novel plans and practiced plans. *Novel plans* are plans that have never been executed beforehand, such as the plan to write this paper or the plan to execute a reaction time task for the first time. *Practiced plans* are plans that have been executed beforehand, such as the plan to execute a familiar reaction time task, reach the office, or prepare an omelet. We argue that the kind of processes involved in representing and storing novel and practiced plans are different from one another in important respects. Consequently,

¹We use the term “reflexivity” instead of the term “automaticity,” because the latter term usually refers to practiced skill and to conditions in which the process is not a part of a pending plan.

the conditions that produce reflexivity in novel plans are quite different from the conditions that produce reflexivity in familiar plans.

The tenet of this paper is that plans are stored in working-memory (WM). However, it is widely appreciated that WM is not a unitary system and there are several different theories describing its subsystems (e.g., Miyake and Shah, 1999). We decided to adopt Oberauer's (2001, 2002, 2010) model [which may be viewed as an extension of Cowan's (1988), model] because of the natural link to reflexivity which this model affords. Following Oberauer, we consider WM as comprising: (1) novel bindings between familiar elements ("region of direct-access," RA) which is a severely limited resource; and (2) temporary heightened accessibility (activation) of familiar representations in long-term memory (LTM), termed "activated LTM" (ALTM), which is much less limited in its capacity (see **Figure 1**). Similar distinctions have been drawn in neuroscience by other theorists including Ruchkin et al. (2003); Postle (2006); Jonides et al. (2008); Bledowski et al., 2010, all suggesting that WM consists of an interaction between attentional systems subserved mostly by prefrontal cortex (PFC) regions and other brain regions involved in perception, semantic processing and action. The link to reflexivity which Oberauer's model affords is related to the fact that, according to Oberauer (2001), RA and ALTM differ not only in their capacity but also in their context sensitivity. While RA is highly sensitive to context, ALTM is not sensitive to context and operates even in inappropriate contexts, i.e., reflexively. To appreciate this link we describe here how Oberauer (2001) operationalized RA and ALTM. In the aforementioned work, Oberauer asked participants to memorize two lists of words and cued them in each trial which list is currently relevant. Following the cue and after a variable interval, a word was presented and the participants' task was to indicate if this probe was a member in the relevant list. Reaction times increased with increasing memory set size, indicating capacity limitations. Additionally, the size of the *irrelevant* list ceased to affect performance when sufficient time (1 s or more) elapsed between the cue and the probe, providing an opportunity to focus

on the relevant list. This result is interpreted as evidence that once the participants knew which one of the two lists is relevant, they focused on this list and removed the irrelevant list from RA. As soon as the irrelevant list was removed from RA it ceased to consume its resources.

The most important finding for the present focus is the *intrusion effect* which, according to Oberauer (2001) indexes ALTM. This effect shows that despite focusing on the relevant list and removing the irrelevant list from RA, participants had difficulty rejecting items that belonged to the irrelevant list. This result indicates that this irrelevant list, while being outside RA, remained highly accessible. In detail, the intrusion effect was defined as the difference in the time taken to reject items from the irrelevant list as compared with completely new items. The intrusion effect was relatively insensitive to set size and the time to focus on the relevant list. Importantly, according to our definition, the intrusion effect indicates the *reflexive* processing of words that belonged to the irrelevant list and thus serves to link ALTM and reflexivity.

THE REPRESENTATION OF PLANS IN WORKING-MEMORY

We now turn to use the distinction between RA and ALTM to describe our hypothesis regarding how plans are represented in WM. Before doing so, we further assume that, because RA resources are scarce (e.g., Cowan, 2001) (1) the use of this resource is avoided as much as possible, and (2) including purging its contents as soon as they become irrelevant or as soon as they can be represented in ALTM.

Based on these considerations, we argue that essential components of novel plans are represented within RA, at least when these plans are meant to be executed in the near future. In order to clarify what we mean by "essential components" let us consider for example the plan to hit the right key in response to a letter from the beginning of the alphabet (see **Figure 2**). This plan has two elements that can be represented in ALTM: the concept

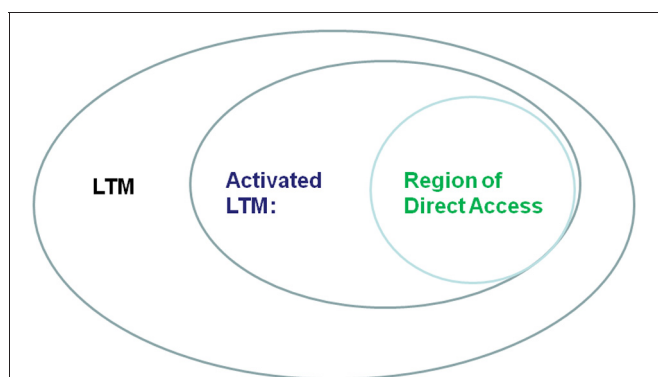


FIGURE 1 | A schematic representation of (our interpretation of) Oberauer's (2001, 2002) model. According to this model, working-memory comprises activated long-term memory representations. The region of direct-access consists of novel bindings between a sub-group of activated long-term memory representations.

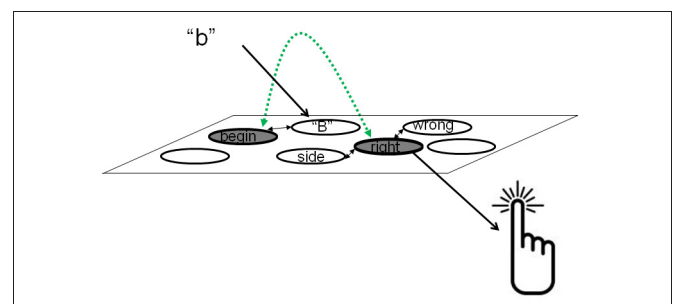


FIGURE 2 | An example for a representation of a simple rule such as IF (the stimulus is a letter from the beginning of the alphabet) THEN (press the right key). When the instructions are given, the long-term memory representations of "letter," "beginning of the alphabet" and "right" are activated. The novel binding between "beginning of the alphabet" and "right" (presented in green color) is in the region of direct-access. When a letter such as "b" is presented, it activates its corresponding representation in long-term memory. Via links in long-term memory, the representation of "beginning of the alphabet" is also further activated, leading to the activation of "right" (via the novel link) and to the key press (via established links in long-term memory).

“right key” and the concept “letter belonging to the beginning of the alphabet.” The reason is that these are familiar concepts, hence concepts that are stored in LTM. Based on our assumptions regarding sparing of RA resources, we assume that the representation of these elements *is* taking place within ALTM. However the link between them is novel and must thus be stored in RA. Had this plan been executed beforehand, all the elements of the plan would have been familiar (represented in LTM) and thus the entire plan could be in ALTM. This would be especially true for plans that have been executed at least several times in the past. Thus, while the entire plan can be represented in ALTM when the plan is practiced and familiar, only parts of it can be represented in this system when the plan is new and some elements of the plan must consume the limited RA resources.

REFLEXIVITY OF PRACTICED PLANS

Hommel (2000) provides an extensive review of the evidence for plan reflexivity, which at the time, was exclusively demonstrated using lengthy experiments. Because the experiments were lengthy and the plans remained the same throughout the experiment, one cannot rule out the possibility that the effects reflected practiced rather than novel plans. In the present review, we focus on evidence (1) linking reflexivity to intention and (2) showing that reflexivity of practiced plans depends on ALTM.

An important piece of evidence concerning reflexivity of practiced plans comes from the task-rule congruency effect found in task-switching experiments. This effect has been first demonstrated by Sudevan and Taylor (1987) who asked participants to switch between 2 numerical tasks performed on digits: size (larger/smaller than 5) and parity. Importantly (1) both tasks were executed on the same set of stimuli (the digits 1–9), and (2) the same right-left keys were used. Thus, there were trials in which the two task-rules were associated with the same response (compatible) or with competing responses (incompatible). Sudevan and Taylor found poorer performance in incompatible trials as compared to compatible trials (although this effect was restricted to the parity task). This result shows that, when the parity task was relevant, the currently irrelevant size rule operated reflexively and activated the response that would have been correct had this rule been relevant. Thus, the task-rule congruency effect provides evidence for the reflexivity of the currently irrelevant task and is quite analogous to Oberauer’s (2001) intrusion effect. Specifically, in both cases, information that is currently irrelevant but may become relevant in the near future influences performance reflexively. Since this demonstration, there were numerous additional papers that reported this effect (e.g., see Meiran and Kessler, 2008, for a partial review). In this section, we focus on two main issues. One is the evidence linking the task-rule congruency effect (and similar effects) to the intention to execute a given task. The other provides evidence for the involvement of ALTM.

LINKING THE REFLEXIVITY OF PRACTICED PLANS TO INTENTION

While the task-rule congruency effect shows that the irrelevant rule operated reflexively, its mere presence does not indicate that this reflexivity is related to the *intention* to execute this rule. An alternative explanation is that the initial execution of this rule

(even during the practice phase of the experiment) leads to the formation of LTM traces, which are known to generate reflexivity (e.g., Logan, 1988). An important piece of evidence linking task-rule congruency effect to intention is the fact that this effect greatly diminishes as soon as participants are told that the task would no longer be required. This result has been demonstrated by several authors including Fagot (1994); Meiran (2000, 2005); Yehene and Meiran (2007); Yamaguchi and Proctor (2011).

Another important demonstration comes from Marble and Proctor (2000, see also Proctor et al., 2000) who compared performance in the Simon task (Lu and Proctor, 1995, for review) in three conditions. In one condition (“pure Simon”) participants reacted to the color of stimuli by pressing right and left keys. They had to ignore the irrelevant location of the stimuli. In this condition, a usual Simon effect was found, showing quicker responses when the (irrelevant) location of the stimulus and the location of the responding hand were compatible rather than incompatible. In the critical conditions, the Simon task was intermixed with a location task, which required participants to respond to right-left locations of white stimuli. In one version of the paradigm, the location task was compatible (e.g., if the location is on the right, press the right key) and in the other version of the paradigm it was incompatible (if the location is on the right, press the left key). Marble and Proctor found that, relative to the pure Simon condition, the Simon effect increased (actually, roughly doubled) when the color task was intermixed with a compatible location task. We interpret this result as evidence that being prepared to execute the location task resulted in reflexive application of this plan even in the color task, when it was not required. Interestingly, the Simon effect (observed in the color task) was reversed when the location task was incompatible. This latter result has two important implications. One is that intention-based reflexivity may sometimes be more potent than automatic behaviors (indicated in the standard Simon effect). Specifically, when the Simon task was intermixed with the incompatible location task, the intention to execute the location task generated a tendency for reversed Simon effect while the automatic tendency was to generate a usual Simon effect. In this case, these two tendencies were opposite in direction and the fact that the intention-based tendency dominated suggests that it is more potent than the automatic tendency. The other implication of the reversed Simon effect is that mixing the color task and the location task did not only result in quicker processing of location information in general. If this were true, the Simon effect should have increased even when the location task was incompatible. The fact that the Simon effect had *reversed* indicates that the instructed link between locations and responses (rather than just location information) became reflexive. Memelink and Hommel (2006) ran a similar study but instead of using one location task, they embedded the Simon trials in blocks in which the other task involved switching between up-down and right-left classification of locations. They showed that the horizontal Simon effect increased when the Simon task immediately followed right-left judgments and that the vertical Simon effect increased when it immediately followed up-down judgments. These results further demonstrate the dependence of reflexivity on (prior) intention.

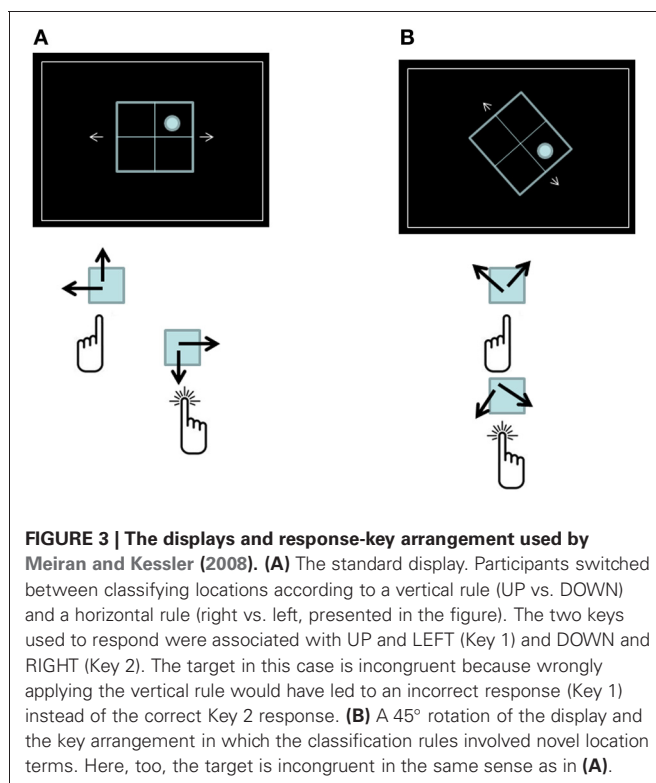
REFLEXIVITY OF PRACTICED PLANS DEPENDS ON ALTM

As mentioned beforehand, we argue that the mere representation of a plan in ALTM provides a sufficient condition for its reflexivity. Thus, the boundary conditions for reflexivity in this case are the same boundary conditions for ALTM representation, namely the prior existence of LTM codes. Meiran and Kessler (2008) noted that the task-rule congruency effect is analogous in many respects to Oberauer's (2001) intrusion effect in the sense that information that is currently irrelevant (but may become relevant soon) is processed nonetheless. To substantiate this analogy, they compared two task-switching paradigms (see **Figure 3**). In one paradigm, participants made up-down and right-left judgments on the location of a target stimulus within a 2×2 grid. The responses that were used were the upper-left key (used to indicate UP and LEFT) and the lower-right key (used to indicate DOWN and RIGHT). In this condition, the task-rules were presumably associated with the LTM-stored codes UP, DOWN, RIGHT, and LEFT. In the other version of the paradigm, the display was rotated by 45° as was the arrangement of the response keys. This rotation maintained the spatial relations of the stimuli, the responses, and the compatibility between responses and their corresponding locations. However, the regions within the 2×2 grid were no longer associated with the familiar (LTM-based) UP, DOWN, RIGHT, and LEFT regions. They probably were quite novel. The results showed that performance in this condition was slightly impaired relative to the standard, upright condition. The critical findings refer to the task-rule congruency effect. This effect was robust in the standard condition (~ 90 ms), showing that being prepared to execute one task-rule results in a reflexive

application of that rule when the other task-rule is required. Most importantly, there was no task-rule congruency effect whatsoever (0 ms) in the rotated condition. This result shows that the reflexivity of the plan depends on the prior availability of task-related response codes (such as UP and LEFT) in LTM and is absent when these response codes are novel. In a further experiment, the authors showed that the task-rule congruency effect in the rotated condition was re-established after one session of practice and remained stable in the remaining sessions. Arguably, practice resulted in the storage of the formerly novel response codes in LTM. Thus, from Session 2 onward, this aspect of the plan was represented in ALTM.

There are two additional pieces of evidence that reflexivity of practiced plans depends on their representation in ALTM. The first line of evidence is insensitivity to preparation. Specifically, Oberauer (2001) showed that the intrusion effect (that operationalized ALTM) was insensitive to the amount of time provided to focus on the relevant word list (and ignore the irrelevant list). Likewise, the task-rule congruency effect has repeatedly been shown to be insensitive to the amount of time allowed to focus on the relevant task and ignore the irrelevant task (e.g., Sudevan and Taylor, 1987; see also Fagot, 1994; Meiran, 1996, 2000, 2005; Meiran et al., 2000; but see Sudevan and Taylor, 1987, regarding preparation-based reduction in task-rule congruency effect after extensive practice). Similarly, as described above, Marble and Proctor (2000) showed that the Simon effect increased in the context of a compatible location task and reversed in the context of an incompatible location task. Importantly, they also showed that the effects remained even when participants received cues indicating which task is currently in effect. This last result shows that participants were unable to remove the plan to execute the location task from ALTM just as Oberauer's (2001) participants were unable to remove the irrelevant list from ALTM.

The last piece of evidence is the lack of sensitivity of the task-rule congruency effect to WM load, a manipulation which presumably exhausts RA limited resources. The rationale here is that if the exhaustion of RA resources does not influence the task-rule congruency effect then this effect must be based on other forms of representation. In detail, Kiesel et al. (2007) studied the task-rule congruency effect in an experiment involving 2 numerical tasks (magnitude and parity). The trials were executed in the retention interval of 2 vs. 5 letters (the WM load manipulation). These authors found that this load manipulation, while being effective (seen in generally poorer performance) did not influence the task-rule congruency effect. Kessler and Meiran (2010) raised the possibility that the load task used by Kiesel et al. might have been ineffective because it may have involved a different WM compartment as that used to store the task plans (Oberauer, 2010, see below). Thus, they used additional tasks as their load manipulation instead of memorizing items. In their experiments, the task-rule congruency effect was measured in the shape and color tasks, performed on colored shapes. The load tasks were 1 or 3 numeric tasks, performed on digits (Experiments 1 and 2) or size, fill and line thickness tasks, performed on completely different shapes. In none of the experiments was there a modulation of the task-rule congruency effect under heavier load, thus substantiating Kiesel et al.'s (2007) conclusions.



REFLEXIVITY OF NOVEL PLANS

According to prevalent theorizing (e.g., Schneider and Shiffrin, 1977; Logan, 1988), reflexivity/automaticity results from extensive and consistent practice. These theories do not postulate reflexivity for novel plans. Moreover, Oberauer's (2001) theory link reflexivity to ALTM but not to the RA. These considerations suggest that plans in RA might *not* be reflexive. However, there is a much older theory in psychology, *Prepared-Reflex* (PR) (Exner, 1879; Woodworth, 1938; see Logan, 1978; Hommel, 2000), suggesting that reflexivity of novel plans is not only possible, but may actually represent the typical scenario (see e.g., Folk et al., 1992, 1993, 1994; Remington and Folk, 2001; Pratt and Hommel, 2003, for related ideas concerning the direction of attention. See Bekkering and Neggers, 2002; Fagioli et al., 2007a,b; Wykowska et al., 2009, for evidence suggesting that reflexive attention orientation may actually represent reflexive action-intentions, as reviewed by Hommel, 2010). According to the PR theory, the representation of novel plans in RA in advance of task execution may be sufficient to cause these plans to become reflexive.

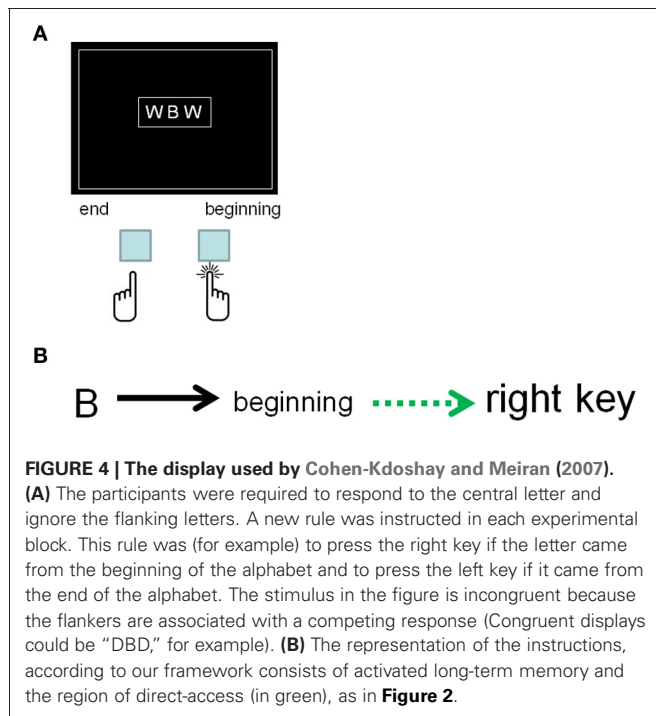
There have only been a few studies that have examined the reflexivity of truly novel plans. The few that did have provided evidence favoring the PR theory by showing that novel plans can be reflexive. These studies also suggest some important boundary conditions for novel plan reflexivity. We review them chronologically. As can be seen, the methodology improved gradually, allowing various alternative explanations to be more carefully ruled out.

de Houwer et al. (2005, Experiment 2) adopted an approach in which plan reflexivity was measured when the participants held the plan active in anticipation of its execution. Specifically, these authors instructed participants to be ready to react to the words "right" and "left" (as well as to equivalent arrows) with the utterances "bee" and "boo." They additionally asked the participants to use these same responses to indicate the color in which rectangles were presented. These rectangles appeared on the right or on the left despite of the fact that location was irrelevant. Importantly, during task performance there were actually no location words or arrows presented, meaning that the plan to react to location information was not practiced. Of interest was whether holding in mind the novel plan to react to locations with utterances would be reflexive. Reflexivity was measured in the color task by comparing compatible and incompatible trials, defined according to whether the reaction to (the irrelevant) location would have been the same as that for (the relevant) color. de Houwer et al. found a significant compatibility effect in their experiment (compatible RT < incompatible RT), indicating the reflexivity of the plan to react to locations. The major limitation of this study has to do with the fact that the plan might not have been truly novel and that reflexivity was built in the course of the experiment. This could have happened if the participants occasionally (and erroneously) reacted to the location of the colors instead of reacting to colors. A similar criticism (with respect to plan novelty) applies to studies that have examined subliminal response priming effects, which are not reviewed here (e.g., Eimer and Schlaghecken, 1998; Kunde et al., 2003). Showing a compatibility effect immediately (i.e., in the very first-trials)

after the task was instructed would have ruled out this possibility. Below we review two series of studies in which this prerequisite was met.

Wenke et al. (2007) asked participants to be prepared to execute a new speeded letter classification task in each trial (Task 2). This task involved an arbitrary mapping of two letters to the right-left key presses (e.g., N → left, K → right). Of interest is that, while being prepared to execute this letter classification task, the participants were given another task (Task 1) in which a pair of letters in different sizes was presented (e.g., "NK"). This task was introduced in order to assess plan reflexivity with responses indicated by pressing the spacebar once or twice. In Experiment 1, the participants were required to indicate the location of the larger letter in the pair while in Experiment 2 they indicated the color in which the larger letter was presented. There were compatible (e.g., "NK"), incompatible (e.g., "KN") and neutral (e.g., "FB") trials. The results indicated slightly quicker reactions to compatible trials than incompatible trials, and were thus interpreted as evidence that the instructions were sufficient to bind letter identity and response location (e.g., bind the letter "N" with the left side). However, these results do not provide evidence that the instructions for the letter classification task (Task 1) operated reflexively. This is partly because the responses in Task 2 (right vs. left key press) were different than the responses in Task 1 (single vs. double press of the space bar), meaning that Task 1 did not enable the full application of Task 2 instructions.

Cohen-Kdoshay and Meiran (2007) study overcomes the aforementioned shortcoming. These authors used a variant of the flanker paradigm (Eriksen and Eriksen, 1974) in which a centrally presented target was flanked by response-(in)compatible noise characters that were always physically different from the target. For example, when classifying target letters as belonging to the beginning/end of the alphabet, the stimuli "WBW" and "EBE" are incompatible and compatible, respectively. While the required response ("beginning of the alphabet") is the same in both of them because the target letter is "B," this target is flanked by letters that are either associated with the same response as the target ("E") or with an opposite response ("W") (see **Figure 4**). The novel aspect about Cohen-Kdoshay and Meiran's design was that a new stimulus set (and stimulus-response binding) was introduced in every experimental block. The key finding was a large (~60 ms) flanker compatibility effect (called first-trials flanker compatibility effect, or "first-trials compatibility effect," for short) that was found in the first block of eight trials immediately following the instructions. By introducing additional procedural changes, Cohen-Kdoshay and Meiran (2009) further showed that the first-trial compatibility effect was found in the very first-trial following the instructions, and that it was numerically larger than the compatibility effect in subsequent trials in the first mini-block of trials. This first-trial(s) compatibility effect indicates reflexivity because the processing of the flankers led to response activation despite the explicit requirement to ignore them. It also overcomes the issue of plan novelty, since the effect was measured immediately after the instructions. Additionally, the methodology is improved over the Wenke et al.'s (2007) study, because in that study the compatible condition constituted a physical representation of the instructions in the sense that the letter associated



with the right key appeared on the right and the letter associated with the left key appeared on the left. This aspect makes it possible that the quicker (slower) responses reflected a (mis)match between the display and the instructions. No such criticism could apply to Cohen-Kdoshay and Meiran’s experiments. Thus, there is greater confidence that the representation of the instructions in RA has reflexively generated a response tendency (or at least, the response identity, e.g., see Hommel, 1998) for the flankers even before those flankers had ever been practiced as a target response. Nonetheless, there still remains an alternative explanation according to which the flankers primed target identity processing (rather than have caused response generation). This alternative account holds because the flankers were semantically associated with the target (e.g., both belonged to the beginning of the alphabet). We do not think this alternative account is likely to be correct because the flanker compatibility effects quickly diminished in the course of the block, a trend that is not expected for semantic priming.

Finally, there is one study that tried but failed to show evidence for novel plan reflexivity. The findings of this study point to important boundary conditions for novel plan reflexivity as discussed below. Waszak et al. (2008) employed a task-switching design in which participants switched between color classification and shape classification. Importantly, some of the (irrelevant) colors used in the shape task and some of the (irrelevant) shapes that were used in the color task were only instructed and were never presented as targets. These authors found a reliable task-rule congruency effect. However, the task-rule congruency effect was only found for the irrelevant features that also were presented as targets (in other trials), but not for the irrelevant features that were merely instructed (and never responded to as targets). Thus, holding the plan to react to a given color/shape in

a certain manner was insufficient for the reflexive application of this plan. The potential reasons for this null finding are discussed now.

BOUNDARY CONDITIONS

The extant literature suggests three boundary conditions for the reflexivity of novel plans. The first is the availability of RA storage resources. The second boundary condition is that the planned task and the task in which reflexivity is measured must be considered to be part of the same processing event. The final condition is the motivation to maintain high preparedness to execute the planned task. We will discuss each one of these factors in turn.

Availability of RA storage resources

The fact that RA storage resources must be available has been demonstrated by Cohen-Kdoshay and Meiran (2007, Experiment 4) who showed that the first-trial compatibility effect was eliminated when RA was loaded by secondary task instructions. Specifically, in addition to instructing participants to carry out a new classification task, each block also included a novel go-nogo to be performed on rare occasions. This task involved clearly distinguishable target stimuli (numbers or number words) and required a “go” response (pressing the spacebar with both thumbs) if the stimulus met a certain criterion (such as being divisible by 3). In these conditions, the first-trial compatibility effect was eliminated, suggesting that only when the novel plan is held in its entirety in RA, reflexivity is found. We have recently addressed an alternative account to this finding according to which the load task did not exhaust storage space, but instead merely introduced increased multitasking demands. In order to rule out this account, Meiran and Cohen-Kdoshay (2012) compared three groups of participants. In one group, there was no additional load. This group showed the usual first-trial compatibility effect. In another group, there was a load task that was changed between blocks, as in Cohen-Kdoshay and Meiran (2007) fourth experiment. The first-trial compatibility effect was eliminated (actually, slightly reversed) in this group, thus replicating our previous work. Of interest is the third group in which the go-nogo task remained the same throughout the experiment. In this group, the load task increased multitasking demands but did not load RA (because the load task could have been represented in ALTM). This group showed a first-trial compatibility effect that was equivalent to that found in the group without load. These results show that what is critical is the representation of the plan in RA.

Ellenbogen and Meiran (2008) studied the involvement of RA in the backward compatibility effect (Hommel, 1998). In their experiments, participants made speeded responses to the letters H and S presented in different colors. The primary task was color classification which was made using right and left key presses. The secondary task that was executed immediately afterwards was letter classification in which the (arbitrarily mapped) responses were the (Hebrew equivalents) of the utterances RED and GREEN. The logic was that executing Task 1 while being prepared to execute Task 2 would lead to compatibility effects such that if the color of the stimulus is the same as the utterance made to letter identity,

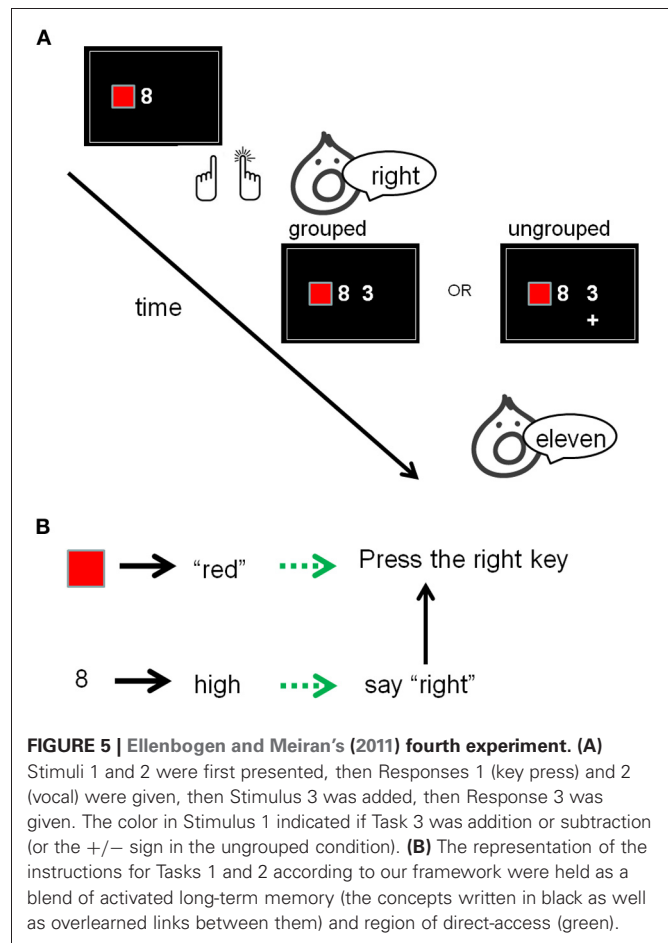
reaction would be relatively quick. This compatibility effect was termed “backward compatibility” by Hommel. Of interest is the load manipulation used in this study, which was the number of possible colors (and thus, the number of color-key associations). When the load was low (two colors) or moderate (four colors) there was a robust backward compatibility effect (thus completely replicating Hommel and Eglau, 2002). However, when the load was extreme (six colors) the backward compatibility effect was eliminated. In a further experiment, the authors showed that the critical factor was not the number of colors, but the number of color-category—to—response rules. The fact that extreme load was needed to eliminate reflexivity already provides evidence that a few dozens of executions are sufficient to reduce the dependence on RA such that extreme load is needed in order to remove plan reflexivity.

Inclusion in an event

We suggest the principle that if a person has a plan in mind to execute a task, this forms an event which lasts from the point of planning to the point of plan execution. We will refer to it as the plan-to-execution event. We further suggest that the measurement of reflexivity of a plan must take place during the plan-to-execution event.

Although the term “event” is rather elusive, there is marked agreement between observers regarding event boundaries, and considerable objective support for the psychological reality of these boundaries (e.g., Zacks et al., 2007). Our notion regarding plan-to-execution events is supported by findings from this literature including those indicating that a goal change demarks event boundaries (Zacks and Swallow, 2007) and the evidence suggesting that the contents of RA are refreshed (i.e., updated) at these boundaries (e.g., Swallow et al., 2009) presumably in order to free these scarce resources as soon as possible.

Ellenbogen and Meiran (2011) showed evidence that plan reflexivity takes place only during the plan-to-execution event. These authors used the backward compatibility effect (described above) as an index of reflexivity. In this paradigm, the relevant plan-to-execution event (of Task 2) presumably starts before Task 1 is executed, when participants are ready to execute both tasks in close succession. However, subtle environmental cues may be sufficient to start planning for Task 2 only after Task 1 has already been performed. We specifically refer to cues indicating that a goal has been achieved. In one of Ellenbogen and Meiran’s experiments (see **Figure 5**), participants were presented with a colored square and a digit. They then reacted to the color by a right/left key press and to the magnitude of the digit by saying “right” or “left.” After responding to both of these stimuli (Tasks 1 and 2, respectively), a second digit was presented and Task 3 was performed. Like in Ellenbogen and Meiran (2008) study, described above, Task 1 was used to assess the reflexivity of the plan to execute Task 2 because Task 2 responses (saying “right” or “left”) were either compatible or incompatible with the key press made in Task 1. Task 3 was added in order to manipulate the duration of the plan-to-execution event. In the experimental group (“grouped”), the color of the square indicated whether Task 3 would involve adding the two digits or subtracting the third digit from the second digit. Under these conditions, color



processing (Task 1) belonged to the same plan-to-execution event as Task 2 because this event ended only when Task 3 has been executed. In the control group, the colored square did not serve as a task cue for the two digits and thus Task 1 belonged to one event while Task 2 and Task 3 belonged to different event. (In this group, the cue for Task 3 was a minus or a plus sign that appeared below the second digit.) Because the second digit (the stimulus for Task 3) was presented only after the response to Task 2 was made, the display and the tasks were identical for the two groups until Task 2 ended, which is when backward compatibility effects were assessed in order to measure plan reflexivity. The results indicated a backward compatibility effect (in Task 1, as usual) but only in the experimental group and not in the control group. This result shows that the plan to execute Task 2 (whose reflexivity was assessed) was retrieved into RA only when Task 1 (in which reflexivity was assessed) was a part of the same plan-to-execute event as Task 2.

Motivation

The motivation to hold the plan in mind seems important as well because when motivation is lacking, the plan may not be represented in RA, for example. To our knowledge, the only motivation-relevant findings are by Wenke et al. (2009) using the same paradigm as Wenke et al. (2007), described above.

Wenke et al. (2009) compared conditions that are associated with differential motivation levels to prepare an action in advance. For example, one such condition involved frequent omission of the requirement to execute the plan. Specifically, their paradigm involved preparing for a choice reaction time task (Task 2) and executing another task (Task 1) that was used to assess plan reflexivity. When Task 2 was frequently omitted, this presumably lowered participants' motivation to prepare toward this task. Accordingly, there was no evidence that the plan to execute Task 2 was reflexive in these conditions. These results suggest that being motivated to maintain high readiness to execute a task is a precondition for reflexivity. Nevertheless, this boundary condition awaits further testing in which motivation levels are manipulated directly (e.g., with incentives).

Individual differences

An additional potential precondition is individual differences in RA capacity, a factor that has not yet been examined. Suggestive evidence comes from Wilhelm and Oberauer (2006) who, in an individual differences study, found a correlation between WM capacity and reaction time. Of interest is the key finding in this paper showing that the aforementioned correlation was significantly higher when the reaction time tasks involved an arbitrary stimulus-response mapping (which presumably would be held in RA, at least until sufficient practice had accrued) as compared with a non-arbitrary (and compatible) mapping in which response identity can be directly retrieved from LTM. Nonetheless, the authors did not examine performance on the tasks immediately following the instructions and therefore it is difficult to tell with certainty if these findings relate to novel task representation.

Relevance of the boundary conditions to the null result of Waszak et al. (2008)

After listing the boundary conditions for plan reflexivity, it becomes clear why there was no evidence for reflexivity in Waszak et al.'s (2008) study. Specifically, in their study, there might not have been sufficient motivation to maintain readiness to react to stimulus values that never served as targets precisely because they never served as targets, a fact that could probably have been detected already at the beginning of the experiment, after a few dozen trials. Additionally, the limited RA storage capacity might have been exhausted because of rather high cognitive demand, involving many stimulus-response pairs and two tasks. Finally, the measurement of reflexivity was performed outside the plan-to-execution event. Specifically, when the task-rule congruency effect was assessed in the color task (for example), it reflected the readiness to execute the shape task, but the shape task was executed in other trials and thus belonged to different events.

PROCEDURAL VS. DECLARATIVE WM

While we adopted Oberauer's (2001, 2002) model, more recently, Oberauer (2010) further suggested that WM has two distinct compartments with analogous structure, procedural WM and declarative WM. Our description of WM subcomponents accords well with Oberauer's (2010) characterization of

declarative WM. However, Oberauer's (2010) characterization of procedural WM indicates that PR behaviors are related to the procedural compartment and not the declarative compartment. We think that given the current knowledge, it is difficult to decide whether a separate procedural WM exists, although the available evidence seems to favor a domain-general view rather than a distinction between two WM systems. One line of evidence concerns individual differences in WM and reaction times (assuming that reaction times are related to procedural WM). Specifically, Schmiedek et al. (2007) tested participants with various declarative WM measures and also tested them on choice reaction time tasks, all involving arbitrary stimulus-response mapping. Using a structural equations modeling approach which permits estimation of correlations between latent variables (as opposed to observed measures), these authors found a strong correlation ($r = 0.90$) between an aspect of the reaction time distribution and WM. Wilhelm and Oberauer (2006) results, described beforehand, lead to a similar conclusion. Along a similar line, Hartstra et al. (2011) who examined the brain areas involved in the representation of novel instructions found that the very same brain region (left lateral PFC) subserved novel task instructions and the storage of novel object-color associations. These results support a domain-general view rather than a separation between procedural and declarative WM.

POSSIBLE NEURO-COGNITIVE MECHANISMS

A theoretical framework that can potentially explain the difference between novel and practiced plans is the dual-mechanisms of control account developed by Braver et al. (2007, 2009; Braver, 2012). This framework distinguishes between two modes of cognitive control: proactive and reactive. The *proactive control* mode involves preparatory activation and sustained maintenance of goal-related information triggered by advance contextual cues (i.e., task instructions). It contrasts with *reactive control*, involving transient, stimulus-triggered retrieval of goal-related information that is based on conflict detection or bottom-up associative links. Importantly, prior work has demonstrated variability or flexibility in these control modes, linked to (sometimes subtle) features of the task or stimuli, as well as stable individual differences (Braver et al., 2007; Braver, 2012). This variability has been detected in terms of brain activity dynamics, using fMRI methods. Thus, under task conditions preferentially associated with proactive control, anticipatory and sustained intention-related activity has been observed in the lateral PFC. In contrast, under conditions involving reactive control, activity was transient, stimulus-triggered and involved not only the lateral PFC, but also brain regions associated with conflict detection (i.e., anterior cingulate cortex) and episodic/associative retrieval (i.e., lateral parietal cortex, anterior PFC, and medial temporal lobes). Additionally, other work has suggested that proactive control is preferentially engaged for the preparation of action-intentions, i.e. plans (vs. preparatory attention; Ruge et al., 2009, 2010). The dual-mechanisms of control framework, therefore, suggest that the critical difference between proactive and reactive control is not just which brain regions are engaged, but moreover the temporal dynamics of activation. Here, we further suggest that plan novelty might also be a potential key feature that promotes

proactive control, whereas practiced plans might rely more on reactive control. One reason why this may be so is that performing a novel task requires sustained binding of the task elements (see below) and thus RA. However, dealing with familiar tasks may be performed “on the fly” by retrieving the already stored binding of task elements. This view predicts that novel plans would be associated with increased anticipatory (i.e., prior to plan implementation) activation within lateral PFC, whereas practiced plans should be associated with greater transient post-stimulus activation (in these and other regions). This would be in line with the idea that novel plans are in a stronger PR type state than practiced plans.

Only very few studies have compared novel and practiced tasks directly. Cole et al. (2010) developed a paradigm that enabled comparison of novel and practiced tasks by permuting a set of rules into many novel rule combinations (i.e., tasks), creating a large space of possible tasks (64). The rules used fall into three categories: semantic (e.g., “is it green?”), decision (e.g., “is the answer to both words the same?”), and response (e.g., “if true, press your left index finger”). Thus, one of the 64 tasks could be, “If the answer to “is it green?” is the same for both words then press your left index finger.” By utilizing 64 different tasks, it is possible to obtain reliable estimates of novel task learning, by examining first-trial performance on each of the tasks. Further, a subset of the 64 tasks is extensively practiced, such that novel and practiced tasks can be directly compared while controlling for the particular stimuli and rules used. Cole et al. found (using functional MRI; fMRI) that practiced task preparation activated the anterior PFC first, then the dorsolateral PFC. In contrast, they found a reversal of this pattern for novel tasks, suggesting that practiced tasks involve a top-down task set retrieval (from LTM) process, while novel tasks involve a bottom-up task set formation process. Thus, it was the pattern of relationship between the lateral PFC and other brain regions (anterior PFC, in this case) that distinguished between practiced and novel tasks, somewhat in line with Braver et al.’s distinction between proactive and reactive control. Using a variety of tasks with unique visual stimuli, Ruge and Wolfensteller (2010) also observed that PFC (among other regions) was involved in novel task learning.

Continuing this line of investigation, Cole et al. (2011) used multivariate pattern analysis of fMRI data to identify the presence of specific rules within PFC. They found that PFC activity patterns associated with specific rules during practice were also present during novel task preparation and execution. This suggests that the PFC gains much of its flexibility by *rigidly* applying old but highly practiced activity patterns to novel contexts. The present review suggests that rigidity may be a general feature of highly prepared states rather than being restricted to the application of familiar rules.

In a related study, Gilbert et al. (2012) used a similar analysis approach to examine the content of delayed intentions. They observed a sustained, intention-related increase in anterior PFC activity, and increased connectivity of this region with posterior cortex regions that showed content-specific representation. This suggests a potentially interactive model, in which sustained PFC input may be needed to shift posterior cortex representations

from the ALTM to RA component of WM or alternatively, to bind the novel pattern of LTM representations in posterior cortex, in accordance with how we defined RA.

Our notion concerning reflexivity of novel plans may seem counterintuitive. Nonetheless, it accords with current theorizing concerning the nature of PFC representations. Specifically, these representations are described as being highly flexible in the sense of representing any novel combination of familiar elements (e.g., O’Reilly et al., 1999; Miller and Cohen, 2001). However, PFC representations are also believed to be highly resistant to interference. What we suggest is that this resistance to interference comes at the cost of losing online flexibility. In this regard, incorporating rapid contextual changes into the existing RA representation may require the formation of an entirely new representation (cf. Kessler and Meiran, 2006, 2008) This new activation-based RA representation is likely fragile (i.e., incoming activity can destroy it), and so it is “locked in” to reduce interference, making us somewhat less responsive to immediate contextual changes. Consistent with this idea, Duncan (2010) suggested that encoding in PFC (and other brain regions which form the Multiple Demand network) is based on the *pattern* of neuronal activation. He further argues that when the represented content changes, so does the *pattern* of neuronal activation. Similarly, O’Reilly et al. (2003) suggested a coarse and distributed coding system. Since the neurons in this coding scheme encode conjunctions of elements, a change in the context seems to again require a change in activation over many neurons. Singer and Gray (1995) and others suggest that novel representations consist of synchronous neuronal firing, which, for WM representations is presumably supported by the PFC (e.g., Ruchkin et al., 2003). Updating such a representation seems to require generating a new pattern of synchronous activity. Braver and Cohen (2000) suggest that PFC representations are updated only on specific occasions, determined by the dopamine-controlled gating system, again implying that there are periods in which representations cannot be updated. Regardless of the differences, all of the theories suggest that RA-based representations, while being extremely flexible in the sense that they can be entirely novel combinations of elements, are inflexible in their responsiveness to rapid contextual changes in the environment, which according to the present thesis is what characterizes intention-based reflexivity.

CONCLUSIONS

In this brief review, we have presented evidence that the intention to carry out a simple plan in the near future may result in paradoxical loss of control, such that the intended plan may be (at least partly) executed prematurely and inappropriately. We distinguished between two types of plans based on the WM compartment that is probably used for their storage. Planning to execute a familiar task may be entirely based on ALTM, thus preserving the scarce RA capacity resources. However, when the plan is novel, aspects of it are probably represented in RA. We further showed that different boundary conditions apply to the reflexivity of novel and practiced plans and suggested the likely neuro-cognitive mechanisms that are being involved. We thus conclude that plan reflexivity provides clues as to the

mechanism underlying the mind's tremendous flexibility during preparation. It seems that whatever mechanism gives us this gift of mental flexibility to allow for rapid novel planning also takes away flexibility as a new plan is prepared to be executed.

The hypotheses that we outlined in this paper lead to many future research directions that stem from currently unresolved questions, such as: is plan reflexivity a (perhaps unwanted) side-effect of planning or is it (also) associated with benefits? In this regard, Gollwitzer (1999) suggested that reflexive plans are more likely to get successfully executed, in part because of reduced dependence on endogenous control inputs. (This hypothesis is still hotly debated, e.g., see Brandstätter et al., 2001; Smith, 2003 vs. e.g., McDaniel and Scullin, 2010.) Additionally, despite the

relatively clear evidence that novel and practiced tasks are subserved by different patterns of brain activity, it is unclear at present whether these differential patterns are related to plan reflexivity. Finally, while we suggested that plan reflexivity is a feature of proactive control, this is merely a speculation at this point and further research is needed to provide direct support for it.

ACKNOWLEDGMENTS

This research was supported by a research grant from the Israel Science Foundation to the first author, and NIH grants R01 MH66078 and MH66078-06A1S1 to the second and third authors. We wish to thank Bernhard Hommel and Klaus Oberauer for helpful comments.

REFERENCES

- Bargh, J. A., and Gollwitzer, P. M. (1994). Environmental control of goal directed action: automatic and strategic contingencies between situations and behavior. *Nebr. Symp. Motiv.* 41, 71–124.
- Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K., and Trötschel, R. (2001). The automated will: non-conscious activation and pursuit of behavioral goals. *J. Pers. Soc. Psychol.* 81, 1014–1027.
- Bekker, H., and Neggers, S. F. W. (2002). Visual search is modulated by action intentions. *Psychol. Sci.* 13, 370–374.
- Bledowski, C., Kaiser, J., and Rahm, B. (2010). Basic operations in working memory: contributions from functional imaging studies. *Behav. Brain Res.* 214, 172–179.
- Brandstätter, V., Lengfelder, A., and Gollwitzer, P. M. (2001). Implementation intentions and efficient action initiation. *J. Pers. Soc. Psychol.* 81, 946–960.
- Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends Cogn. Sci.* 16, 106–113.
- Braver, T. S., and Cohen, J. D. (2000). “On the control of control: the role of dopamine in regulating prefrontal function and working memory,” in *Attention and Performance XVIII* eds S. Monsell and J. Driver (Cambridge, MA: MIT Press), 713–737.
- Braver, T. S., Gray, J. R., and Burgess, G. C. (2007). “Explaining the many varieties of working memory variation: dual mechanisms of cognitive control (Chapter 4),” in *Variation in Working Memory*, eds A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake and J. N. Towse (New York, NY: Oxford University Press), 76–106.
- Braver, T. S., Reynolds, J. R., and Donaldson, D. I. (2003). Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron* 39, 713–726.
- Braver, T. S., Paxton, J. L., Locke, H. S., and Barch, D. M. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proc. Natl. Acad. Sci.* 106, 7351–7356.
- Cohen-Kadosh, O., and Meiran, N. (2007). The representation of instructions in working memory leads to autonomous response activation: evidence from the first trials in the flanker paradigm. *Q. J. Exp. Psychol.* 60, 1140–1154.
- Cohen-Kadosh, O., and Meiran, N. (2009). The representation of instructions operates like a prepared reflex: flanker compatibility effects found in first trial following S-R instructions. *Exp. Psychol.* 56, 128–133.
- Cole, M. W., Bagic, A., Kass, R., and Schneider, W. (2010). Prefrontal dynamics underlying rapid instructed task learning reverse with practice. *J. Neurosci.* 30, 14245–14241.
- Cole, M. W., Etzel, J. A., Zacks, J. M., Schneider, W., and Braver, T. S. (2011). Rapid transfer of abstract rules to novel contexts in human lateral prefrontal cortex. *Front. Hum. Neurosci.* doi: 10.3389/fnhum.2011.00142
- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychol. Bull.* 104, 163–191.
- Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci.* 24, 1, 87–114.
- de Houwer, J., Beckers, T., Vandorp, S., and Custers, R. (2005). Further evidence for the role of mode-independent short-term associations in spatial Simon effects. *Percept. Psychophys.* 67, 659–666.
- Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn. Sci.* 14, 172–179.
- Eimer, M., and Schlaghecken, F. (1998). Effect of masked stimuli on motor activation: behavioral and electrophysiological evidence. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 1737–1747.
- Ellenbogen, R., and Meiran, N. (2008). Working memory involvement in dual task performance: evidence from the backward compatibility effect. *Mem. Cogn.* 36, 968–978.
- Ellenbogen, R., and Meiran, N. (2011). Objects and events as determinants of parallel processing in dual tasks: evidence from the backward compatibility effect. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 152–167.
- Eriksen, B., and Eriksen, C. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept. Psychophys.* 16, 143–149.
- Exner, S. (1879). “Physiologie der Grosshirnrinde,” in *Handbuch Der Physiologie*, Vol. 2, ed L. Hermann (Leipzig, Germany: Vogel), 189–350.
- Fagioli, S., Ferlazzo, F., and Hommel, B. (2007a). Controlling attention through action: observing actions primes action-related stimulus dimensions. *Neuropsychologia* 45, 3351–3355.
- Fagioli, S., Hommel, B., and Schubotz, R. I. (2007b). Intentional control of attention: action planning primes action-related stimulus dimensions. *Psychol. Res.* 71, 22–29.
- Fagot, C. (1994). *Chronometric Investigations of Task Switching*. Ph.D. thesis, University of California, CA, San Diego.
- Folk, C. L., Remington, R. W., and Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *J. Exp. Psychol. Hum. Percept. Perform.* 18, 1030–1044.
- Folk, C. L., Remington, R. W., and Johnston, J. C. (1993). Contingent attentional capture: a reply to Yantis (1993). *J. Exp. Psychol. Hum. Percept. Perform.* 19, 682–685.
- Folk, C. L., Remington, R. W., and Wright, J. H. (1994). The structure of attentional control: contingent attentional capture by apparent motion, abrupt onset, and color. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 317–329.
- Gilbert, S. J., Armbruster, D. J. N., and Panagiotidi, M. (2012). Similarity between brain activity at encoding and retrieval predicts successful realization of delayed intentions. *J. Cogn. Neurosci.* 24, 93–105.
- Gollwitzer, P. M. (1999). Implementation intentions: strong effects of simple plans. *Am. Psychol.* 54, 493–503.
- Goschke, T., and Dreisbach, G. (2008). Conflict-triggered goal shielding: response conflicts attenuate background-monitoring for concurrent prospective memory cues. *Psychol. Sci.* 19, 25–32.
- Hartstra, E., Kuhn, S., Verguts, T., and Brass, M. (2011). The implementation of verbal instructions: an fMRI study. *Hum. Brain Mapp.* 32, 1811–1824.
- Hommel, B. (1998). Automatic stimulus-response translation in dual-task performance. *J. Exp.*

- Psychol. Hum. Percept. Perform.* 24, 1368–1384.
- Hommel, B. (2000). “The prepared reflex: automaticity and control in stimulus response translation,” in *Attention and Performance 18: Control of Cognitive Processes*, eds S. Monsell and J. Driver (Cambridge, MA: MIT Press), 247–273.
- Hommel, B. (2010). “Grounding attention in action control: the intentional control of selection,” in *Effortless Attention: A New Perspective in the Cognitive Science of Attention and Action*, ed B. J. Bruya (Cambridge, MA: MIT Press), 121–140.
- Hommel, B., and Eglau, B. (2002). Control of stimulus-response translation in dual-task performance. *Psychol. Res.* 66, 260–273.
- Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., and Moore, K. S. (2008). The mind and brain of short-term memory. *Annu. Rev. Psychol.* 59, 193–224.
- Kessler, Y., and Meiran, N. (2006). All updateable objects in working memory are updated whenever any of them is modified: evidence from the memory updating paradigm. *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 570–585.
- Kessler, Y., and Meiran, N. (2008). Two dissociable updating processes in working memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 1339–1348.
- Kessler, Y., and Meiran, N. (2010). The reaction-time task-rule congruency effect is not affected by working memory load: further support for the activated long-term memory hypothesis. *Psychol. Res.* 74, 388–399.
- Kiesel, A., Wendt, M., and Peters, A. (2007). Task switching: on the origins of response congruency effects. *Psychol. Res.* 71, 117–125.
- Koch, I., and Allport, A. (2006). Cue-based preparation and stimulus-based priming of tasks in task switching. *Mem. Cogn.* 34, 433–444.
- Kunde, W., Kiesel, A., and Hoffmann, J. (2003). Conscious control over the content of unconscious cognition. *Cognition* 88, 223–242.
- Logan, G. D. (1978). Attention in character classification: evidence for the automaticity of component stages. *J. Exp. Psychol. Gen.* 107, 32–63.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychol. Rev.* 95, 492–527.
- Lu, C.-H., and Proctor, R. W. (1995). The influence of irrelevant location information on performance: a review of the Simon and spatial Stroop effects. *Psychon. Bull. Rev.* 2, 174–207.
- Marble, J. G., and Proctor, R. W. (2000). Mixing location-relevant and location-irrelevant trials in choice-reaction tasks: influences of location mapping on the Simon effect. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 1515–1533.
- McDaniel, M. A., and Scullin, M. K. (2010). Implementation intention encoding does not automatize prospective memory responding. *Mem. Cogn.* 38, 221–232.
- Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 1423–1442.
- Meiran, N. (2000). Modeling cognitive control in task-switching. *Psychol. Res.* 63, 234–249.
- Meiran, N. (2005). Task rule congruency and Simon-like effects in switching between spatial tasks. *Q. J. Exp. Psychol. A* 58, 1023–1041.
- Meiran, N., Chorev, Z., and Sapir, A. (2000). Component processes in task switching. *Cogn. Psychol.* 41, 211–253.
- Meiran, N., and Cohen-Kdoshay, O. (2012). Working memory load but not multitasking eliminates the prepared reflex: further evidence from the adapted flanker paradigm. *Acta Psychol.* 139, 309–313.
- Meiran, N., and Daichman, A. (2005). Advance task preparation reduces task error rate in the cueing task-switching paradigm. *Mem. Cogn.* 33, 1272–1288.
- Meiran, N., and Kessler, Y. (2008). The task rule congruency effect in task switching reflects activated long term memory. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 137–157.
- Memelink, J., and Hommel, B. (2006). Tailoring perception and action to the task at hand. *Eur. J. Cogn. Psychol.* 18, 579–592.
- Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202.
- Miyake, A., and Shah, P. (Eds.) (1999). *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. New York, NY: Cambridge University Press.
- Oberauer, K. (2001). Removing irrelevant information from working memory: a cognitive aging study with the modified Sternberg task. *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 948–957.
- Oberauer, K. (2002). Access to information in working memory: exploring the focus of attention. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 411–421.
- Oberauer, (2010). Design for a working memory. *Psychol. Learn. Motiv.* 51, 45–100.
- O’Reilly, R. C., Braver, T. S., and Cohen, J. D. (1999). “A biologically-based computational model of working memory,” in *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, eds A. Miyake and P. Shah (Cambridge, MA: Cambridge University Press), 102–134.
- O’Reilly, R. C., Busby, R. S. and Soto, R. (2003). “Three forms of binding and their neural substrates: alternatives to temporal synchrony,” in *The Unity of Consciousness: Binding, Integration, and Dissociation*, ed A. Cleeremans (Oxford: Oxford University Press), 168–192.
- Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience* 139, 23–28.
- Pratt, J., and Hommel, B. (2003). Symbolic control of visual attention: the role of working memory and attentional control settings. *J. Exp. Psychol. Hum. Percept. Perform.* 5, 835–845.
- Proctor, R. W., Marble, J., and Vu, K.-P. (2000). Mixing incompatibly mapped location-relevant trials with location-irrelevant trials: effects of stimulus mode on performance. *Psychol. Res.* 64, 11–24.
- Remington, R. W., and Folk, C. L. (2001). Dissociation between attention and selection. *Psychol. Sci.* 12, 511–515.
- Ruchkin, D. S., Grafman, J., Cameron, K., and Berndt, R. S. (2003). Working memory retention systems: a state of activated long-term memory. *Behav. Brain Sci.* 26, 709–728.
- Ruge, H., Braver, T. S., and Meiran, N. (2009). Attention, intention, and strategy in preparatory control. *Neuropsychologia* 47, 1670–1685.
- Ruge, H., Müller, S. C., and Braver, T. S. (2010). Anticipating the consequences of action: an fMRI study of intention-based task preparation. *Psychophysiology* 47, 1019–1027.
- Ruge, H., and Wolfensteller, U. (2010). Rapid formation of pragmatic rule representations in the human brain during instruction-based learning. *Cereb. Cortex* 20, 1656–1667.
- Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H. M., and Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *J. Exp. Psychol. Gen.* 136, 414–429.
- Schneider, W., and Shiffrin, R. M. (1977). Controlled and automatic human information processing: 1. Detection, search, and attention. *Psychol. Rev.* 84, 1–66.
- Singer, W., and Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.* 18, 555–586.
- Smith, R. E. (2003). The cost of remembering to remember in event-based prospective memory: investigating the capacity demands of delayed intention performance. *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 347–361.
- Sudevan, P., and Taylor, D. A. (1987). The cueing and priming of cognitive operations. *J. Exp. Psychol. Hum. Percept. Perform.* 13, 89–103.
- Swallow, K. M., Zacks, J. M., and Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *J. Exp. Psychol. Gen.* 138, 236–257.
- Tzelgov, J. (1997). “Automatic but conscious: that is how we act most of the time,” in *Advances in Social Cognition*, Vol. X, eds R. Wyer (Mahwah, NJ: Erlbaum), 217–230.
- Waszak, F., Wenke, D., and Brass, M. (2008). Cross-talk of instructed and applied arbitrary visuomotor mappings. *Acta Psychol.* 127, 30–35.
- Wenke, D., Gaschler, R., and Nattkemper, D. (2007). Instruction-induced feature binding. *Psychol. Res.* 71, 92–106.
- Wenke, D., Gaschler, R., Nattkemper, D., and Frensch, P. A. (2009). Strategic influences on implementing instructions for future actions. *Psychol. Res.* 73, 587–601.
- Wilhelm, O., and Oberauer, K. (2006). Why are reasoning ability and working memory capacity related to mental speed? An investigation of stimulus-response compatibility in choice-reaction-time tasks. *Eur. J. Cogn. Psychol.* 18, 18–50.
- Woodworth, R. S. (1938). *Experimental Psychology*. New York, NY: Holt, Rinehart and Winston.
- Wykowska, A., Schubö, A., and Hommel, B. (2009). How you move is what you see: action planning biases selection in visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 1755–1769.

- Yamaguchi, M., and Proctor, R. W. (2011). Automaticity without extensive training: the role of memory retrieval in automatic implementation of task-defined rules. *Psychon. Bull. Rev.* 18, 347–354.
- Yehene, E., and Meiran, N. (2007). Is there a general task switching ability? *Acta Psychol.* 120, 169–195.
- Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., and Reynolds, J. R. (2007). Event perception: a mind/brain perspective. *Psychol. Bull.* 133, 273–293.
- Zacks, J. M., and Swallow, K. M. (2007). Event segmentation. *Curr. Dir. Psychol. Sci.* 16, 80–84.
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 19 January 2012; paper pending published: 28 February 2012; accepted: 09 April 2012; published online: 08 May 2012.
- Citation: Meiran N, Cole MW and Braver TS (2012) When planning results in loss of control: intention-based reflexivity and working-memory. *Front. Hum. Neurosci.* 6:104. doi: 10.3389/fnhum.2012.00104
- Copyright © 2012 Meiran, Cole and Braver. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.