



Attentional influences on memory formation: A tale of a not-so-simple story

J. Ortiz-Tudela¹ · B. Milliken² · L. Jiménez³ · J. Lupiáñez¹

Published online: 12 January 2018
© Psychonomic Society, Inc. 2018

Abstract

Is there a learning mechanism triggered by mere expectation violation? Is there some form of memory enhancement inherent to an event mismatching our predictions? Across seven experiments, we explore this issue by means of a validity paradigm. Although our manipulation clearly succeeded in generating an expectation and breaking it, the memory consequences of that expectation mismatch are not so obvious. We report here evidence of a null effect of expectation on memory formation. Our results (1) show that enhanced memory for unexpected events is not easily achieved and (2) call for a reevaluation of previous accounts of memory enhancements based on prediction error or difficulty of processing. Limitations of this study and possible implications for the field are discussed in detail.

Keywords attention · familiarity in recognition memory · implicit memory · memory · recognition

What are the long-term consequences of encountering an unexpected or conflicting situation? Are those episodes especially salient and encoded into long-term memory? Recent research and cognitive modeling have pointed at some form of cognitive conflict as a learning inducer (Krebs, Boehler, De Belder, & Egner, 2015; Rosner, D'Angelo, MacLellan, & Milliken, 2015a; Verguts & Notebaert, 2009). For example, Krebs et al. (2015) presented a group of participants with a gender categorization task in which they had to respond to the gender of human faces with a superimposed word. Critically, the word was a gender label that could be congruent, neutral, or incongruent with the response. The authors found impaired performance (i.e., longer response times and higher error rates) for incongruent than for congruent word–face pairs. More important, when their participants were later asked to remember previously presented faces in a surprise recognition test, they found better recognition for faces that were part of incongruent than congruent study items (see the full paper for converging neuroimaging findings). In a

conceptually similar study, Rosner et al., (2015a, b) asked participants to name aloud a list of target words that were displayed with matching or mismatching distractor words (e.g., the target word *piano* could be interleaved with either that same word or with the word *truck*). The authors reported that on a later surprise recognition memory test, recognition memory was superior (i.e., greater hit rates and fewer false alarms) for items that were part of incongruent word pairs in the study phase. Rosner et al. claimed that conflict detection might have recruited cognitive control which, in turn, may have up-regulated memory encoding for incongruent items. This set of results can easily be accommodated by Verguts and Notebaert's adaptation by binding model of conflict mediated learning. According to this model, conflict detection (arguably taking place in the anterior cingulate cortex), releases norepinephrine throughout the brain and thus strengthens the connections between coactive representations.

Another appealing and compatible interpretation of these effects comes from understanding incongruity as a mismatch between the predictions that we make and the outcomes that we actually encounter. Several models of predictive coding have proposed that prediction error can indeed trigger learning mechanisms that gather new evidence to update previously established associations. These updated representations would, in turn, improve future predictions (Henson & Gagnepain, 2010; van Kesteren, Ruiters, Fernández, & Henson, 2012).

Despite their surface appeal, several factors prove challenging for this set of ideas that map conflict or prediction error to episodic memory encoding. One such factor is that the direct

✉ J. Ortiz-Tudela
fjavierortiz@correo.ugr.es

¹ Department of Experimental Psychology, Universidad de Granada, Granada, Spain

² Department of Psychology, McMaster University, Hamilton, Canada

³ Department of Psychology, Universidad de Santiago, Santiago de Compostela, Spain

connection between incongruity and prediction error, at least in the studies outlined above, is not straightforward. A second factor of note is that it is relatively easy to find sets of data in the literature that support a putatively contrary principle, that is, that structured, consistent, congruent information is remembered better than incongruent information (Ortiz-Tudela, Milliken, Botta, LaPointe, & Lupiáñez, 2016; van Kesteren et al., 2012). Lastly, another factor worth noting is that, in most of the studies described above, the manipulations included at study necessarily imply a processing difficulty that is confounded with the expectation mismatch itself. Indeed, in those studies it is difficult to disentangle the separate contributions of prediction error (or conflict detection, or expectation mismatch) from additional processing that often follows as a consequence of prediction error.

Hence, in light of the inconsistent evidence and lack of a direct correspondence between prediction error and conflict, we developed a paradigm in which recognition memory is assessed as a function of fulfillment (or not) of an expectation, with those expectations realized either through explicit voluntary prediction (Experiments 1, 2, 3, and 4) or through bottom-up anticipation from a cue to a target (Experiments 5, 6, and 7). In this study we attempt to test directly the assumption that there is a core learning mechanism triggered solely by expectation mismatch. In order to do so, our experimental paradigm equates the conditions of interest in all features other than the expectation fulfillment. Namely, our expectation matching (i.e., valid) and mismatching (i.e., invalid) trials only differed from one another in that the anticipation about the upcoming location of the stimulus is either met or broken; other variables such as cognitive control exertion, selective attention deployment, or cognitive effort cannot be used to explain any differential results between valid and invalid trials since the two conditions of interest rely on those mechanisms equally. According to this approach, and in line with the conceptual overlap between our study and prior studies on conflict and memory encoding (Krebs et al., 2015; Rosner et al., 2015a, b), superior recognition memory ought to occur when an anticipatory cue results in an expectation that is violated. Specifically, our method focuses on cueing the location of an upcoming stimulus, which either matches or mismatches the actual location of that stimulus, and the straightforward prediction is that we should find better recognition memory for the stimuli appearing in an unexpected location. In the following experiments we thoroughly test this intuitive idea that stimuli appearing at unexpected locations are later better remembered, across a variety of manipulations of the standard validity paradigm (Chica, Martín-Arévalo, Botta, & Lupiáñez, 2014; Posner, 1980) and using a recognition memory test.

Generating (and breaking) expectancies

The most common way of inducing an expectation about a particular stimulus feature (e.g., its category, its color, its

location on the screen) is to validly cue said feature prior to the stimulus' appearance. The standard validity paradigm has been widely used to induce spatial expectancies (Posner, 1980). In its most common form, a visual cue is given prior to the presentation of a peripheral target stimulus. Although the cue always carries some form of spatial information, the extent to which that information relates to the actual location of the target defines the validity of the trial. Namely, on a valid trial, the cue correctly anticipates the location of the target; conversely, on an invalid trial, the cue directs attention toward a location different from that of the target. On invalid trials, when attention is misdirected, the appearance of the stimulus at an unexpected location triggers the reorienting system to disengage attention from the cued location, to shift attention toward the actual target location, and to reengage attention on the new target (Corbetta, Patel, & Shulman, 2008).

Another key aspect of validity paradigms is the nature of the anticipatory cue. The general consensus is that peripherally presented cues intrinsically carry spatial information (by their mere peripheral appearance) and automatically capture attention at their location, regardless of their informative value (i.e., exogenous orienting). On the other hand, centrally presented cues must contain some form of symbolic information (although see Lambert, Roser, Wells, & Heffer, 2006, for a discussion on symmetrical vs. asymmetrical cues) about a specific location and therefore require conscious and volitional action to properly shift attention (i.e., endogenous orienting). These two ways of orienting attention have different impacts on the processes that take place once the stimulus appears (Chica et al., 2014; Funes, Lupiáñez, & Milliken, 2007; Hauer & MacLeod, 2006; Markant & Amso, 2014).

Our aim was to combine these methods of generating a spatial expectation with a straightforward verbal memory task. To our knowledge, there are only two prior studies that have addressed related issues, although they do so indirectly, with a similar paradigm (Hauer & MacLeod, 2006; Markant & Amso, 2014). In a study by Markant and Amso (2014), the authors showed that spatial cueing effects in a study phase can predict recognition memory in a following test phase. More concretely, they demonstrated that participants who engaged in greater suppression of cued locations showed improved memory performance. Although they did not report individual memory performance separately for valid and invalid trials, their results can be taken broadly as evidence that processes that impact spatial cueing effects can impact memory encoding. In another study, Hauer and MacLeod (2006) used either central symbolic (endogenous) cues or peripheral abrupt onset (exogenous) cues to assess the influence of these two cueing methods at study on later memory performance. Their results indicate that endogenously attended words were remembered better than exogenously attended words. Crucial to our interests, although their study focused only on valid words (invalid words were to be ignored during the task), they

did report a small tendency towards better memory for invalid words in spite of the fact that they were not even attended. The present study builds on this prior work but explicitly requires our participants to attend (and respond) to both validly and invalidly cued words.

The present experiments

The overall structure of all of the experiments reported here is as follows. First, in a study phase, participants were presented with some variation of a validity paradigm. Namely, an anticipatory visual cue was presented with the purpose of generating an expectation about the location in which a given target word was going to appear. For Experiments 1, 2, 3, and 4, we used informative centrally presented arrows as cues; for Experiment 5 and onwards, we used uninformative peripherally presented asterisks as cues (see Fig. 1). Following each cue, a word was presented to which participants responded, with the word appearing either at the anticipated location (valid trials) or at the opposite location (invalid trials). The timing parameters were adjusted for each experiment so that the target would appear at a temporal interval following the cue that would be expected to produce facilitation on valid trials (Chica et al., 2014). In a second phase, a distracter task was completed for about 10 minutes consisting of math calculations either on a paper sheet (Experiments 1, 2, 3, 6, and 7) or on a computer display (Experiments 4 and 5). Finally, a surprise recognition test was administered for the words presented during the incidental study phase. The specific details of each experiment are described in the corresponding section.

Indeed, in the study phase, the processing of stimuli presented at cued locations was facilitated (i.e., faster response times and/or higher accuracy) relative to uncued locations in all experiments (see Table 1). As these results were anticipated and not central to our research question, and in order to ease reading of the manuscript, they are not discussed further.

Table 1 Summary of RTs (ms) from the study phases of all seven experiments reported in this manuscript

Experiment	Valid	Neutral	Invalid	Stats	<i>p</i> value	η_p^2
1	814	844	893	$F(2, 78) = 28.76$	<.000	0.42
2	755	774	803	$F(2, 56) = 23.66$	<.000	0.458
3	783	801	822	$F(2, 26) = 3.73$	0.037	0.223
4	601	645	675	$F(2, 38) = 26.34$	<.000	0.581
5	556	640	685	$F(2, 38) = 84.28$	<.000	0.816
6	512	–	666	$F(1, 29) = 88.53$	<.000	0.75
7	532	–	692	$F(4, 68) = 3.16$.019	0.157

Note. For Experiments 6 and 7, RTs for the different levels of repetition are collapsed in this table. See Figure 2 for a breakdown

However, these basic validity effects did allow us to test our main prediction concerning memory for words as a function of whether or not they appeared at the cued location. For the test phase, we hypothesized that if prediction error triggers better encoding of expectation-mismatching items, then memory performance ought to be better for invalid location trials than for valid location trials.

We use the term *prediction* to refer to any configuration of the system in response to an internal or external cue that anticipates a given stimulus feature. In other words, a prediction will occur when, in the presence of a cue that signals a given stimulus feature, the system reconfigures to efficiently respond to upcoming stimuli. From this perspective, any mismatch between the anticipated input and the actual stimulus will be considered a prediction error (Henson & Gagnepain, 2010). The specific empirical issue addressed in our study is whether such prediction errors regulate learning and memory broadly—will a strictly spatial prediction error regulate learning and memory in a task that has no obvious spatial component? To the extent that spatial information plays an important role in event encoding (Kahneman, Treisman, & Gibbs, 1992), spatial prediction errors may well play a key role in event segmentation processes known to impact both perception and memory (Kurby & Zacks, 2008). If this is the case, then we might also expect spatial prediction errors to impact performance in a simple verbal memory task.

Series A: Endogenous attention

Experiment 1

Voluntary shifts of attention towards a specific location are often studied with centrally presented symbolic cues. Among the types of cues used, arrows are the most common. Our initial goal was to evaluate the influence of explicit spatial expectation on memory encoding; as such, this first series of experiments used centrally presented arrows predictive of the target word location.

Participants

In order to meet counterbalancing requirements (see below), 40 undergraduate students (36 female; mean age = 22 years, $SD = 3.40$) from the Universidad de Granada were recruited to participate in the study in exchange for course credits. All participants in this and other experiments reported here signed a consent form approved by the local ethics committee, and the experiments were conducted according to the ethical standards of the 1964 Declaration of Helsinki.

Details of the procedure

All stimuli in both phases were presented, and responses were recorded using E-Prime 2.0 software (Psychology Software Tools, 2012).

Study phase Each trial of the study phase began with an “X” presented in the center of the screen for one second that served as fixation. On 80% of the trials, the fixation cross was replaced by an informative cue; that is, by two colored arrow heads pointing left and right, respectively. The two arrow heads were presented in different colors, and participants were instructed to pay attention to only one of them (colors were counterbalanced across participants). Participants were told that the relevant arrow would point to the location at which the upcoming stimulus was most likely to appear. On the rest of the trials (neutral trials), an uninformative cue appeared, in which the fixation cross was replaced by randomly oriented colored lines that provided no directional information (Martín-Arévalo, Lupiáñez, Botta, & Chica, 2015). Five hundred ms after onset of the cue, a colored (red or green) target word appeared in either the valid (75% of the trials) or the invalid (25% of the trials) location; across the entire experimental session, 60% of the trials were valid, 20% were invalid, and 20% were neutral. Words subtended approximately 2° vertically and between 5° and 7° horizontally of visual angle. Participants were instructed to use two keys of a standard QWERTY keyboard to categorize the word as either “natural” or “artificial.” The word remained on the screen until participants made a response, and speed and accuracy were both encouraged.

A total of 80 Spanish words representing natural objects and 80 words representing artificial objects were used for the study phase. Words were extracted from the Alameda and Cuetos database (Alameda & Cuetos, 1995) and we tried to equate them as much as possible in frequency and length, resulting, however, in artificial words being slightly longer (6.4 vs. 5.9) and more frequent (101 vs. 19 appearances per million words) than natural words, $t(158) = 1.90$, $p = .059$, Cohen’s $d = .297$, and $t(158) = 3.08$, $p = .002$, Cohen’s $d = .487$, respectively. In spite of these differences, it is important to note that validity was counterbalanced across participants so that every word would appear in every condition across the entire sample, and for every participant, half of the words were extracted from each category for each condition.

Test phase After the study phase, participants completed a surprise recognition test in which they judged centrally presented words as OLD (i.e., already presented in the study phase) or NEW (i.e., not presented in the study phase). The 160 words from the study phase and 40 new words were used in the memory test.

In addition to the OLD/NEW judgment, participants reported the confidence of their responses on a 4-point scale

(i.e., *I’m completely sure it is OLD, I think it is OLD, I think it is NEW, and I’m completely sure it is NEW*). After judging whether each test word was old or new, participants reported their memory for the color and location of the words presented in the study phase, again on a 4-point scale (i.e., *I’m completely sure the word was presented in red, I think it was presented in red, I think it was presented in green, and I’m completely sure the word was presented in green*; and a corresponding judgment about left-right location). As none of the results from these 4-point scales offered any useful information in this or any of the other experiments (in fact, memory for location and color responses was always at chance), these results are not discussed further.¹

Results and discussion

Separate repeated-measures ANOVAs were conducted for study and test phases. Both analyses included validity (valid vs. neutral vs. invalid) as a within-participants factor. For the study phase, response times to categorize the words served as the dependent variable; for the test phase, as it was impossible to assess independent false-alarm rates for valid, neutral, and invalid trials, hit rates were used as our measure of memory performance.

Study phase The percentage of correct responses for valid, neutral and invalid trials was 96%, and did not differ across conditions. However, a strong expectation effect for target location was reflected in the pattern of response times for valid (814 ms), neutral (844 ms), and invalid (893 ms) words, $F(2, 78) = 28.76$, $p < .001$, $\eta_p^2 = .42$.

Test phase Overall d' (1.70) indicated that participants accurately discriminated between OLD and NEW words; overall hit rate was .61 and overall false-alarm (FA) rate was .10. Unexpectedly, there was no significant difference between the hit rate for words presented at valid (.62), neutral (.62), and invalid (.58) locations, $F(2, 78) = 2.85$, $p = .064$, $\eta_p^2 = .068$.

Experiment 2

Despite the expected difference in response times between valid and invalid trials in the study phase, recognition did not differ for valid and invalid trials in the test phase. This unexpected result is inconsistent with our initial hypothesis, and surprising in light of the results of several previous studies. In these studies, an attentional manipulation during the

¹ Confidence ratings were analyzed on independent ANOVAs and collapsed after not finding any differences between these measures. “I’m completely sure” and “I think” responses were collapsed together for both old and new words.

study phase produced superior memory performance for the condition with higher attention demands at study (Krebs et al., 2015; Mulligan, Spataro, & Picklesimer, 2014; Rosner, D'Angelo, et al., 2015a; Rosner, Davis, & Milliken, 2015b; Swallow & Jiang, 2010). For example, consider the study of Krebs et al. (2015). They found better memory for faces presented in the context of incongruent than congruent face–word Stroop stimuli. A potentially important aspect of their procedure was the preexposure of all the materials before the study phase as a way to equate the subjective familiarity of all the stimuli. Although words from our stimulus set all had an intermediate normative frequency, small variations in perceived familiarity could have increased noise in our results, thus making it more difficult to find a true effect. In Experiment 2, we mimicked Krebs et al.'s procedure to reduce variation in the familiarity of our items.

Participants

In order to optimize resources, but still complete the six counterbalancing cycles, we reduced the total sample to 30 participants; all of them were undergraduate students from the Universidad de Granada and volunteered to participate in the study in exchange for course credits. One participant did not complete the entire experimental session so the final sample was composed of 29 students (25 female; mean age = 22.9 years, $SD = 3.92$).

Details of the procedure

The procedure and stimuli used in Experiment 2 were similar to Experiment 1 with the exception of an initial phase in which we preexposed all stimuli to participants. This initial phase consisted of the same categorization task used for the following study phase, but now involving all words that appeared later either as study words in the study and test phases or as lures in the test phase only. In this initial phase, the words were presented centrally in black. The rest of the procedure was identical to Experiment 1. During the test phase, participants were instructed to consider as old only those words that appeared in color, to the left or right of fixation, during the study phase.

Results and discussion

As expected, preexposing the words in the initial phase led to a large increase in false alarms in this experiment (.50) compared to Experiment 1. At the same time, preexposing the words prior to the study phase had no appreciable influence on the results of primary interest.

In the study phase, there was a validity effect of 47 ms in the response-time analysis, $F(2, 56) = 23.66$, $p < .001$, $\eta_p^2 = .458$, but no effect of validity in the accuracy analysis (95%

correct responses for each of the three conditions). Recognition sensitivity in the test phase was somewhat lower in this experiment than in Experiment 1, with overall hit and FA rates of .70 and .49, respectively. Nonetheless, the mean d' of .6 was significantly different from zero, $t(28) = 8.60$, $p < .001$, Cohen's $d_s = 2.26$. Importantly, however, the effect of validity on hit rates was not significant, $F < 1$, with hit rates not differing for valid (.70), neutral (.72), and invalid (.70) words.

Experiment 3

The absence of a validity effect in the memory tests of Experiments 1 and 2 was puzzling. However, it has been shown that when a single retrieval cue is associated with many memories, the cue can lose its utility in activating particular memories. This phenomenon is often found in the literature under the name of *cue-overload* (Watkins & Watkins, 1975). As Experiments 1 and 2 displayed words in just two locations on the screen (i.e., left and right), and a total of 160 different words were presented, the word-location ratio was 1:80. Ratios this large might diminish the distinctiveness of available cues, and thereby hurt memory performance (Logan, 1998).

In Experiment 3, we addressed this issue by presenting target words in four different locations (i.e., left, right, up, and down with respect to the fixation point) and by reducing the number of words to 80 in order to decrease the word-location ratio (note that d' was much reduced in Experiment 2, with 160 words).

Details of the procedure

The procedure of Experiment 3 was identical to Experiment 2, except for the following details. Instead of one out of two possible arrow cues we now presented one out of four possible arrow cues on each trial (i.e., pointing up, pointing down, pointing left, or pointing right). On half of the trials (i.e., valid trials), the target words appeared in the location indicated by the arrow cue, on 25% of the trials (i.e., invalid trials) the target word appeared in a location randomly chosen from among the three not-cued locations, and on the rest of the trials the same neutral cues from Experiments 1 and 2 were used (i.e., neutral trials). In addition, from the original pool of 160 words, we randomly selected a total 80 words. This method reduced the number of words presented at a single location from 80 to 20.

Participants

Due to the changes in the proportions of trials in each category, a sample size that was a multiple of four would produce complete counterbalancing of items across conditions. We

aimed for a sample size of 16 but fell short by two, ending up with a final sample of 14 undergraduate students (13 female; mean age = 22.43 years, $SD = 3.99$) from the Universidad de Granada, who also participated in the study in exchange for course credits.

Results and discussion

Reducing the cue load did not influence the results in any important way. We observed a validity effect for RTs in the study phase, 39 ms, $F(2, 26) = 3.73$, $p = .04$, $\eta_p^2 = .223$, but not for accuracy (95% for invalid, 96% for neutral, and 97% for valid). Overall memory sensitivity in the test phase ($d' = .87$) was significantly greater than zero, $t(13) = 6.36$, $p < .001$, Cohen's $d_s = 2.40$, hits = .75, FA = .54. However, memory performance was not affected by validity, $F < 1$; hit rates were similar for the valid (.77), neutral (.72), and invalid (.76) conditions.

Discussion of Series A

To this point, the results suggest that enhancement of memory due to a mismatch between target expectation and actual target appearance is not easy to observe. However, being aware of the problems of drawing conclusions from null results, and also being confident of the very straightforward prediction outlined in the Introduction, we decided to explore the issue further.

Several factors might have been driving the null effects found in Series A. One of the factors responsible for the null effects in this series of experiments may be the depth of processing of the words during the study phase. It has been shown that deep, semantic processing of stimuli usually leads to better encoding and remembering than shallow, structural processing (i.e., the levels of processing effect; Craik & Lockhart, 1972). In the prior experiments, relatively deep processing (natural vs. artificial) during the study phase may have made it difficult to observe an effect related to mismatches with expectation. In the following series we addressed this issue by changing the encoding task during the study phase to one with less emphasis on meaning.

Series B: Task and cue type

Rosner et al., (2015a, b) required participants to name words in the study phase, a shallow encoding task. In contrast, in the experiments in Series A, participants categorized words as “natural” or “artificial,” a deeper encoding task that requires access to semantic representations. These two types of encoding do produce a levels of processing effect, with better memory following semantic categorization than naming (Craik & Lockhart, 1972). It seemed possible that the

semantic categorization task used in the experiments in Series A incidentally improved memory for all of our stimuli to a level at which the expectation manipulation was unable to boost memory any further. In the present series, we addressed this issue.

Experiment 4

To equate our procedure with that of Rosner et al., (2015a, b) as much as possible, we conducted a spatial cueing study like that in Series A, but in the same lab, using the same task, and also the very same stimulus set as Rosner et al., (2015a, b).

Details of the procedure

The procedure used for Experiment 4 was identical to that of Experiment 1, except for the following. Instead of categorizing the words, upon presentation participants were required to name aloud the words as quickly as possible. The onsets of vocal responses were detected by a microphone and coded online by the experimenter as “correct” or “incorrect.” Note, however, that most (above 96%) of the words were read correctly and the few coded errors were caused by coughs or other extraneous noises. In addition, instead of Spanish words, a total of 240 high-frequency English (borrowed from Rosner et al., 2015a, b) words were used; 120 were presented in both the study and test phases and 120 served as lures for the test phase only. Additionally, in this experiment, as in the study of Rosner et al., (2015a, b), instead of presenting words until response, words were displayed for 1 s regardless of participants' response times. Stimuli were presented and responses were recorded using Psychopy software (Peirce, 2007). Lastly, all of the words included were counterbalanced not only in the study phase, but also between study and test phase across participants.

Participants

The new counterbalancing conditions required sample size multiples of 10. Twenty undergraduate students (12 female; mean age = 18.85 years, $SD = 1.63$) from McMaster University voluntarily participated in the study in exchange for course credits.

Results and discussion

The pattern of results obtained in this experiment corresponded closely to that found in previous experiments. A validity effect of 74 ms was observed for RTs, $F(2, 38) = 26.34$, $p < .001$, $\eta_p^2 = .581$, while accuracy was similar across conditions (99% correct on valid and 98% correct on neutral and invalid trials). Overall recognition sensitivity in the test phase was well above chance (hits = .56, FAs = .23 and $d' = 1.06$), and there were no differences in memory performance

for words presented at the valid (.56), neutral (.57), and invalid (.54) locations, $F < 1$. The only difference between these results and those reported previously was an expected decrease in overall hit rates, arguably due to naming rather than categorizing the target words—a successful levels-of-processing manipulation. In summary, level of processing, language differences, and other parameters that differed between this and prior experiments do not appear to explain the differences between our null result (no memory difference for the valid and invalid conditions) and that of Rosner et al., (2015a, b; better memory for the incongruent than congruent condition).

Experiment 5

Having set aside the task as responsible for the null results reported above, we explored further the underlying rationale for our prediction. As noted in the Introduction, different ways of orienting attention might have different impacts on the processing of subsequent stimuli (Chica et al., 2014; Hauer & MacLeod, 2006; Markant & Amso, 2014). Indeed, with a paradigm very similar to the one we used, Hauer and MacLeod showed that endogenous and exogenous orienting had very different impacts on later remembering of cued stimuli. The authors reported that endogenous orienting facilitated stimulus encoding and improved memory (Hauer & MacLeod, 2006). Therefore, it seemed possible that the null effects in our experiments result from an enhancement of memory due to endogenous cues that overrides the effect of expectation mismatching, thus rendering a null cue validity effect on memory. If this is the case, then forcing an automatic shift of attention, by means of an exogenous cue, might reveal an underlying validity effect. In Experiment 5, we addressed this hypothesis by using an exogenous cueing paradigm.

Details of the procedure

There were only two differences between the current and previous experiment. First, on each trial, instead of a centrally presented arrow, a row of three asterisks was presented at the center, to the left of center, or to the right of center of the screen for 200 ms. Second, this anticipatory cue was unpredictable of the location of the following target word, which appeared equally often to the left and right of center. All other details as well as the stimulus set used were the same as in the preceding experiment.

Participants

Although a complete counterbalancing cycle in Experiment 5 required six participants, a sample size of 20 was used to match that in Experiment 4. Twenty undergraduate students (18 female; mean age = 20.25 years, $SD = 3.93$) from

McMaster University participated in the study in exchange for course credits.

Results and discussion

Shifting from an endogenous to an exogenous cueing paradigm did not change the previous pattern of results. Once again, we obtained a validity effect on RT, 129 ms, $F(2, 38) = 84.28$, $p < .001$, $\eta_p^2 = .816$, no validity effect on accuracy (99% on valid, neutral and invalid trials), better than chance memory sensitivity, hits = .48, FAs = .21, and $d' = .8$, $t(19) = 8.69$, $p < .001$, Cohen's $d_s = 2.75$, but no recognition difference between words presented at the valid (.46), neutral (.50), and invalid (.47) locations, $F < 1$. Note that the overall hit rate decreased in this experiment (.48) relative to Experiment 4 (.55); however, this decrease did not reach statistical significance, $t(20) = 2.41$, $p = .055$, Cohen's $d = .40$. This result is consistent with previous literature, which has shown better overall memory performance when endogenous rather than exogenous attentional cueing is involved (Hauer & MacLeod, 2006). However, and importantly, there was no difference observed between the valid and invalid conditions.

Discussion of Series B

Neither the depth of processing of the target words nor the nature of the cueing manipulation could account for the absence of a validity effect obtained so far. Also, these experiments rule out the attribution of the null validity effect to some idiosyncrasy of our Spanish stimulus materials or our population of Spanish students. Indeed, Spanish and English are very different languages with respect to reading. Although Spanish is an almost completely transparent language, the structure of English requires readers to access the proper lexical representation to read words correctly (Bajo, Burton, Burton, & Canas, 1994). Yet we find null validity effects with both Spanish and English stimulus materials and participants.

Although none of the variables manipulated to this point changed our results, we considered another possibility. Krebs et al. (2015) tested a similar hypothesis with faces as stimuli, and they obtained a benefit for encoding on incongruent trials. One of the differences between their procedure and ours is the number of presentations of the stimuli during the study phase. They presented each of their faces up to three times, whereas we have relied on a single presentation approach. It may be that a single presentation of the target words did not allow an enhancement in encoding for invalid trials to be captured by our memory test. In the last experimental series, we explored whether repeated presentation of the target words, and also the valid/invalid status of those words, could result in a cumulative enhancement of the effect.

Series C. Number of presentations

Here we tested the effect of repeated presentation of words during the study phase. The number of presentations of words varied from one to five. We expected memory performance to improve with increased presentations of the words at study. More important, if encoding enhancement on invalid trials accrues with increased presentations at study, then we ought to find that learning curves across repeated presentation differ for valid and invalid trials.

Experiment 6

This experiment included repetition of study words as a variable to explore whether the learning curves for valid and invalid words differed.

Details of the procedure

The procedure of this experiment was similar to that in Experiment 5 in that we used peripheral cues and a naming task. The only changes were the following. First, we used Spanish words with Spanish participants; a list of 120 high-frequency words was selected in accord with LEXESP (Sebastián-Gallés, 2000). Second, and more important, the stimulus list was divided into five sets so that each set of 24 words was presented one, two, three, four, or five times in the study phase. As a consequence, the total number of trials was 360. Importantly, trial sequences were generated so that word repetitions were distributed relatively evenly across the entire sequence. That is, if a word was presented, for example, three times, the first presentation always took place within the first third of the stimulus list, the second presentation took place during the second third of the stimulus list, and the third presentation took place in the final third. This procedure prevented words being clustered together for a given participant in a specific part of the experiment. Both the validity and location of each word were held constant across word repetitions within each participant, but the assignment of the words to each of the repetition conditions, each of the validity conditions, and each location on the screen was counterbalanced across participants. No neutral trials were included in this experiment. The study phase duration increased from approximately 10 minutes in previous experiments to almost 30 minutes in this experiment.

Participants

The repetitions of the words in this experiment required a sample size multiple of 15, and therefore, for Experiments 6 and 7, 30 undergraduate students (26 female; mean age = 22.16 years, $SD = 3.83$) were recruited from the Universidad

de Granada to participate in the study in exchange for course credits.

Results and discussion

We conducted a 5 (presentations) \times 2 (validity) repeated-measures ANOVA on the RTs from the study phase. The main effect of validity was again highly significant, $F(1, 29) = 88.53$, $p < .001$, $\eta_p^2 = .75$. The interaction between presentation and validity was also significant, $F(4, 116) = 4.65$, $p = .001$, $\eta_p^2 = .14$, showing that the validity effect, although significant, $F(1, 29) = 6.09$, $p = .019$, $\eta_p^2 = .17$, was smaller for words presented five times (81 ms) than for words presented only once (158 ms).

Importantly, the test phase results were similar to those found in previous experiments. Memory sensitivity was again significantly different from zero, hits = .79, FAs = .17, $d' = 1.92$, $t(29) = 17.10$, $p < .001$, Cohen's $d_s = 4.42$, and higher than in previous experiments, as expected given that words were repeated several times in this experiment. To test our predictions, we conducted a two-way repeated-measures ANOVA on hit rates, with presentations (1 to 5) and validity (valid vs. invalid) as within-subject factors. The repetition manipulation proved to be effective, as there was a significant main effect of repetition, $F(4, 116) = 10.58$, $p < .001$, $\eta_p^2 = .267$; words presented more times were recognized better than words presented fewer times. This result was, of course, not surprising, but it was crucial to allow us to test whether the repetition manipulation influenced the validity effect on memory. The Validity \times Presentations interaction was far from significance (see Fig. 2), $F < 1$. The main effect of validity was also not significant, $F < 1$. Therefore, once again, the breaking of an expectation (in this case, a bottom-up expectation produced by an exogenous cue) produced a strong validity effect on RTs for words presented in the study phase, but did not affect memory for those words in the test phase, even when expectation was broken up to five times.

Experiment 7

Although repeating target words several times did improve memory in Experiment 6, this method did not produce a benefit for invalid trials relative to valid trials. Nevertheless, we deemed it possible that participants' recognition judgments may have hinged more on familiarity rather than on recollection, in particular given the length of the study phase. Indeed, other similar effects have been found to be sensitive to this distinction (Krebs et al., 2015). In addition, repeatedly cuing invalid locations might have an impact on the way the word-location compound is encoded. Specifically, the increased number of expectation mismatches might strengthen memory for the location of invalidly cued words.

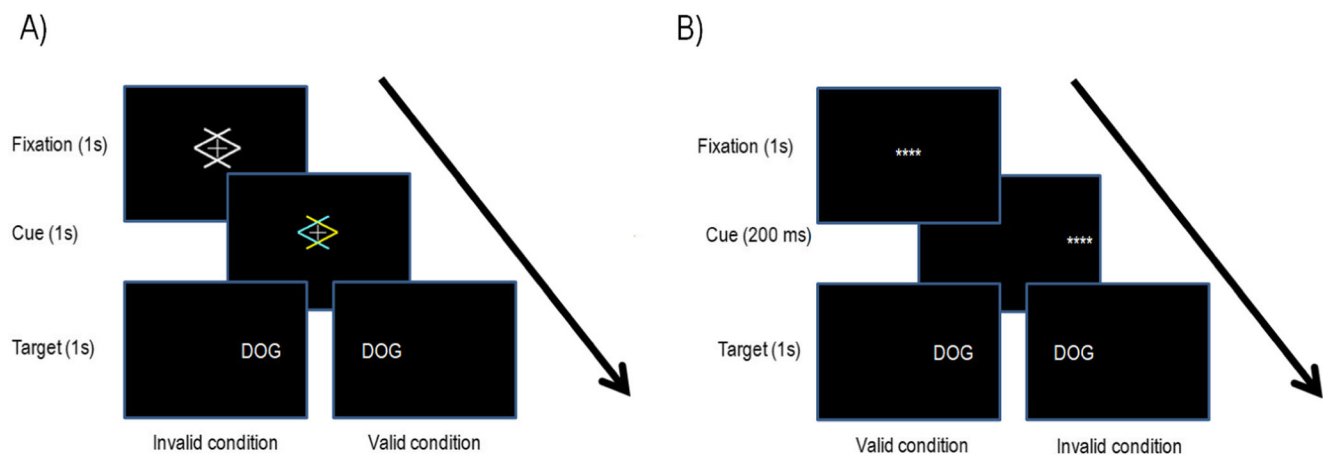


Fig. 1 Illustration of the experimental procedure for Experiments 1, 2, 3, and 4 (a) and Experiments 5, 6, and 7 (b)

We aimed to load the task heavily toward recollection of episodic details by testing our participants not only on the identity of the words but also on the location at which the words appeared during the study phase. Although prior experiments had shown performance in this type of task to be at chance, here we tested the idea that multiple study presentations might lift performance in this source task off floor. It was also an aim in this experiment to ensure that participants' recognition strategy at test was biased toward recollection of source information.

Details of the procedure

The procedure of the present experiment was identical to that of Experiment 6, except for the critical change in the memory test. On each trial, and after having judged each word as “OLD” or “NEW,” participants were asked to report the location (i.e., left or right) in which the word had been presented.

Participants

Thirty undergraduate students from the Universidad de Granada volunteered to participate in the study in exchange for course credit. One of them was unable to finish the experimental session, so the final sample size was 29 (27 female; mean age = 21.66 years, $SD = 3.39$).

Results and discussion

As in Experiment 6, we conducted a 5 (presentation) \times 2 (validity) repeated-measures ANOVA on the study phase naming times. The interaction between presentation and validity was not significant this time, $F < 1$, but, importantly, both main effects of presentation and validity were significant, $F(4, 68) = 3.16, p = .019, \eta_p^2 = .157$, and $F(1, 17)^2 = 67.24, p < .001, \eta_p^2 = .798$, respectively. Invalidly cued words were

² Due to a microphone malfunction, response times of 11 participants were not recorded.

always responded to more slowly than validly cued words, and response times increased with repetitions (from 562 ms with one presentation up to 638 ms with five presentations). Memory sensitivity was again significantly different from zero, hits = .77, FAs = .15, and $d' = 2.10, t(28) = 16.94, p < .001$, Cohen's $d_s = 5.24$. As in Experiment 6, we submitted the hit rates to a two-way ANOVA. The main effect of presentation was again significant, $F(4, 112) = 16, p < .001, \eta_p^2 = .363$. Neither the main effect of validity, $F(1, 28) = 2.18, p = .15, \eta_p^2 = .072$, nor the interaction between presentation and validity, $F < 1$, were significant. Memory for the location of the words was again at chance. Therefore, increasing the emphasis on recollection of episodic (or source) details did not affect the results.

Discussion of Series C

Repeated presentation of the target words up to five times in the study phase, with repeated words occurring in the same validity condition across the repetitions, did not improve the encoding of invalid words relative to valid words, in spite of producing a general improvement in memory. This result replicates and extends findings from our previous series, and rules out the possibility that a single exposure at study produces a weak boost in memory encoding that would be cumulative and measurable with multiple stimulus presentations.

Overall analyses of all experiments

Unexpectedly, none of the manipulations introduced across the experiments produced a validity effect on the memory for target words. Indeed, evidence from the analyses we conducted on the seven experiments points to an absence of influence of a mismatch between expected and actual location on memory sensitivity. Nevertheless, being aware that one cannot assume *evidence of absence* from *absence of evidence*, we used a Bayesian approach to reanalyze our data

Table 2 Summary of the analyses of HIT rates in the seven experiments

Experiment	Valid	Invalid	NHST	BF ₁₀	Interpretation	Posterior odds ₀₁
1	.62	.58	$F(2, 78) = 2.85$.052	Support for H_0 .	19.23
2	.70	.70	$F < 1$.169	Support for H_0 .	113.79
3	.77	.76	$F < 1$.447	Inconclusive evidence.	254.57
4	.56	.54	$F < 1$.152	Support for H_0 .	1,674.78
5	.46	.47	$F < 1$.414	Inconclusive evidence	40,45.37
6	.79	.79	$F < 1$.196	Support for H_0 .	20,639.65
7	.79	.77	$F(1, 28) = 2.18$.728	Inconclusive evidence	28,351.17

Note. Both NHST and Bayesian analyses are reported. Interpretations are based on Jarosz and Wiley (2014). Note that, although evidence from some of the experiments is tagged as *inconclusive*, the continuous nature of the BF and the combined analyses clearly supports the absence of an effect

(Wagenmakers, 2007). The Bayesian approach to statistics has two critical advantages over traditional null hypothesis significance testing (NHST). First, the output of the analyses has a more straightforward interpretation than that of a standard p value, which has often been subject to misinterpretation (Dienes, 2011; Fisher, Anderson, Peng, & Leek, 2014). Second, and particularly important with patterns of data such as the one described in this manuscript, Bayesian statistics easily allow for assessment of how much support we have for the null hypothesis. More specifically, most Bayesian analyses output a value often called Bayes factor (BF₁₀), which represents a ratio between the likelihood of two hypotheses given a set of data (i.e., how strongly a specific piece of evidence supports one hypothesis over the other). In this case, we compared our initial hypothesis (i.e., better memory for invalid trials, H₁) against the standard null hypothesis (i.e., no difference between valid and invalid trials, H₀). BF₁₀ ranges from zero to infinity, and its common interpretation is as follows: (i) when BF₁₀ is greater than 1, the greater its value, the stronger is the evidence for H₁ when compared with H₀, and (ii) BF₁₀ values ranging from zero to one indicate evidence in

favor of H₀ with smaller values indicating stronger evidence for H₀ (Wagenmakers, 2007). The standard practice is to use two thresholds that define three ranges of values for BF₁₀: values from zero to .33 can be considered as evidence favoring the absence of an effect, values ranging from .33 up to 3 are indicators of inconclusive evidence (e.g., low statistical power), and values above 3 should be interpreted as considerable evidence of the existence of an effect (i.e., a similar interpretation to that made from a p value smaller than .05 on traditional NHST; Jarosz & Wiley, 2014).

Table 2 summarizes the BFs for the seven experiments reported in this article. All Bayesian analyses were conducted with JASP software (JASP Team, 2016). As one can easily see, most of the experiments offer substantial evidence for a null effect. Take, for instance, Experiment 1. The BF₁₀ for the t test on the validity effect is .052. Since this value falls below .33 we must interpret it as reflecting that our data strongly supports H₀. In contrast, consider Experiment 7. The BF₁₀ value for that validity comparison is .728. According to the interpretation commonly accepted, we can conclude that evidence in this particular case is not informative of whether H₁ or H₀ is true.

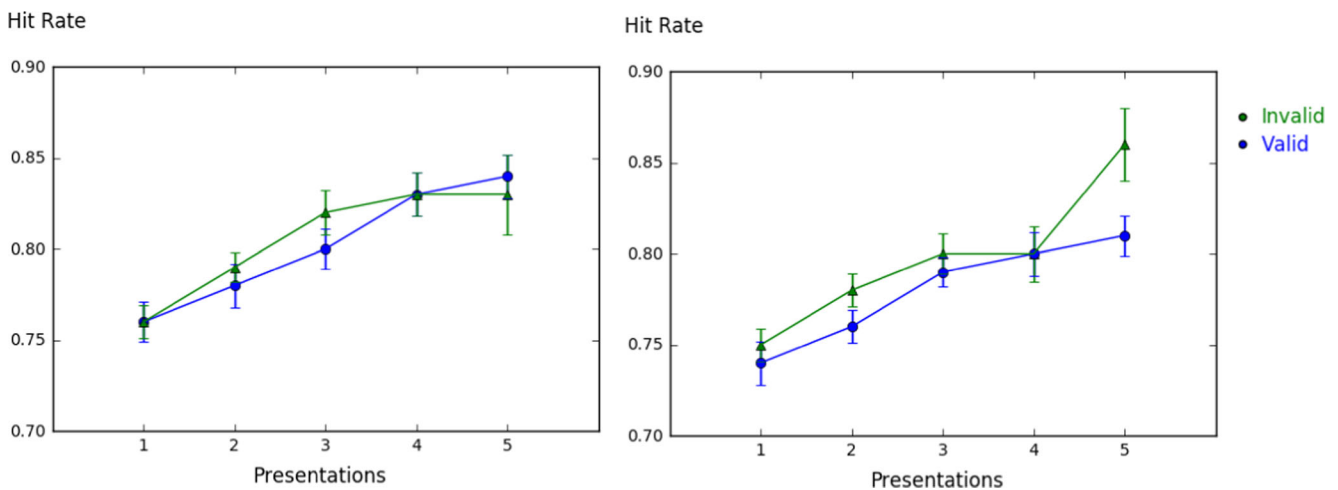


Fig. 2 Mean HIT rates for Experiments 6 and 7 across the number of presentations (from 1 to 5). Data from Experiment 6 are plotted in the left panel and data from Experiment 7 in the right panel. Error bars displaying

standard error of the variance with between-subjects differences removed (Cousineau & O’Brien, 2014)

Moreover, another important feature of Bayes' theorem is that it is relatively easy to combine sets of results to draw a broader picture. In its basic formulation, Bayes' theorem states that the posterior probability (i.e., how much we should change our beliefs over a given phenomenon after having collected the data) can be obtained by multiplying the prior probability (i.e., our beliefs regarding the phenomenon before any evidence is gathered) by the likelihood ratio (i.e., the BF). When no prior beliefs are assumed about the possible outcome of a given experiment, the prior probability is 1, and therefore the posterior probability and the BF are equivalent. This is a conservative but fair assumption to make when considering either an isolated experiment or the first one in a series of experiments.

However, prior probability can be updated to incorporate results of a new experiment that explores the same phenomenon (Ly, Etz, Marsman, & Wagenmakers, 2017); that is, we can use what we have learnt from Experiment 1 as a prior belief for Experiment 2. More concretely, the BF_{10} for Experiment 1 was .052 meaning that, if we would assume an equal prior probability for H_0 and H_1 , H_1 was .052 times more likely than H_0 , or conversely, H_0 was 19.23 times more likely than H_1 . For Experiment 2, the evidence gathered rendered a BF_{10} of .169;

however, we now know that H_0 is more likely than H_1 . When the obtained BF_{10} is multiplied by our updated beliefs from Experiment 1, the likelihood of H_0 being true grows to 113.79 times larger than the likelihood of H_1 . Subsequent steps can be taken to incorporate results from the rest of the experiments (see Table 2). As can be seen in Figure 3, the gathered evidence increases support for the null hypothesis exponentially. It is therefore evident that the data reported here clearly favor the absence of an effect.

In conclusion, with the current setup, namely, an anticipatory cue in the study phase, a 10-minute distracter interval and a recognition memory test, one can strongly conclude that there are no differences in memory sensitivity for valid and invalid trials (Hauer & MacLeod, 2006).

General discussion

The main objective of this study was to assess whether memory encoding is enhanced as a consequence of a mismatch between the expected location (either as straightforward top-down prediction or as bottom-up anticipation from cue to

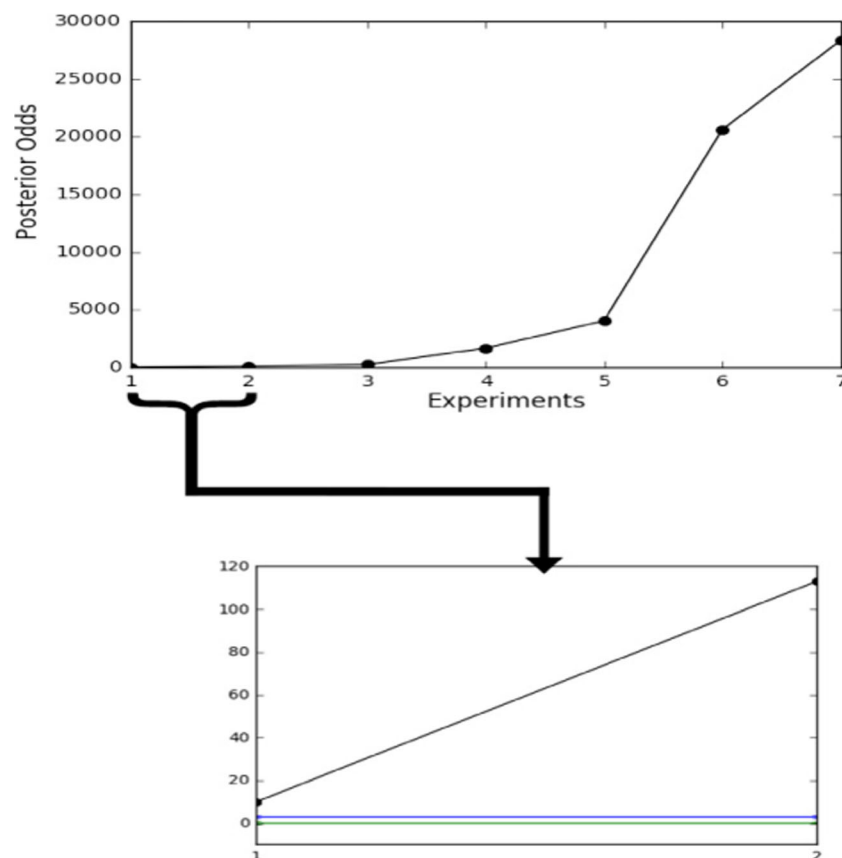


Fig. 3 Graphic representation of the evolution of the posterior odds across the experimental series (upper panel) and a zoom in depiction of the analyses of Experiments 1 and 2 for illustration purposes. Colored lines reflect the usually accepted likelihood ratio decision boundaries: 3 for accepting the null hypothesis (blue line) and .3 for substantial

evidence in favor of the alternative (green line). Note that the likelihood of the null hypothesis being true has been plotted here against the likelihood of the alternative hypothesis being true rather than the opposite; it should be interpreted then as higher values reflect stronger support for the null hypothesis. (Color figure online)

target) of a stimulus and its actual location. Expectation mismatch in our study is defined broadly to refer to any form of prediction error; that is, any divergence between how the system was set up to encode and what it actually encounters. This definition of prediction error is domain general and ought to be applicable across a range of very different paradigms that would include manipulations of proportion congruity, oddball paradigms, as well as the spatial cueing procedures. To explore this idea, we used a spatial validity paradigm in which we presented a word peripherally preceded by a cue that validly or invalidly signaled the location of the target word; by doing so, we were able to generate an expectation that was completely independent of the nature of the tested materials. Across seven different experiments we explored the nature of the cue and its orienting properties, the task participants had to accomplish once the word had appeared, the number of repetitions of the words during the study phase, and many other variables.

The results of the seven experiments converged on the same result: a mere spatial expectation mismatch does not trigger an enhancement in recognition memory. This surprising finding is not well accounted for by predictive coding or conflict-driven learning theories (Henson & Gagnepain, 2010; Krebs et al., 2015; Verguts & Notebaert, 2009). Likewise, accounts of attention that hinge on resource allocation would also have problems explaining why we found no encoding benefits for items that were correctly predicted and attended when compared with those in which attention was initially misdirected and then had to be refocused. Across the set of experiments, a great deal of evidence was collected that explored many nuances of the orienting paradigm; none of those many variations of the basic orienting method changed the results in a meaningful way, in spite of all of them producing the expected spatial attentional-orienting effects. Moreover, in addition to the independent results from each experiment revealing no difference between memory for validly and invalidly cued words, the combined analysis with updated Bayesian a priori probabilities across all experiments offered a very clear finding—the results here undeniably favor the absence of an effect of validity on memory.

Van Kesteren et al. (2012) describe a framework for how encoding of objects might be impacted positively both by matches with schemas and by novelty. Although the paradigm described in van Kesteren et al. (2012) is very different from the one used in our research, both share the idea that encoding might be impacted by a mismatch between the input the system expects and that which it actually encounters. Many different cognitive processes may precede and follow this mismatch in processing. For instance, when using visual scenes to establish a schema, the system is tuned to receive a visual representation of the upcoming stimulus and, when it eventually appears, this particular set of the system needs to be modified and updated to account for the new stimulus.

Likewise, once the new stimulus has been processed, a readjustment of the beliefs about the general structure of stimulus presentation is required; some rumination about what has just happened most surely also takes place. Although any particular experimental paradigms will rely more heavily on some processes than others, all such experimental paradigms share the commonality that the stimulus the system was prepared for does not match the stimulus that actually occurs. This is the broad notion of expectation mismatch that we refer to here, rather than the narrower notion associated with the conscious experience of an unfulfilled belief.

Previous studies that have found memory to be affected by some form of expectation mismatch or conflict may have differed in some critical way from the method used in our study. Namely, those studies may have included not only an expectation mismatch or cognitive conflict but also some associated increase in processing difficulty. For instance, in the study of Krebs et al. (2015) participants were required to perform a gender categorization task on face–word Stroop-like stimuli. Responding to this sort of stimulus entails exerting some form of cognitive control to solve the incompatibility that arises when two incompatible representations are active at the same time. This putative up-regulation in cognitive control may make this method a poor one for measuring the isolated contribution of expectation mismatch to memory encoding. Likewise, stimuli in the study of Rosner et al., (2015a, b) were constructed by combining either two instances of the same word (i.e., congruent pairs) or two different words (i.e., incongruent pairs). Responding to incongruent words may also involve an up-regulation in cognitive control, or may involve other processes that are distinct in nature from those involved in responding to congruent words (e.g., increased elaboration; Craik & Lockhart, 1972). Once again, the important point here is that this method also may be a poor one for measuring the isolated effect of expectation mismatch—the results of our seven experiments demonstrate that expectation mismatch on its own does not modulate recognition.

Of course, the above comments are not meant to imply that the memory encoding effects observed in the above studies are uninteresting. Differential elaboration of encoding in response to conflict implies an interesting link between transient shifts in cognitive control (Botvinick et al., 2001) and memory encoding. Moreover, further research on this topic may help to add some specificity to the *desirable difficulty principle*. Bjork and Bjork (1992) have noted a range of well-studied cognitive phenomena that appear to fit the notion that processing difficulty enhances retention. What is critically needed for this principle to be of practical use in guiding new research is an accepted and noncircular definition of the “types” of difficulty that enhance retention. As illustrated in the present study, if difficulty is defined simply as measurably worse performance in one condition than another in a study phase, then it

is quite clear that processing difficulty does not always enhance memory encoding. However, additional research on links between conflict-induced control adaptations and memory encoding may point to a more restricted and process-specific definition of difficulty that does reliably predict when memory encoding will be enhanced.

In any case, it seems reasonable to conclude that expectation mismatches in the study phase of the present experiments were very likely produced by mechanisms that are qualitatively different from the mechanisms that produced processing difficulty effects on memory in other studies (e.g., Krebs et al., 2015; Rosner et al., 2015a, b). Valid and invalid trials in our study differ specifically and solely in terms of whether an expectation generated by the cue is fulfilled or broken by a following target. It is this isolation of the influence of expectation mismatches from other, more complex forms of cognitive control adjustment that we view to be a merit of the present study. The results reported here make it clear that the readjustment of the system in response to a target that mismatches the prediction set up in response to a prior spatial cue does not lead to enhanced memory encoding of that target.

Conclusions and future directions

We find the pattern of results uncovered by this line of research interesting for both empirical and theoretical reasons. From an empirical perspective, several recent studies have demonstrated ways in which attentional manipulations at study impact memory significantly at test. Across these studies, memory performance has been shown to be sensitive to selective attention congruency (Krebs et al., 2015; Rosner, D'Angelo, et al., 2015a), perceptual encoding difficulty (Rosner, Davis, et al., 2015b), and to transient shifts in temporal attention (Swallow & Jiang, 2010). A quick look at the literature gives one the impression that it is a fairly simple exercise to impact memory encoding with attentional manipulations. Furthermore, there is a relatively straightforward mapping between several of those manipulations, the cue validity method used in the present study, and several cognitive and neural models (Henson & Gagnepain, 2010; Verguts & Notebaert, 2008) that would lead us to predict superior memory performance for invalid trials in our study. Clearly, this prediction was not supported in a straightforward manner by the results of our experiments. As such, our results point to a broad empirical issue worthy of further study. From a theoretical perspective, models of prediction error-based learning need to more exhaustively define what counts as a prediction and what counts as a prediction error. It appears from the present study that not every anticipatory cue will produce prediction error that up-regulates memory encoding in spite of affecting performance considerably. Does the task relevance of the stimulus feature associated with the anticipatory

cue matter? Does the processing level (e.g., semantic vs. nonsemantic) of the prediction matter? Does increasing the number of instances of a specific prediction error impact memory encoding (Henson & Gagnepain, 2010; van Kesteren et al., 2012)? At the same time, accounts of memory performance that focus on encoding difficulty (e.g., the desirable difficulty principle) must address how it is that two conditions, valid and invalid, with clearly different impacts on response outcome, produce no difference whatsoever in recognition memory—a clear definition of *difficulty* itself is needed (Bjork & Bjork, 1992; de Winstanley & Bjork, 1997).

Acknowledgements The present research was funded by the Spanish Ministerio de Economía y Competitividad with a research grant to Juan Lupiáñez (PSI2014-52764-P) and a research grant to Luis Jiménez (PSI2015-70990-P).

References

- Alameda, J. R., & Cuertos, F. (1995). Diccionario de frecuencias de las unidades lingüísticas del castellano [Spanish dictionary of the frequencies of linguistic units]. Oviedo, Spain: Universidad de Oviedo, Servicio de Publ.
- Bajo, M. T., Burton, A., Burton, E., & Canas, J.-J. (1994). Word recognition across orthographies: Another look at the interaction between context and degradation. *European Journal of Cognitive Psychology*, 6(2), 171–193. <https://doi.org/10.1080/09541449408520142>
- Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *Essays in honor of William K. Estes, Vol. 1. From learning theory to connectionist theory; Vol. 2. From learning processes to cognitive processes* (pp. 35–67). Hillsdale, NJ: Erlbaum.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological review*, 108(3), 624.
- Chica, A. B., Martín-Arévalo, E., Botta, F., & Lupiáñez, J. (2014). The spatial orienting paradigm: How to design and interpret spatial attention experiments. *Neuroscience and Biobehavioral Reviews*, 40, 35–51. <https://doi.org/10.1016/j.neubiorev.2014.01.002>
- Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron*, 58(3), 306–324. <https://doi.org/10.1016/j.neuron.2008.04.017>
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684. [https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X)
- de Winstanley, P. A., & Bjork, E. L. (1997). Processing instructions and the generation effect: A test of the multifactor transfer-appropriate processing theory. *Memory (Hove, England)*, 5(3), 401–421. <https://doi.org/10.1080/741941392>
- Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, 6(3), 274–290. <https://doi.org/10.1177/1745691611406920>
- Fisher, A., Anderson, G. B., Peng, R., & Leek, J. (2014). A randomized trial in a massive online open course shows people don't know what a statistically significant relationship looks like, but they can learn. *PeerJ*, 2, e589.
- Funes, M. J., Lupiáñez, J., & Milliken, B. (2007). Separate mechanisms recruited by exogenous and endogenous spatial cues: Evidence from

- a spatial Stroop paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, 33(2), 348–362. <https://doi.org/10.1037/0096-1523.33.2.348>
- Hauer, B. J. A., & MacLeod, C. M. (2006). Endogenous versus exogenous attentional cuing effects on memory. *Acta Psychologica*, 122(3), 305–320. <https://doi.org/10.1016/j.actpsy.2005.12.008>
- Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus*, 20(11), 1315–26. <https://doi.org/10.1002/hipo.20857>
- Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving*, 7, 2–9. <https://doi.org/10.7771/1932-6246.1167>
- JASP Team. (2016). JASP (Version 0.7. 5.5) [Computer software]. Retrieved from <https://jasp-stats.org>
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, 24(2), 175–219. [https://doi.org/10.1016/0010-0285\(92\)90007-0](https://doi.org/10.1016/0010-0285(92)90007-0)
- Krebs, R. M., Boehler, C. N., De Belder, M., & Egner, T. (2015). Neural conflict-control mechanisms improve memory for target stimuli. *Cerebral Cortex (New York, N.Y.: 1991)*. <https://doi.org/10.1093/cercor/bht283>
- Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, 12(2), 72–79. <https://doi.org/10.1016/J.TICS.2007.11.004>
- Lambert, A., Roser, M., Wells, I., & Heffer, C. (2006). The spatial correspondence hypothesis and orienting in response to central and peripheral spatial cues. *Visual Cognition*, 13(1), 65–88. <https://doi.org/10.1080/13506280500264460>
- Logan, G. D. (1998). What is learned during automatization? II. Obligatory encoding of spatial location. *Journal of Experimental Psychology: Human Perception and Performance*, 24(6), 1720–1736. <https://doi.org/10.1037/0096-1523.24.6.1720>
- Ly, A., Etz, A., Marsman, M., & Wagenmakers, E. J. (2017). Replication Bayes factors from evidence updating. Retrieved from <https://psyarxiv.com/u8m2s>
- Markant, J., & Amso, D. (2014). Leveling the playing field: Attention mitigates the effects of intelligence on memory. *Cognition*, 131(2), 195–204. <https://doi.org/10.1016/j.cognition.2014.01.006>
- Martín-Arévalo, E., Lupiáñez, J., Botta, F., & Chica, A. B. (2015). Endogenous attention modulates attentional and motor interference from distractors: Evidence from behavioral and electrophysiological results. *Frontiers in Psychology*, 6, 132. <https://doi.org/10.3389/fpsyg.2015.00132>
- Mulligan, N. W., Spataro, P., & Picklesimer, M. (2014). The attentional boost effect with verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(4), 1049–1063. Retrieved from <http://cat.inist.fr/?aModele=afficheN&cpsid=28599188>
- Ortiz-Tudela, J., Milliken, B., Botta, F., LaPointe, M. R., & Lupiáñez, J. (2016). A cow on the prairie vs. a cow on the street: Long-term consequences of semantic conflict on episodic encoding. *Psychological Research*. <https://doi.org/10.1007/s00426-016-0805-y>
- Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1/2), 8–13. <https://doi.org/10.1016/j.jneumeth.2006.11.017>
- Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, 32(1), 3–25. <https://doi.org/10.1080/00335558008248231>
- Psychology Software Tools. (2012). E-Prime 2.0 [Computer software]. Pittsburgh, PA. Retrieved from <http://www.pstnet.com>
- Rosner, T. M., D'Angelo, M. C., MacLellan, E., & Milliken, B. (2015a). Selective attention and recognition: Effects of congruency on episodic learning. *Psychological Research*. <https://doi.org/10.1007/s00426-014-0572-6>
- Rosner, T. M., Davis, H., & Milliken, B. (2015b). Perceptual blurring and recognition memory: A desirable difficulty effect revealed. *Acta Psychologica*, 160, 11–22. <https://doi.org/10.1016/j.actpsy.2015.06.006>
- Sebastián-Gallés, N. (2000). *LEXESP: Léxico informatizado del español* [Spanish informatized lexicon]. Edicions Universitat Barcelona.
- Swallow, K. M., & Jiang, Y. V. (2010). The attentional boost effect: Transient increases in attention to one task enhance performance in a second task. *Cognition*, 115(1), 118–132. <https://doi.org/10.1016/j.cognition.2009.12.003>
- van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, 35(4), 211–219. <https://doi.org/10.1016/j.tics.2012.02.001>
- Verguts, T., & Notebaert, W. (2008). Hebbian learning of cognitive control: Dealing with specific and nonspecific adaptation. *Psychological Review*, 115(2), 518–25. <https://doi.org/10.1037/0033-295X.115.2.518>
- Verguts, T., & Notebaert, W. (2009). Adaptation by binding: a learning account of cognitive control. *Trends in Cognitive Sciences*, 13(6), 252–257. <https://doi.org/10.1016/j.tics.2009.02.007>
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14(5), 779–804. <https://doi.org/10.3758/BF03194105>
- Watkins, O. C., & Watkins, M. J. (1975). Build up of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning & Memory*, 1(4), 442–452. <https://doi.org/10.1037/0278-7393.1.4.442>