

# Previously acquired cue–outcome structural knowledge guides new learning: Evidence from the retroactive-interference-between-cues effect

David Luque<sup>1,2,3,4</sup> · Joaquín Morís<sup>2,3,4</sup> · Francisco J. López<sup>2,3</sup> · Pedro L. Cobos<sup>2,3</sup>

Published online: 12 April 2017  
© Psychonomic Society, Inc. 2017

**Abstract** The effect of retroactive interference between cues predicting the same outcome (RIBC) occurs when the behavioral expression of a cue–outcome association (e.g., A→O1) is reduced due to the later acquisition of an association between a different cue and the same outcome (e.g., B→O1). In the present experimental series, we show that this effect can be modulated by knowledge concerning the structure of these cue–outcome relationships. In Experiments 1A and 1B, a pretraining phase was included to promote the expectation of either a one-to-one (OtO) or a many-to-one (MtO) cue–outcome structure during the subsequent RIBC training phases. We hypothesized that the adoption of an OtO expectation would make participants infer that the previously learned A→O1 relationship would not hold any longer after the exposure to B→O1 trials. Alternatively, the adoption of an MtO expectation would prevent participants from making such an inference. Experiment 1B included an additional condition without pretraining, to assess whether the OtO structure

was expected by default. Experiment 2 included control conditions to assess the RIBC effect and induced the expectation of an OtO or MtO structure without the addition of a pretraining phase. Overall, the results suggest that participants effectively induced structural expectations regarding the cue–outcome contingencies. In turn, these expectations may have potentiated (OtO expectation) or alleviated (MtO expectation) the RIBC effect, depending on how well these expectations could accommodate the target A→O1 test association. This pattern of results poses difficulties for current explanations of the RIBC effect, since these explanations do not consider the incidence of cue–outcome structural expectations.

**Keywords** Associative learning · Interference · Interference between cues · Structural knowledge · Top-down

The ability to predict future outcomes from present cues is a task that organisms need to face on a daily basis and constitutes an important prerequisite for adaptive behavior. Consequently, a great deal of research has been devoted to how humans learn cue–outcome contingencies (human contingency learning, or HCL hereafter; De Houwer & Beckers, 2002; Shanks, 2010). Correct understanding of HCL requires an explanation of when and how humans acquire new contingency information, as well as when and how previously acquired contingency information needs to be revised by new information. This topic has greatly benefited from the study of phenomena such as interference (e.g., Pineño & Miller, 2007). *Interference* refers to a difficulty in the recall of specific information due to the acquisition, at a different time, of some other information (Anderson, 2003). Among the different interference effects described in the literature, one is particularly intriguing: namely, the retroactive cue - interference effect (also known as *retroactive*

---

**Electronic supplementary material** The online version of this article (doi:10.3758/s13421-017-0705-4) contains supplementary material, which is available to authorized users.

---

✉ David Luque  
david.luque@gmail.com

<sup>1</sup> School of Psychology, UNSW Sydney, Sydney 2052, NSW, Australia

<sup>2</sup> Departamento de Psicología Básica, University of Málaga, Málaga, Spain

<sup>3</sup> Instituto de Investigación Biomédica de Málaga (IBIMA), Málaga, Spain

<sup>4</sup> Department of Psychology, University of Oviedo, Oviedo, Spain

*interference between cues of the same outcome*, or RIBC hereafter). RIBC occurs when the behavioral expression of an association between a cue and an outcome (e.g., cue A → Outcome 1, O1) is reduced by later acquisition of an association between a different cue and the same outcome (B → O1; see Table 1). Interestingly, although there has been recent interest in this phenomenon within the HCL field, this effect was first studied in the paired-associate learning literature several decades ago (e.g., Keppel, Bonge, Strand, & Parker, 1971).

Despite the simplicity of this effect, most of the relevant HCL theories cannot account for the basic RIBC effect without post-hoc assumptions (Cobos, López, & Luque, 2007; Escobar, Pineño, & Matute, 2002; Matute & Pineño, 1998a, 1998b; Miller & Escobar, 2002; Vadillo, Castro, Matute, & Wasserman, 2008). These ad-hoc RIBC accounts consist of adaptations of models originally designed to explain other phenomena, such as backward blocking (Dickinson & Burke, 1996; Van Hamme & Wasserman, 1994) or extinction (Bouton, 1993). As a consequence, they rely either on learning of the new B → O1 relationship reducing the previously learned A → O1 association (i.e., the *learning deficit* account) or on the former association interfering with the retrieval of the latter one (i.e., the *retrieval deficit* account).

It is not necessary to describe the specific details of these explanations, but will suffice to say that they are essentially based on the following ideas. The learning deficit account is based on the idea that the activation of the mental representation of O1 during B → O1 learning can elicit the activation of the previously learned A → O1 relationship. Then, to reduce interference, the memory trace for the retrieved A → O1 relationship is weakened (Matute & Pineño, 1998a, 1998b). On the other hand, the retrieval deficit account is based on two ideas: First, interference occurs if two learned associations have a common element in a common temporal location (whether this be the cue or the outcome location). Second, the most recent learned relationship is primed at the expense of other older associations with which it shares a common element, which can either be the cue or the outcome (Miller & Escobar, 2002). It is worth noting that both accounts are based on the operation of associative learning and associative

retrieval mechanisms. Associative mechanisms are thought to be fast acting, automatic, and require little cognitive resources to operate (e.g., Cobos, Gutiérrez-Cobo, Morís, & Luque, 2017; Morís, Cobos, Luque, & López, 2014; Shanks, 2007; Wagner, 1981). Because of this, they are prototypic of the low-level processes proposed by two system theories, as opposed to high-level inference processes (see Darlow & Sloman, 2010, for a review).

Importantly, the evidence used so far to support these explanations does not discard the possibility that the RIBC effect is, at least in part, due to inference processes and based on representations that are not compatible with the standard associative view of HCL.<sup>1</sup> In fact, there is evidence showing that the RIBC effect may depend on inference processes based on abstract representations of how cues and outcomes are expected to relate to each other. For instance, some studies have shown top-down influences of general causal knowledge on the RIBC effect. RIBC becomes weak or nonexistent when participants are instructed that the cues (A and B in Table 1) are independent causes (e.g., two different medicines) producing a certain effect (e.g., headache relief; O1 in Table 1) (Cobos et al., 2007; cf. Waldmann & Holyoak, 1992). Causal content modulations of RIBC have been explained within the general framework of causal model theory (Waldmann & Holyoak, 1992). This model emphasizes the use of general causal knowledge in order to learn new causal relationships. Thus, since it is well known that two causes can independently produce the same effect, given a causal scenario with two independent causes producing one effect, participants will not perceive any interference or conflict between the relationships learned during a RIBC experiment. As a consequence, a cause → effect scenario will protect the retrieval of the A → O1 relationship from any interference, and thus no RIBC is expected. Conversely, when cues and outcomes are interpreted as effects and causes, respectively, the RIBC effect is promoted since causes (O1) are not expected to produce negatively correlated effects (Cues A and B). Interestingly, most experiments on RIBC conducted so far have been framed within scenarios that are susceptible to causal interpretations that may favor the RIBC effect through reasoning and abstract representations of cue–outcome relationships (see Cobos et al., 2007, for a systematic review).

**Table 1** RIBC design

Group	Phase 1	Phase 2	Test
Experimental	A → O1	B → O1	A?
	F → O3	F → O3	
Control	A → O1	B → O2	A?
	F → O3	F → O3	

The letters denote cues (cue A, B, and F [filler cues]), and the letter “O” followed by a number is used to denote the outcomes (Outcome 1, O1; Outcome 2, O2; Outcome 3, O3). Question marks indicate that this cue was presented at test (and therefore no outcome was presented)

<sup>1</sup> The causal-reasoning explanation for the RIBC effect does not postulate that *memory interference* is the cause of the effect. This might be confusing, since the name of the effect itself (i.e., retroactive *interference* between cues of the same outcome) mentions “interference,” which might suggest that the effect *has to be produced by memory interference*. Thus, it is important to make clear that the term *RIBC* should be used only for the empirical phenomenon or effect as it is described at the end of the first paragraph of the introduction (see also Table 1), independently of what mechanisms actually produce it (see De Houwer, 2007). Thus, the RIBC effect might be caused by memory interference, as is postulated by the retrieval deficit account, or by other psychological mechanisms—such as causal reasoning, as was proposed by Cobos et al. (2007).

Although the RIBC account based on the causal model theory can explain most of the RIBC effects found within the HCL literature (see Cobos et al., 2007; Luque, Cobos, & López, 2008), it cannot explain why RIBC has been found in several HCL experiments in which the instructions did not describe cue–outcome relationships as *causal* relationships (Luque, Morís, & Cobos, 2010; Luque, Morís, Cobos, & López, 2009; Luque, Morís, Orgaz, Cobos, & Matute, 2011). However, in our opinion, even when no causal interpretation of the task may reasonably be expected, the RIBC effect may still be the result of reasoning processes based on abstract representations of cue–outcome relationships. We propose here a new hypothesis for the RIBC effect that does not require a causal framing for cue–outcome relationships. The main hypothesis of the present study is that RIBC is promoted when learners are encouraged to expect cues and outcomes to relate following a one-to-one (OtO) cue–outcome structure. The OtO structure entails that each cue is related to only one outcome, and that each outcome is related to only one cue. An OtO structure expectation may lead learners to experience a conflict in a RIBC experiment, as learners experience  $A \rightarrow O1$  trials in Phase 1 followed by  $B \rightarrow O1$  trials in Phase 2. To adhere to the OtO structure expectation, some learners may solve the conflict by inferring that the  $A \rightarrow O1$  relationship does not hold any longer during Phase 2, since  $O1$  is being consistently predicted by cue  $B$  during the whole of Phase 2. This inference would then produce a decrease in the expression of the previously learned  $A \rightarrow O1$  relationship in the test phase (it should be noted that the test and second phase are usually indistinguishable to participants). If a sufficient number of participants embrace the OtO expectation during a RIBC experiment, and therefore engage in the reasoning processes described above, then a decrease in the average responses anticipating  $O1$  (in short, hereafter,  $O1$  responses) is expected during the test—that is, RIBC.

A straightforward prediction from our hypothesis is that RIBC should be reduced if learners are somehow encouraged to expect a many-to-one (MtO) cue–outcome structure. In such a case, no conflict would be perceived between the  $A \rightarrow O1$  and the  $B \rightarrow O1$  trials, and no inference affecting the reevaluation of the former relationship would be needed.

To reiterate, the learning deficit and retrieval deficit accounts claim that RIBC is the consequence of a memory deficit (due either to a weak memory trace, or to a retrieval deficit, respectively), produced by the operation of associative mechanisms. For this reason, these accounts are not able to explain the formation of abstract beliefs about cue–outcome relationships, or their possible influence on future learning about other cues and outcomes. Our proposal, on the other hand, is based on reasoning processes. It claims that, because of this reasoning operating during HCL, participants infer that a specific cue–outcome structure applies, even for completely new cue–outcome relationships. This kind of abstract learning goes

beyond the scope of associative models (Waldmann & Holyoak, 1992). The proposed hypothesis also departs from the explanation based on causal model theory, since it is not constrained to causal learning scenarios, which is important, given that RIBC has been reported in noncausal learning tasks (e.g., Luque et al., 2009).

The main aim of the present study was to test whether the RIBC effect changes depending on whether participants are trained to expect an OtO or MtO cue–outcome structure. In Experiments 1A and 1B, we altered the participants' structural belief by including a pretraining phase that was consistent with either an OtO structure (OtO group) or an MtO structure (MtO group; see Table 2). In Experiment 1B we also investigated whether an OtO structure expectation might be contemplated by default. For this purpose, Experiment 1B included an additional condition without pretraining (*no-pretraining* group). Experiment 2 addressed similar questions with an improved design. First, additional control conditions to assess the RIBC effect were included. In this way, the effect of promoting structural expectations about the magnitude of the RIBC effect could be more adequately addressed. Structural beliefs were not trained during a pretraining phase in Experiment 2. Instead, participants received training trials that could be consistent with either an OtO (OtO group) or an MtO (MtO group) cue–outcome structure during both the first and the second learning phases. The magnitude of the RIBC found in the OtO group was then compared with that of the RIBC effect found in the MtO group.

In all experiments we predicted more  $O1$  responses to cue  $A$  during testing in MtO than in OtO conditions. Interestingly, the learning deficit and retrieval deficit accounts of RIBC predict no differences at all between such groups, since all relationships used for the OtO or MtO training never shared any common element (cue or outcome) with the relationships involved in the RIBC effect.

## Experiments 1A and 1B

### Method

**Participants and apparatus** In all, 163 and 175 participants from the University of Málaga took part in Experiments 1A and 1B, respectively, for course credits. The number of participants was determined by the following stopping rule. The recruitment phase lasted for one week, and made use of a web-based recruitment platform. All of the participants who signed up and successfully completed the experiment were included in the final sample. The experiments were carried out in a quiet room with 10 semi-isolated cubicles equipped with Windows XP PCs (Microsoft, Redmond, WA, USA). The task was programmed using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA).

**Table 2** Design of Experiments 1A and 1B

Group	Pretraining	Phase 1	Phase 2	Test
Many-to-One (Exps. 1A & 1B)	C→O2 (x10)	A→O1 (x10)	B→O1 (x10)	A?
	D→O2 (x10)	F→O3 (x10)	F→O3 (x10)	
	F→O3 (x10)			
One-to-One (Exps. 1A & 1B)	C→O2 (x20)	A→O1 (x10)	B→O1 (x10)	A?
	F→O3 (x10)	F→O3 (x10)	F→O3 (x10)	
No -pretraining (Exp. 1B)		A→O1 (x10)	B→O1 (x10)	A?
		F→O3 (x10)	F→O3 (x10)	

The letters denote cues (cue A, B, C, D, and F [filler cues]), and the letter “O” followed by a number is used to denote the outcomes (Outcome 1, O1; Outcome 2, O2; Outcome 3, O3). The number of repetitions of each type of trial is shown between parentheses. For instance, A→O1 (x10) indicates that cue A followed by the Outcome 1 was presented ten times. Question marks indicate that this cue was presented at test (and therefore no outcome was presented)

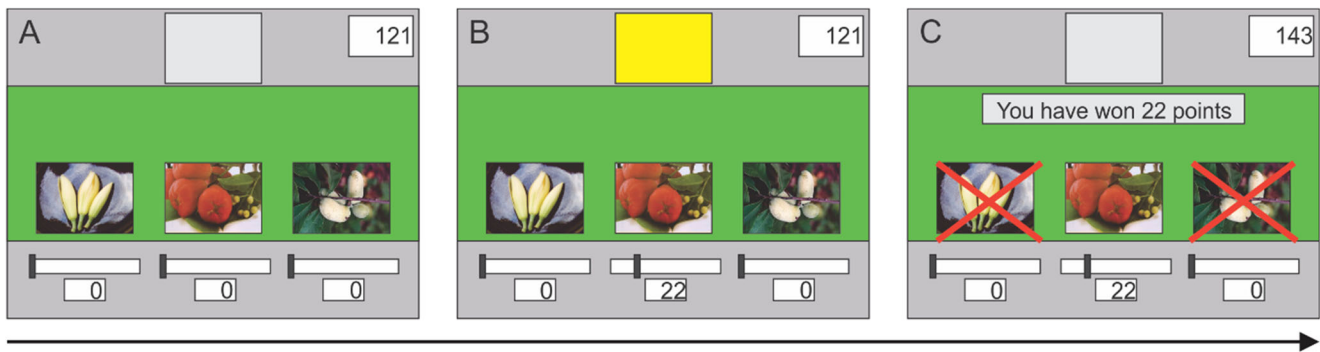
**Design** The design of both experiments included an independent variable, *Pretraining*, with two between-subjects groups in Experiment 1A (OtO and MtO groups) and three between-subjects groups in Experiment 1B (OtO, MtO, and no-pretraining groups) (see Table 2). In the pretraining phase, participants in the OtO groups learned two independent cue–outcome relationships that shared neither the cue nor the outcome. In contrast, participants in the MtO groups learned three relationships in which two cues on their own predicted the same outcome, whereas the remaining cue predicted a different outcome. Later, all participants were exposed to the experimental condition of a RIBC design. Accordingly, they learned that a cue predicted an outcome during the first training phase, and that a different cue also predicted the same outcome during the second training phase. Additional filler cues were included throughout the pretraining and the two training phases (F→O3 in Table 2) to force participants to discriminate between cues in all phases and groups.

**Procedure** The procedure was equivalent to that used by Luque et al. (2009, Exp. 2). First, participants read the instructions of the experiment and were given the opportunity to resolve any doubts they might have had regarding the experiment. They were told that the objective of the task was to earn as many points as possible by choosing the correct response option on the basis of the cue presented on each trial. To do so, participants had to learn the relationship between each of several cues and each of several outcomes on a trial- by-trial basis (the original instructions and their translation can be found in the [supplemental materials](#)). Once the participants had read the instructions and resolved any doubts, the experimental task began.

Colored rectangles (magenta, yellow, light green, cyan and red) were used as cues, and pictures of fictitious plants (labeled as *Kollin*, *Dobe*, and *Yamma*) as outcomes. The role of the stimuli was counterbalanced across participants following a Latin square design.

Figure 1 shows a trial example. Each trial began with the presentation of the three possible outcomes at the bottom of

the screen, in three different positions (left, center, and right). After 2 s, a cue appeared centrally in the upper section of the screen. Participants were required to make a response within 2.5 s by pressing the key corresponding to one of the three outcomes. The left, center, and right outcomes were assigned to the response keys “1,” “2,” and “3,” respectively, from the numeric keyboard. As such, the relative spatial locations for the response keys were the same as for the corresponding spatial locations of the outcomes on the screen. Participants could monitor their responses for each option by means of a horizontal scroll bar under each outcome that moved from the left to the right as the number of responses to the corresponding outcome increased. There was also a small text box on top of each plant picture that showed the exact number of responses for that outcome. Responses for an outcome increased either by repeatedly pressing the corresponding response key or by just keeping the key pressed, the latter method allowing for fast, continuous responding: +1 point each 45 ms. Therefore, the maximum amount of responses on a trial was ~55. Once the 2.5 s had elapsed, all incorrect outcomes disappeared, and only the correct one remained visible on the screen. As additional feedback, the number of points earned on a particular trial was displayed on the screen, which was the number of responses for the correct outcome minus the sum of the responses for the other two outcomes. Participants could then move on to the next trial by pressing the “X” key. For all experimental conditions, pretraining and Phase 1 shared the same physical context, which was different from the physical context that was used in Phase 2 and the test (the test context was always the same as the Phase 2 context). These physical contexts consisted of a vertically centered rectangle with a width that was equal to the screen size and a height equal to a quarter of the screen’s height. The rectangle was either green or red, and the assignment of each color to each phase was counterbalanced across participants, following a Latin square design. This contextual cue was included in order to keep the procedure as similar as possible to that of previous experiments that have successfully found the RIBC effect in our laboratory (e.g., Luque et al., 2009).



**Fig. 1** Trial structure for all experiments. The cue is shown as a rectangle at the top of the screen (initially in gray, panel A). The total points earned throughout the task are shown in the text box at the top-right corner of the screen (in this example, the trial starts with 121 points already earned, panel A). The cue appears 2 s after the beginning of the trial—in this example, the color yellow (B). The cue is presented for 2.5 s, during which participants can respond to the different outcomes—the plants at the bottom of the screen. During the actual experiment, the plant pictures

were labeled with fictional names. For the sake of clarity, these names are not represented in this figure. As can be seen in panel B, the participant in this example is trying to win 22 points by predicting that the outcome at the center is the correct response. After that, the incorrect response options are each marked with a red X, the total points are updated, and the cue disappears (C). Participants could then move on to the next trial by pressing the “X” key

Trials were pseudorandomly ordered within each learning phase, with the constraint that no more than two consecutive trials of the same type could occur. During the pretraining phase, the MtO group received ten C→O<sub>2</sub>, ten D→O<sub>2</sub>, and ten F→O<sub>3</sub> trials, whereas the OtO group received 20 C→O<sub>2</sub> and ten F→O<sub>3</sub> trials. All groups received ten A→O<sub>1</sub> trials intermixed with ten F→O<sub>3</sub> trials during Phase 1, and ten B→O<sub>1</sub> trials intermixed with ten F→O<sub>3</sub> trials during Phase 2. After completion of the second phase, participants were presented with cue A on an additional test trial. The transition from the second training phase to the test trial was not signaled to participants in any way. The only procedural differences between this test trial and the previous trials were that participants had 5 s to respond and that the test trial ended without any outcome presentation or feedback.

## Results

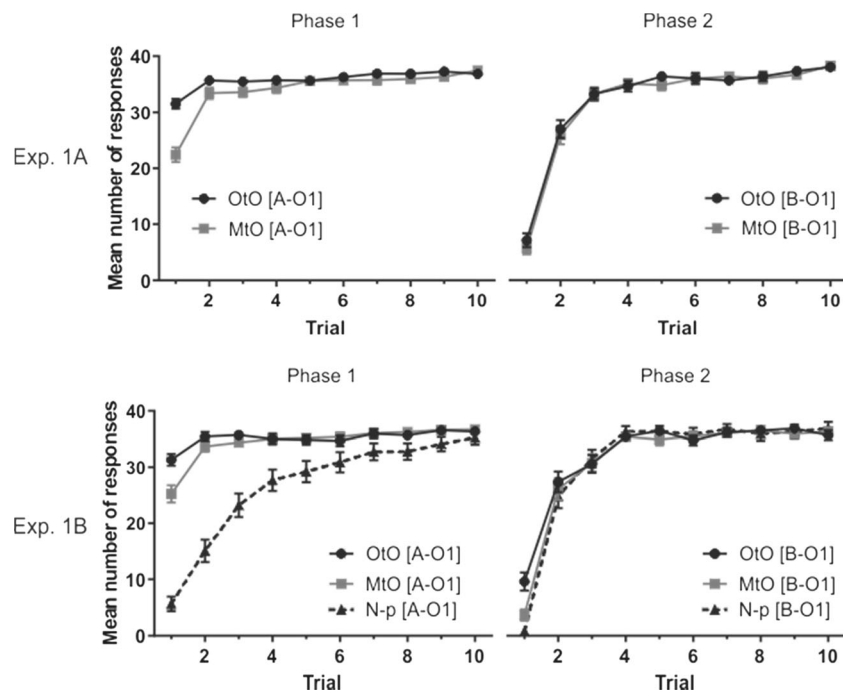
For all experiments (Exps. 1A, 1B, and 2), the level of significance was set at  $\alpha = 0.05$ . In addition, homoscedasticity (Levene’s test) and sphericity (Mauchly’s test) were tested, and the number of degrees of freedom was corrected when necessary (by using Student’s *t* tests for unequal variances or Greenhouse–Geisser corrections). The analysis of variance (ANOVA) model used was a full factorial model using a Type III sums of squares.

We excluded from the analyses the data from participants who did not learn the contingencies programmed or did not understand the instructions. First, we excluded the participants who responded more to the incorrect outcomes than to the correct outcome in any of the two last A→O<sub>1</sub> training trials from Phase 1 (Trials 9 and 10). Second, previous experiments ran in our laboratory showed that a proportion of participants do not follow the instructions provided and do not repeatedly

press the response keys, even when they already know the correct response. As a consequence, the inclusion of these participants leads the response distribution to be considerably skewed, reducing the statistical power of the measure (see also Luque et al., 2009). Hence, participants that did not press a response option key at least ten times on at least one training trial were excluded from further analysis. The application of these criteria led to the exclusion of one participant (<1%) in Experiment 1A and ten participants in Experiment 1B (5.7%). The final sample of Experiment 1A included 80 and 82 participants in the OtO and MtO groups, respectively; the sample in Experiment 1B consisted of 57, 58, and 50 participants in the OtO, MtO, and no-pretraining groups, respectively.

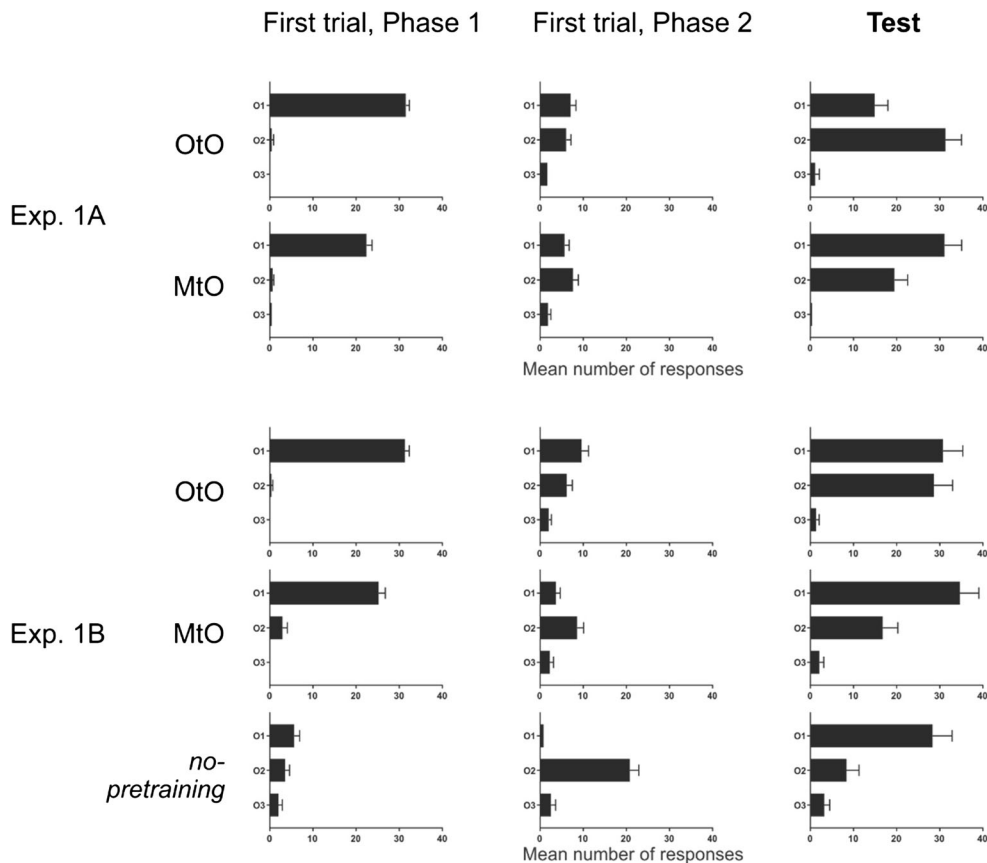
Figure 2 displays the mean numbers of correct responses on each trial for the A→O<sub>1</sub> and B→O<sub>1</sub> trial types (i.e., excluding filler cues). In both experiments, performance appears to be better in the OtO groups, at least at the beginning of Phase 1. In Experiment 1B, the figure shows a clear difference in performance between the no-pretraining and the other groups, particularly in Phase 1. These differences become more evident when we depict the main responses to the three response options on the first learning trial of Phase 1 (Fig. 3).

To confirm these impressions, we first conducted an ANOVA on the numbers of O<sub>1</sub> responses in A→O<sub>1</sub> trials during the first training phase for each experiment. These analyses included Trial as a repeated measures factor (ten levels) and the various Pretraining conditions (OtO vs. MtO for Exp. 1A, and OtO vs. MtO vs. no-pretraining for Exp. 1B) as a between-subjects factor. These analyses yielded significant main effects of Trial in Experiment 1A,  $F(5.27, 859.42) = 57.20, p < .001, \eta_p^2 = .26$ , and Experiment 1B,  $F(4.43, 500.22) = 39.96, p < .001, \eta_p^2 = .24$ , a significant main effect of Pretraining in Experiment 1A,  $F(1, 160) = 6.63, p < .05, \eta_p^2 = .04$ , but not in Experiment 1B,  $F(1, 113) < 1$ ; and significant



**Fig. 2** Learning curves for Experiments 1A and 1B. The panels show the acquisition curves for the correct response (O1) during Phase 1 (left) and Phase 2 (right). Therefore, this figure shows the mean numbers of anticipatory responses to O1, given the cue A (Phase 1) or the cue B (Phase 2).

The data from Experiments 1A and 1B are shown in the top and bottom rows, respectively. Lines denote the different experimental conditions: OtO, one-to-one; MtO, many-to-one; N-p, no-pretraining (see Table 2). Error bars represent the standard errors of the means



**Fig. 3** Mean responses on the first learning trial for each training phase and test phase for Experiments 1A and 1B. Responses for each of the three response options are shown (O1, O2, O3). The left and right panels

depict the responses given to cue A, whereas the central panels depict the responses to cue B. Error bars represent the standard errors of the means

Trial  $\times$  Pretraining interactions in both experiments: Experiment 1A,  $F(5.27, 859.42) = 11.95$ ,  $p < .001$ ,  $\eta_p^2 = .07$ ; Experiment 1B,  $F(9, 1017) = 6.42$ ,  $p < .001$ ,  $\eta_p^2 = .05$ .

Figure 3 suggests that these results, particularly the interaction, are due to a higher rate of O1 responses in the OtO groups than in the MtO groups on the first trial of Phase 1. We conducted pairwise comparisons (Student's  $t$  tests) on O1 responses in the first cue-A trial of Phase 1. For Experiment 1A, this test showed significant differences between OtO and MtO (OtO > MtO),  $t(136.93) = 6.01$ ,  $p < .001$ ,  $d = 0.94$ . For the data of Experiment 1B these analyses included comparisons between the three levels of the variable pretraining (OtO, MtO, no-pretraining). Again, OtO and MtO groups differed on the first cue-A trial (OtO > MtO),  $t(98.81) = 3.30$ ,  $p < .005$ ,  $d = 0.62$ . Participants in the no-pretraining group elicited fewer anticipatory responses to O1 than did the two pretraining groups (see Fig. 3; all  $ps < .001$ ). These results confirm the differences between OtO and MtO found in Experiment 1A, and also indicate that participants in both of the two groups with pretraining started Phase 1 with a higher number of O1 responses than the no-pretraining group.

The good performance of the groups with pretraining as compared with the no-pretraining group at the beginning of Phase 1 could be attributed to a lack of familiarity with the task for the no-pretraining group. However, it is also possible that this result was a consequence of inference processing. This result is compatible with a bias to select the response option that has not been used in the last learning trials (during the entire pretraining phase in this case) when a nonexpected cue is presented and therefore the correct response is not clear (cue A in this case). This bias would explain the high number of O1 responses to the first appearance of cue A in OtO and MtO groups.

Importantly, the OtO and MtO groups also differed from each other at the beginning of Phase 1, with more O1 responses in the OtO than in the MtO group. This difference cannot be attributed to a general bias toward recently unused response options, since both conditions had the same outcomes presented during pretraining. On the other hand, the difference between OtO and MtO groups could be explained if our treatment was effective, and more participants in the OtO group assumed the OtO belief, as compared with the MtO group. If the learner believes that cue–outcome relationships have to follow an OtO structure, given the new cue A at the beginning of Phase 1, the necessary conclusion is that none of the recently experienced outcomes can be the correct response. Therefore, the new cue A *must* be related to some other response option—O1 in this case. This inference process might overlap with a more general bias toward recently unused response options, resulting in the observed results.

These findings obtained from the beginning of Phase 1 were not anticipated; however, they can be seen as evidence of the success of our structural belief training. Interestingly,

the application of this idea to the beginning of Phase 2 would predict an increased number of anticipatory responses for O2 given the new cue B (since only O1 and O3 were used during Phase 1). Thus, for Phase 2 data, we analyzed responses to O1 (correct responses) and also to O2. A visual inspection of the progression of O1 responses across trials suggest that there were no differences between the conditions in Experiment 1A; however, in Experiment 1B a trend was observed on the first trial (Fig. 2). To confirm these impressions, we conducted two ANOVAs on the numbers of correct responses (O1) in B→O1 trials during the second training phase, one for each experiment. These analyses included Trial as a repeated measures factor (ten levels) and the different Pretraining conditions as a between-subjects factor (OtO vs. MtO for Exp. 1A, OtO vs. MtO vs. no-pretraining for Exp. 1B). These analyses yielded significant main effects of Trial in both Experiment 1A,  $F(5.44, 870.78) = 256.46$ ,  $p < .001$ ,  $\eta_p^2 = .62$ , and Experiment 1B,  $F(4.93, 799.08) = 285.97$ ,  $p < .001$ ,  $\eta_p^2 = .64$ , and a significant Trial  $\times$  Pretraining interaction for Experiment 1B,  $F(18, 1458) = 2.33$ ,  $p < .005$ ,  $\eta_p^2 = .03$ ; this interaction was not significant in Experiment 1A,  $F < 1$ . An inspection of the learning curves from Experiment 1B suggests that the Trial  $\times$  Pretraining interaction is due to more anticipatory O1 responses in the OtO group, as compared with the other groups, and more correct responses in the MtO than in the no-pretraining group.  $T$ -test pairwise comparisons conducted on the data from the first trial confirmed these impressions; more correct responses were made in the OtO than in the MtO group,  $t(94.67) = 3.13$ ,  $p < .005$ ,  $d = 0.58$ , and than in the no-pretraining group,  $t(66.28) = 5.24$ ,  $p < .001$ ,  $d = 1.02$ . Responses in the MtO and no-pretraining groups were also significantly different (MtO > no - pretraining),  $t(81.67) = 2.53$ ,  $p < .05$ ,  $d = 0.49$ .

Anticipatory responses to O2 during the second phase were also analyzed in the same way as the O1 responses. Thus, two Trial  $\times$  Pretraining analyses were conducted, one for each experiment. These analyses yielded main effects of Trial, Experiment 1A,  $F(9, 1440) = 35.78$ ,  $p < .001$ ,  $\eta_p^2 = .18$ ; Experiment 1B,  $F(9, 1458) = 93.77$ ,  $p < .001$ ,  $\eta_p^2 = .37$ . The main effect of pretraining, and the Trial  $\times$  Pretraining interaction were found to be significant only in Experiment 1B, Pretraining  $F(9, 1458) = 93.77$ ,  $p < .001$ ,  $\eta_p^2 = .37$ , Trial  $\times$  Pretraining  $F(2, 162) = 15.32$ ,  $p < .001$ ,  $\eta_p^2 = .16$  (in Exp. 1A, lower  $p$  value = .143). The data from Experiment 1B suggest that the Pretraining main effect and the Trial  $\times$  Pretraining interaction in both experiments are due to more anticipatory responses to O2 in the no-pretraining group than in the OtO and MtO groups, this difference being clear on the first trial (see Fig. 3). We analyzed possible differences between groups on the first trial of the second phase using pairwise  $t$  - tests. Anticipatory responses to O2 were significantly more frequent in the no-pretraining group compared with the OtO group,  $t(83.81) = 5.95$ ,  $p < .001$ ,  $d = 1.15$ , and the MtO group,

$t(96.63) = 4.74, p < .001, d = 0.91$ , whereas the OtO and MtO groups did not differ,  $t(113) = 1.12, p = .234, d = 0.21$ .

In summary, the results from Phase 2 are not as readily interpreted as those from Phase 1. Some of the effects reported (e.g., the Trial  $\times$  Pretraining interaction in the B $\rightarrow$ O1 responses) were only significant in one of the experiments. We will return to this issue in the Discussion section (see below).

Regarding the analysis of the test phase, a visual inspection of Fig. 3 suggests that responses to O1 differed between conditions, especially in Experiment 1A. A clear trend is observed in O2 responses in both experiments. To confirm these impressions, we conducted a 3 (Outcome: O1, O2, & O3)  $\times$  2 (Pretraining: OtO vs. MtO) ANOVA for Experiment 1A, and a 3 (Outcome: O1, O2, & O3)  $\times$  3 (Pretraining: OtO vs. MtO vs. no - pretraining) ANOVA for Experiment 1B. Both tests yielded a significant main effect of Outcome, Experiment 1A,  $F(1.26, 201.67) = 35.25, p < .001, \eta_p^2 = .18$ ; Experiment 1B,  $F(1.27, 205.65) = 40.71, p < .001, \eta_p^2 = .20$ , and a significant Outcome  $\times$  Pretraining interaction, Experiment 1A,  $F(2, 320) = 9.53, p < .001, \eta_p^2 = .06$ ; Experiment 1B,  $F(4, 324) = 2.48, p = .044, \eta_p^2 = .03$ . The main effect of Pretraining was only significant in Experiment 1B,  $F(2, 162) = 8.77, p < .001, \eta_p^2 = .10$  (Exp. 1A,  $F < 1$ ).

In Experiment 1A, the important Outcome  $\times$  Pretraining interaction was driven by fewer responses anticipating O1 and more responses anticipating O2 in the OtO group in comparison with the MtO group. This interpretation was confirmed by planned pairwise comparisons, which yielded a significant effect of Pretraining on O1 responses,  $t(151.42) = 3.24, p = .001, d = 0.51$ ; a significant effect on O2 responses,  $t(154.12) = 2.45, p = .015, d = 0.38$ ; and no significant effect on O3 responses,  $t < 1$ . We also conducted pairwise comparisons between the response options (O1, O2, O3) within each group (OtO, MtO). In the OtO group, the pattern of responses can be summarized as O2 > O1 > O3: O2 > O1,  $t(79) = 2.77, p = .007, d_z = 0.31$ ; O1 > O3,  $t(79) = 4.26, p < .001, d_z = 0.48$ . In the MtO group, the pattern was O1 = O2 > O3: O1 = O2,  $t(81) = 1.87, p = .066, d_z = 0.21$ ; O1 > O3,  $t(81) = 7.70, p < .001, d_z = 0.85$ ; O2 > O3,  $t(81) = 6.13, p < .001, d_z = 0.68$ .

In Experiment 1B, only responses anticipating O2 differed across groups, these being most frequent in the OtO condition, with no differences between MtO and no - pretraining: OtO versus MtO,  $t(113) = 2.12, p = .036, d = 0.39$ ; OtO versus no - pretraining,  $t(105) = 3.76, p < .001, d = 0.73$ ; MtO versus no - pretraining,  $t(106) = 1.79, p = .075, d = 0.34$ ; other  $ps > .18$ . Pairwise comparisons between response options (O1, O2, O3) within the OtO group yielded an O1 = O2 > O3 pattern: O1 = O2,  $t(56) = 0.25, p = .801, d_z = 0.03$ ; O1 > O3,  $t(56) = 6.21, p < .001, d_z = 0.82$ ; O2 > O3,  $t(56) = 6.04, p < .001, d_z = 0.80$ . In the MtO group the pattern was different, O1 > O2 > O3: O1 > O2,  $t(57) = 2.54, p = .014, d_z = 0.33$ ; O1 > O3,  $t(57) = 6.90, p < .001, d_z = 0.91$ ; O2 > O3,  $t(57) = 3.87, p < .001, d_z = 0.51$ . Finally, an O1 > O2 = O3 pattern was obtained in the no-

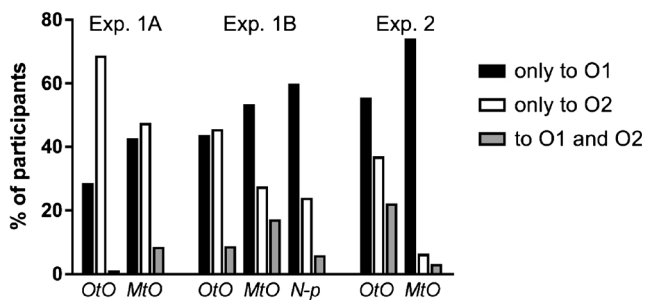
pretraining group: O1 > O2,  $t(49) = 3.24, p = .002, d_z = 0.46$ ; O1 > O3,  $t(49) = 5.01, p < .001, d_z = 0.71$ ; O2 = O3,  $t(49) = 1.56, p = .126, d_z = 0.22$ .

Finally, we also analyzed possible differences in the numbers of participants who made O1 responses only, O2 responses only, or both. These data are shown in Fig. 4. Since only a very small number of participants responded to both response options, we excluded this condition from the following analyses. We compared the relative frequencies of “only O1” and “only O2” response strategies among experimental conditions by using chi-square tests. Regarding Experiment 1A, the number of participants who made only O1 responses was the same in the OtO group as in the MtO group,  $\chi^2(1) = 3.42, p = .073$ , though numerically the difference favored the MtO group (i.e., MtO = 35, OtO = 23). In the same vein, more participants chose O2 as the only possible outcome in the OtO than in the MtO group (OtO = 55, MtO = 39),  $\chi^2(1) = 7.47, p = .007$ . A similar pattern was found in Experiment 1B. More participants made only O1 responses in the MtO than in the OtO group (OtO = 25, MtO = 31); the number of these responses in the no-pretraining group fell between the numbers in the other two groups (no-pretraining = 30). However, this difference was not significant,  $\chi^2(2) = 2.84, p = .241$ . The groups differed regarding O2 responses,  $\chi^2(2) = 6.73, p = .035$ . Specifically, these differences were due to more participants choosing O2 as the only possible outcome in the OtO than in the MtO group (OtO = 26, MtO = 16), though the difference was only marginally significant,  $\chi^2(1) = 4.03, p = .054$ , and more in the OtO than in the no-pretraining group,  $\chi^2(1) = 5.43, p = .026$  (no-pretraining = 12). There were no differences between the no-pretraining and MtO groups,  $\chi^2(1) < 1$ . In summary, more participants responded by uniquely anticipating O2 (and not O1) in the OtO than in the MtO group.

## Discussion

The results from Experiments 1A and 1B partially support the structural-belief account of RIBC described in the introduction. A prediction from this hypothesis was that fewer O1 responses should be observed in the OtO than in the MtO group on the test trial. As predicted, in both experiments, O1 responses were numerically more frequent in the MtO than in the OtO group, although this difference only reached significant levels in Experiment 1A. However, we found additional results supporting our hypothesis. First, incorrect responses at test were biased toward O2 rather than randomly distributed between the O2 and O3 response options (Fig. 3). This bias could be explained by using a simple rule such as *if a nonexpected cue appears, then select the response option not used in the previous phase*. However, in both experiments the number of O2 responses was greater in the OtO group than in the other groups, despite the fact that Outcome 2 had never





**Fig. 4** Response distribution at test. The figure shows the percentage of participants responding only to O1, O2, or to O1 and O2 for Experiments 1A, 1B, and 2. Percentages are shown for each experimental condition: OtO, one-to-one; MtO, many-to-one; N-p, no-pretraining. For Experiment 2, only the data from the experimental condition (OtO and MtO) are shown

been paired with any cue during the previous phase in all groups. Our structural-belief hypothesis can give a more complete account of the pattern of results regarding the O2 responses at test. Given cue A at test, those participants with a strong OtO belief would show a greater tendency to disregard any of the recently used response options, O1 and O3. These response options could not be the correct because other cues had been paired with them relatively recently. If these two response options were discarded, the only one remaining was O2, which was the response option chosen by many participants in the OtO groups (at least more often than by participants in the MtO groups). Further, a similar pattern was found in the first A→O1 trial of Phase 1. Given the new cue, A, participants responded more frequently to the response option not used during the previous phase—that is, O1. Importantly, this bias was more pronounced in the OtO groups than in the MtO groups. And finally, more participants chose only the O2 response option at test in the OtO than in the MtO group. Again, this result can be explained in terms of an inference process guided by the participants' structural belief.

In some respects, there are limitations to the evidence provided by Experiments 1A and 1B. One concern that should be addressed is that the effect of pretraining on the expression of the supposedly interfered-with association, A→O1, was significant only in Experiment 1A. Additionally, it is difficult to assess the effect of our pretraining manipulation on the RIBC effect itself, because the assessment of such an effect must involve the inclusion of a control condition in which the target association, A→O1, did not receive an interference treatment (see Table 1). Finally, the reason for changing the contextual cue between Phase 1 and Phase 2 was to ensure consistency with the procedure used in previous studies in our laboratory that had been successful in observing the RIBC effect. However, our structural-belief hypothesis does not attribute any crucial role to this procedural aspect in the explanation of the RIBC effect. Indeed, the inclusion of a contextual change between Phase 1 and Phase 2 might have affected the responses on the first trial of Phase 2. Consequently, more

compelling evidence in support of our hypothesis could be found by omitting this contextual change in a new experiment.

## Experiment 2

As in the previous experiments, the main aim of Experiment 2 was to clarify further whether participants' structural beliefs (i.e., OtO or MtO) can modulate the RIBC effect. In addition, to overcome the limitations stated previously, Experiment 2 included control conditions designed to take these problems into account, and we assessed the RIBC effect in a procedure in which no context change was programmed to occur in the transition between phases. In this experiment, OtO and MtO structural beliefs were promoted through the very same training trials that had formed part of Phases 1 and 2. Thus, no pretraining was used in Experiment 2.

## Method

**Participants and apparatus** A total of 122 participants from the University of Málaga took part in Experiment 2 for course credits. The number of participants was determined by the same stopping rule that we described for Experiments 1A and 1B. Experiment 2 was conducted in the same room, and all apparatus was the same as in Experiments 1A and 1B.

**Design** Experiment 2 included four different experimental groups in a 2 × 2 between-subjects design: The independent variables were Structural Training, OtO and MtO groups, and Interference, which comprised the experimental and control conditions. In Experiment 2, Structural Training was implemented during the two phases of a standard RIBC design (see Table 3). As is shown in Table 3, the design includes the usual A→O1 (Phase 1); B→O1 (Phase 2) training for the experimental condition. The control condition also includes the standard A→O1 (Phase 1); B→O2 (Phase 2) training procedure that is routinely used in a typical RIBC design (Tables 1 and 3). In addition to these relationships, participants in the MtO groups learned two additional relationships, in which two independent cues separately predicted the same outcome in Phases 1 and 2 (C→O3, D→O3), but participants in the OtO groups learned about one additional cue–outcome relationship (C→O3) during Phases 1 and 2 (Table 3).

**Procedure** The procedure was similar to that used in Experiments 1A and 1B (Fig. 1), except that the color of the context was determined at random at the beginning of the experiment for each participant. This context color was kept constant during the experiment. Four colors were used for the cues: magenta, yellow, light green, and cyan (counterbalanced across participants following a Latin square design).

**Table 3** Design of Experiment 2

Group	Phase 1	Phase 2	Test
Experimental–MtO	A→O1 (x10)	B→O1 (x10)	A?
	C→O3 (x10)	C→O3 (x10)	
	D→O3 (x10)	D→O3 (x10)	
Control–MtO	A→O1 (x10)	B→O2 (x10)	A?
	C→O3 (x10)	C→O3 (x10)	
	D→O3 (x10)	D→O3 (x10)	
Experimental–OtO	A→O1 (x10)	B→O1 (x10)	A?
	C→O3 (x20)	C→O3 (x20)	
Control–OtO	A→O1 (x10)	B→O2 (x10)	A?
	C→O3 (x20)	C→O3 (x20)	

The letters denote cues (cue A, B, C, D), and the letter “O” followed by a number is used to denote an outcome (Outcome 1, O1; Outcome 2, O2; Outcome 3, O3). The number of repetitions of each type of trial is shown between parentheses. For instance, A→O1 (x10), indicates that cue A followed by the Outcome 1 was presented ten times. Question marks indicate that this cue was presented at test (and therefore no outcome was presented).

## Results

The application of the same exclusion criteria as in Experiments 1A and 1B led to the exclusion of ten participants (~8%). The final sample of Experiment 2 included 31 participants in the experimental–OtO group (one exclusion), 29 in the control–OtO group (one exclusion), 27 in the experimental–MtO group (four exclusions), and 26 in the control–MtO group (four exclusions).

Figure 5 displays the mean numbers of correct responses on each trial for the A→O1 and B→O1 trial types in the experimental groups and for A→O1 and B→O2 in the control groups. Upon inspection of these figures, it is apparent that there were drastic differences between conditions, particularly at the beginning of Phase 2 (Fig. 6).

To confirm these impressions, we first conducted an ANOVA on the numbers of correct response (O1) in A→O1 trials during the first training phase. This analysis included Trial as a repeated - measures factor (Trial 1–10), and Structural Training (OtO, MtO) and Interference (experimental, control) as between-subjects factors. This analysis yielded significant main effects of Trial,  $F(6.17, 672.07) = 130.41, p < .001, \eta_p^2 = .55$ , and Structural Training,  $F(1, 109) = 10.70, p = .001, \eta_p^2 = .09$ . The ANOVA also revealed a significant Trial  $\times$  Structural Training interaction,  $F(9, 981) = 1.91, p = .047, \eta_p^2 = .02$  (all remaining effects were not significant  $F_s < 1.2$ ). An inspection of Fig. 5 suggests that the Structural Training main effect was due to faster learning in the OtO groups. The Trial  $\times$  Structural Training interaction is likely to reflect the fact that the Structural Training effect was particularly evident on certain trials (between Trials 4 and 6; see Fig. 5).

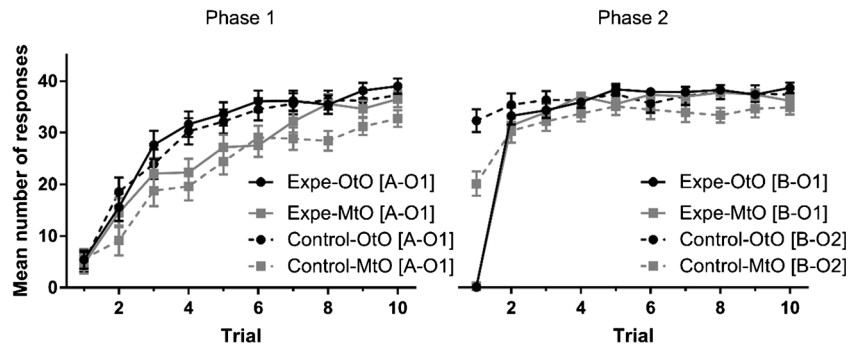
Regarding the results from Phase 2, visual inspection of Figs. 5 and 6 suggests that the Phase 1 training had a relevant impact on participants’ patterns of responding during Phase 2, particularly on the first trial. To confirm these impressions, we

analyzed O1 and O2 responses to cue B trials in two separate sets of analyses. First, we conducted a 10 (Trial)  $\times$  2 (Structural Training)  $\times$  2 (Interference) ANOVA on the numbers of O1 responses in B trials during the second training phase. This analysis yielded significant main effects of Trial,  $F(4.63, 505.22) = 217.36, p < .001, \eta_p^2 = .67$ , and Interference,  $F(1, 109) = 1196.93, p < .001, \eta_p^2 = .98$ . The Trial  $\times$  Interference interaction was also significant,  $F(9, 981) = 254.41, p < .001, \eta_p^2 = .70$ , showing that the Interference main effect was stronger at the beginning of Phase 2. All remaining effects were not significant ( $F_s < 1.1$ ). Next we repeated the same analysis using the numbers of O2 responses on B trials. This ANOVA yielded significant main effects of Trial,  $F(5.33, 581.01) = 57.83, p < .001, \eta_p^2 = .35$ , and of both Interference,  $F(1, 109) = 1116.56, p < .001, \eta_p^2 = .91$ , and Structural Training (OtO > MtO),  $F(1, 109) = 5.17, p = .025, \eta_p^2 = .04$ . We again found a significant Trial  $\times$  Interference interaction,  $F(9, 981) = 186.88, p < .001, \eta_p^2 = .63$ . More importantly, in this analysis we also found a significant Trial  $\times$  Structural Training interaction,  $F(9, 981) = 5.72, p < .001, \eta_p^2 = .05$  (all remaining effects were not significant; largest  $F = 2.88, p = .092$ ).

Inspection of Figs. 5 and 6 suggests that the Trial  $\times$  Structural Training interaction was due to more O2 responses in the OtO groups, particularly on the first trial of Phase 2. Analysis of responses on this first trial supports this impression: More O2 (incorrect) responses were made in the experimental–OtO than in the experimental–MtO group,  $t(37.42) = 2.95, p = .006, d = 0.78$ . The same pattern was found in the control groups: More O2 (correct) responses were made in the control–OtO than in the control–MtO group,  $t(36.14) = 3.61, p = .001, d = 0.97$ .

Regarding the analysis of the test phase, visual inspection of Fig. 6 suggests that the RIBC effect (fewer O1 responses in the experimental than in control conditions) was smaller in the MtO condition. In addition, it appears that O1 responses were more frequent in the experimental–MtO than in the experimental–OtO group. Planned analyses included a 3 (Outcome: O1, O2, O3)  $\times$  2 (Structural Training)  $\times$  2 (Interference) ANOVA on participants’ responses at test, which yielded a significant main effect of Outcome,  $F(1.26, 136.92) = 258.80, p < .001, \eta_p^2 = .70$ ; a significant main effect of Interference (experimental < control),  $F(1, 109) = 5.78, p = .018, \eta_p^2 = .50$ ; and a significant Outcome  $\times$  Interference interaction,  $F(2, 218) = 14.61, p < .001, \eta_p^2 = .12$  (this effect was likely also a consequence of the RIBC effect, which should not be evident for O3 responses—hence, the interaction with outcome). Importantly, the Outcome  $\times$  Structural Training  $\times$  Interference interaction was significant,  $F(2, 218) = 5.40, p = .005, \eta_p^2 = .05$ .

Next we tested whether the RIBC effect (experimental < control for O1 responses) was larger in the OtO than in the MtO condition. For this, we conducted a 2 (Structural Training)  $\times$  2 (Interference) ANOVA on the O1 responses at



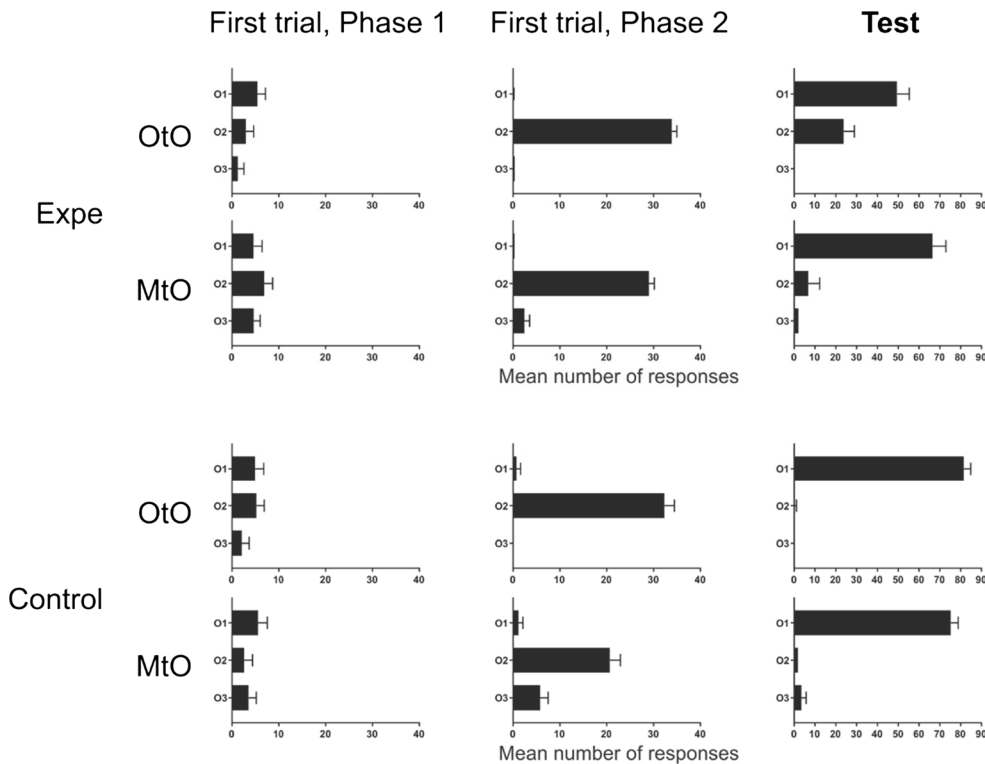
**Fig. 5** Learning curve for Experiment 2. The figure shows the acquisition curves for the correct response (O1) during Phase 1 (left) and Phase 2 (right) (O1 for the experimental condition and O2 for the control condition). The lines denote the different experimental conditions:

Black and gray lines represent the one-to-one and the many-to-one conditions, respectively. Solid and dashed lines represent the experimental and the control conditions, respectively. Error bars represent the standard errors of the means

test. This analysis yielded a main effect of Interference (experimental < control),  $F(1, 109) = 15.94, p < .001, \eta_p^2 = .13$ , which indicates that the RIBC effect was found, and importantly, a significant Structural Training  $\times$  Interference interaction,  $F(1, 109) = 5.24, p = .024, \eta_p^2 = .05$ . Pairwise comparisons showed that the standard RIBC effect was significant in the OtO condition,  $t(32) = 4.68, p < .001, d = 1.21$ , but not in the MtO condition,  $t(51) = 1.18, p = .242, d = 0.32$ . There were also marginally significant differences between the experimental–OtO and experimental–MtO groups,  $t(54.62) = 1.99, p = .051, d = 0.50$ ; an equivalent contrast for control

conditions was not significant, control–OtO versus control–MtO,  $t(28) = 1.20, p = .239, d = 0.21$ .

Similar analyses were conducted on the O2 responses at test. A 2 (Structural Training)  $\times$  2 (Interference) ANOVA yielded a marginally significant main effect of Structural Training (OtO > MtO),  $F(1, 109) = 3.73, p = .056, \eta_p^2 = .03$ ; a main effect of Interference (experimental > control),  $F(1, 109) = 8.11, p < .001, \eta_p^2 = .18$ ; and a significant Structural Training  $\times$  Interference interaction,  $F(1, 109) = 5.72, p = .019, \eta_p^2 = .05$ . Pairwise comparisons showed that O2 responses were significantly more frequent in the OtO than



**Fig. 6** Mean responses on the first learning trial for each training phase and test phase in Experiment 2. Responses to each of the three response options are shown (O1, O2, O3). Error bars represent the standard errors of the means

in the MtO conditions within the experimental condition,  $t(49.45) = 2.34, p = .023, d = 0.62$ , but not within the control condition,  $t(25) = 1.12, p = .275, d = 0.30$ . O2 responses were more frequent in the experimental than in the control condition within the OtO group,  $t(30) = 3.86, p = .001, d = 1.00$ , but not within the MtO group,  $t(34.82) = 1.19, p = .243, d = 0.33$ .

We also conducted pairwise comparisons between the response options (O1, O2, O3) within each RIBC group (experimental–OtO and experimental–MtO). In both groups, O1 responses were more frequent than O2 responses [experimental–OtO,  $t(30) = 2.06, p < .05, d_z = 0.37$ ; experimental–MtO,  $t(26) = 6.87, p < .001, d_z = 1.32$ ] and O3 responses [experimental–OtO,  $t(30) = 7.25, p < .001, d_z = 1.32$ ; experimental–MtO,  $t(26) = 11.16, p < .001, d_z = 2.15$ ]. O2 responses were more frequent than O3 responses in the experimental–OtO group,  $t(30) = 3.81, p = .001, d_z = 0.68$ , but not in the experimental–MtO group,  $t(26) = 1.13, p > .02, d_z = 0.22$ .

Finally, we also analyzed differences in the numbers of participants responding uniquely to O1, to O2, or to both response options (Fig. 4). This analysis was conducted only within the experimental condition, since almost all participants in the control groups made O1 responses. We compared the relative frequencies of these response strategies between experimental conditions by using chi-square tests. This analysis showed that (1) the numbers of participants who only made O1 responses differed between the OtO and MtO groups (OtO = 15, MtO = 23),  $\chi^2(1) = 8.65, p = .005$ ; (2) the numbers of participants who only made O2 responses differed between the OtO and the MtO groups (OtO = 10, MtO = 2),  $\chi^2(1) = 5.43, p < .05$ ; and (3) the numbers of participants who made both types of responses did not differ between the OtO and MtO groups (OtO = 6, MtO = 1),  $\chi^2(1) = 3.33, p > .1$ .

## Discussion

Experiment 2 further investigated the influence of training in structural beliefs on the RIBC effect. The pattern of results provided strong evidence supporting the structural belief hypothesis described in the introduction. First, the basic RIBC effect (i.e., fewer O1 responses during test in the experimental than in the control group) was greater in the OtO than in the MtO condition. Second, the number of O1 responses at test was greater in the experimental–MtO than in the experimental–OtO group (although this difference was marginally significant,  $p = .051$ ). Third, the frequencies of O2 and O3 responses at test were not randomly distributed, as indicated by the fact that O2 responses were much more frequent than O3 responses. Importantly, O2 responses were more frequent in the experimental–OtO than in the experimental–MtO group. And fourth, more participants uniquely made O1 responses in the experimental–MtO than in the experimental–OtO group, but the opposite result was found in the case of O2 responses: More participants uniquely chose the O2 response option in

the experimental–OtO group than in the experimental–MtO group.

The present complex pattern of results is both highly informative and parsimoniously consistent with our structural-belief hypothesis. According to this account, participants in the experimental–OtO group inferred that the A→O1 relationship learned during Phase 1 was no longer valid after receiving B→O1 training, which in turn lowered the number of O1 responses at test. Moreover, our results provide compelling evidence to suggest that the participants in the experimental–OtO group were applying a strategy at test that was also predicted by our hypothesis: When cue A was unexpectedly presented at test, participants could discount O1 and O3 as possible correct outcomes, since (1) these outcomes had been recently predicted by other cues during Phase 2, and (2) according to the OtO belief, participants expected each outcome to be related to only one cue. Consequently, participants inferred that the correct response option was O2, since this was the only free outcome, and therefore the only way of maintaining coherence with the OtO belief. Importantly, the differences between the OtO and MtO groups provide converging support for the ideas that this response strategy was based on an OtO belief and that the belief training provided during the training phases of the learning task actually modulated participants' expectations about the cue–outcome relationship structure.

Further independent evidence supporting the idea that the structural-belief training was successful comes from the first trial of Phase 2. A bias to choose the unused response option on the first B trial was evident. Importantly, as in the test phase, this bias was affected by the structural-belief training. As can be seen in Fig. 6, O2 responses were more frequent in the OtO than in the MtO group on the first B trial of Phase 2. This result may be taken as evidence for the successful promotion of structural beliefs, as, indeed, more participants were expecting a one-to-one structure in the OtO group.

Experiment 2 did not include any contextual change in the transition between Phases 1 and 2; this procedural feature of Experiments 1A and 1B was not necessary to observe the RIBC effect and to study the effect of our belief training. Moreover, removing the context change may have helped us obtain clearer results, since the mere change in context could have made participants disregard what had been learned in the presence of a given contextual cue when learning in the presence of a different contextual cue.

## Meta-analysis of Experiments 1A, 1B, and 2

In Experiments 1A, 1B, and 2 (experimental groups), we assessed the effects of training in structural beliefs (OtO vs. MtO) on the number of O1 responses at test. In all experiments we found (numerically) fewer responses when the

participants were trained in an OtO structural belief. However, whereas this effect was statistically significant in Experiment 1A, it was only marginally significant in Experiment 2, and not significant in Experiment 1B. To clarify whether this was a genuine effect, we performed a meta-analysis on the relevant data from the three experiments. It might be claimed that Experiment 1A was, for some unknown reason, different from the others in some critical aspect that made its results qualitatively different, and therefore that the meta-analysis is invalid. To address this possibility, we also conducted an analysis of heterogeneity in order to determine whether there were systematic differences across studies.

The mean numbers of O1 responses at test were submitted to a random-effect meta-analysis, using the Review Manager 5.3 software (Cochrane Collaboration, 2014). The data included in the meta-analysis and a forest plot are shown in Fig. 7. The meta-analysis showed that the effect was clearly significant when the data from all three experiments are considered,  $z = 2.90$ ,  $p < .005$ . Importantly, the analysis of heterogeneity showed that there were no systematic differences across studies,  $I^2 = 25\%$ ,  $\chi^2(2) = 2.66$ ,  $p = .26$ .

## General discussion

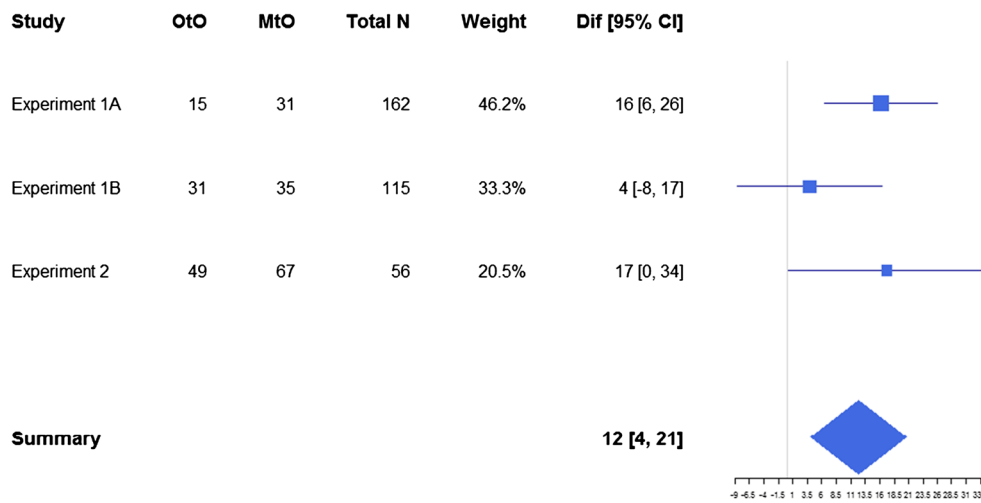
The results from Experiments 1A, 1B, and 2 strongly suggest that participants in a contingency learning experiment generate beliefs about the cue–outcome structure that better accommodate the perceived data, and that such beliefs modulate (Exps. 1A and 1B) or even conditionalize (Exp. 2) the RIBC effect. Moreover, these structural beliefs and expectations lead to inferences about the most probable correct response given an uncertain situation. For instance, in our RIBC experiments, when the cue A was unexpectedly presented at test, many participants responded only to the O2 response option. This occurred in spite of the fact that cue A had *never* been paired with O2; in fact, it had been consistently paired with O1 during the whole of Phase 1. This inference can be explained as the consequence of a discount process, motivated by the belief that the cues and the outcomes relate to each other following an OtO structure (i.e., each cue relates to only one outcome and each outcome relates to only one cue). Therefore, when cue A was presented at test, the participants could discount O1 and O3 as possible correct outcomes, since these outcomes had recently been predicted by other cues during Phase 2 (see Tables 2 and 3), prompting the conclusion that the correct outcome had to be O2. Importantly, this effect (among others) was affected by our structural belief manipulation. This discount strategy was more evident in OtO than in MtO groups. This result indicates that the discount strategy was not merely a general bias toward unused response options.

This result also suggests that we have effectively modified participants' structural beliefs in these experiments. The

strategy for doing so was to show additional evidence consistent (OtO conditions) or not (MtO conditions) with the OtO belief. A number of differences between the OtO and MtO conditions across the three experiments support the notion that this structural-belief training was effective. In addition to the aforementioned results during the test phase, this account receives additional support from the results regarding the first A→O1 trial of Phase 1 in Experiments 1A and 1B. On that trial, participants had to decide for the first time about what the correct outcome was for cue A. Although cue A had never been presented before, the number of correct O1 responses was unusually high. Importantly, this number of O1 responses was even higher for the OtO groups. According to the inference process advocated here, when cue A was presented for the first time on that trial, O1 was the only outcome that had not been previously associated with other cues during the pretraining phase. Thus, participants in the MtO group could not make any inference regarding cue A on this first trial as they had learned that the same outcome could be related to more than one cue during the pretraining phase. However, participants in the OtO group could expect O1 to be the correct outcome, as this would be the only possible outcome consistent with an OtO structure. This would explain why the number of correct responses on the first A→O1 trial of Phase 1 in the OtO group was larger than in the MtO group. Notably, a similar inference process might explain the result from the first B→O1 trial in Phase 2 from Experiment 2. In this case, an unusually high number of O2 responses were given to the completely new cue, B, in the OtO groups (Fig. 6).

The ways in which participants allocate their responses at test suggest that our manipulation of belief training did affect the number of participants who confidently used the discount strategy described above. More participants chose only the O1 response option in the MtO groups than in the OtO groups, whereas the opposite pattern was found regarding O2 responses: More participants chose only the O2 response option in the OtO than in the MtO groups. This pattern of results was observed in all experiments (Fig. 4), and it was significant in Experiment 2. These results suggest that our manipulation affected the RIBC effect not by simply reducing participants' confidence in the O1 response option, but by increasing the number of participants who chose the alternative O2 response option.

These results have theoretical implications for current accounts of the RIBC effect. In particular, the results from Experiment 2 suggest that the RIBC effect might be potentiated by inferences based on structural beliefs. These inferences would lead participants to reduce the number of responses anticipating O1 at test in the experimental condition, and also to increase them in the control condition, resulting in a strong RIBC effect. This explanation of the effect represents a significant departure from previous accounts. According to the learning deficit and retrieval deficit accounts, the RIBC



**Fig. 7** Meta-analysis of the differences in the numbers of O1 responses at test when comparing the one-to-one and many-to-one conditions. The forest plot (figure at the right) show the standardized main differences

and their 95% confidence intervals for Experiments 1A, 1B, and 2. At the bottom (the diamond), the results of the random-effects meta-analysis are shown

effect occurs either because the A→O1 memory trace becomes weaker during the second phase or because the retrieval of the A→O1 memory is inhibited by the activation of the competing B→O1 memory trace, respectively (Matute & Pineño, 1998a, b; Miller & Escobar, 2002). In the present experiments, however, the expression of the A→O1 memory trace improved when participants learned that cues C and D were independent predictors of Outcome O3. The relationships used to promote a MtO structure expectation (i.e., C→O3 and D→O3) did no share any element (cue or outcome) with the relationships directly involved in the RIBC effect (i.e., A→O1 and B→O1). Consequently, the learning deficit and retrieval deficit accounts cannot explain these results. To explain the effects of our MtO/OtO training manipulation, it is necessary to postulate that participants learn structural aspects of the task that go beyond the specific cues and outcomes involved in the actual training.

The causal model theory account of RIBC also has difficulty in explaining the present findings. According to this account, RIBC should be evident only if cues A and B are described to participants as *effects* from the same *cause*, Outcome O1 (see Cobos et al., 2007; Luque et al., 2008). The learning task used in the present experiments did not describe cues and outcomes as part of any causal scenario in which cues and outcomes were regarded as effects and causes, respectively. Instead, the instructions simply pointed out that the participant’s goal was to earn as many points as possible by choosing on each trial which outcome corresponded with the presented cue. It is therefore not easy to reconcile our results with a causal model theory account of the RIBC effect. Interestingly, the hypothesis advocated here is consistent with previous results favoring causal model theory, particularly if our structural belief account is enriched with the principles of causal model theory. These previous results have shown that

RIBC occurs when cues are described to participants as effects from the same cause but not when cues are described as independent causes of the same effect (Cobos et al., 2007). It might be argued that a *causes-to-effect* scenario should facilitate the assumption of a MtO structural belief, at least to a greater extent than an *effects-to-cause* scenario. The reason for this is that it is easier to conceive a situation in which many independent causes would produce one single effect (i.e., MtO), than another in which different effects are produced by the same cause at different times. The evident similarity between this explanation and that provided by Cobos et al. (2007) and Luque et al. (2008) suggests that the causal model theory account may be considered as an application of our more general structural belief account to the special case of learning tasks framed within causal scenarios.

One of the contributions of the present study is to show that choosing an alternative response option (O2 in our design) could contribute to the RIBC effect when an OtO structure belief is assumed. In this case, the RIBC effect would be promoted by the possibility of choosing an alternative response option at test. In other words, tests with multiple response options should be more sensitive to RIBC effects than one-response-option tests. This idea is consistent with the fact that, to the best of our knowledge, all previous studies reporting null RIBC effects (Cobos et al., 2007; Lipp & Dal Santo, 2002; Luque et al., 2008) are based on the use of one response option. Conversely, experiments using multiple response options have always reported RIBC effects (e.g., Keppel et al., 1971; Luque et al., 2009; Vadillo et al., 2008).

The inference process proposed in our structural belief hypothesis is cognitively demanding, and might contribute to the RIBC effect, particularly when experimental conditions favor the deployment of multiple cognitive resources. For instance, the operation of highly demanding reasoning processes should

be particularly favored in experiments in which only a few deterministic relationships are to be learned and participants are under time pressure to respond during the test (e.g., Escobar et al., 2002; Luque et al., 2008; Luque et al., 2010; Luque et al., 2011; Matute & Pineno, 1998a). It should be noted that this is not the case in the experiments published in the paired-associate learning literature (e.g., Keppel et al., 1971). In these experiments, participants learn many different cue–outcome relationships (e.g., 12 per learning phase in Keppel et al., 1971). Thus, it seems unlikely that inferential reasoning has significantly contributed to these instances of RIBC. Rather, it seems more likely that some sort of purely associative-based memory interference underpins these RIBC effects. Thus, we are not claiming that the inferences and abstract representations envisaged in our structural belief hypothesis are the only possible processes leading to a RIBC effect. This is a very simple behavioral effect, but it might be produced by very different mechanisms that operate concurrently.

Nevertheless, our study opens up the possibility that the RIBC effect studied so far in the HCL literature is entirely due to a reasoning process. For instance, one possible explanation of our results is that learners are fully able to retrieve Outcome 1 from the presence of cue A at test. However, they infer that the A→O1 relationship is no longer valid after having been recently exposed to B→O1 trials. In such a case, the burden of the explanation of the RIBC effect would concentrate on reasoning, leaving memory aside. Another possibility is that the inference that the A→O1 relationship is no longer valid is made early during Phase 2, which in turn would lead either to the weakening of the A–O1 association or to the storage of a new memory of cue A that would interfere with the retrieval of the A–O1 association at test. This latter possibility places part of the explanatory burden on memory-based processes, which requires the assumption that the RIBC effect is the result of the interaction between reasoning and memory. Our results do not allow us to discriminate between these two possibilities, but instead suggest that the common assumption that RIBC is a memory effect may need further empirical support. In particular, future experiments should be conducted to determine the role played by memory and reasoning processes in the RIBC effect.

The present study raises important questions that could be addressed in future experiments. For instance, in further research it might be worth investigating to what extent, and under what circumstances, the RIBC effect can be produced by other mechanisms different to those described here. The results of these future experiments might illuminate whether the RIBC effect found in paired-associate experiments is the consequence of inference processes, or whether they are the consequence of some kind of associative-based memory interference. Variables such as the cognitive load associated with the task, or the number of responses available at test, are likely

to be important factors in elucidating the mechanisms underlying any RIBC effect. It will also be important to study whether certain participants are unaffected by the structural-belief training (as Fig. 4 seems to suggest). A particularly intriguing possibility is that individual differences might be mediating the effects described in the present experiments; one possible factor is that individuals could differ in terms of their cognitive styles (Riding & Rayner, 1998).

In summary, the present results provide the first evidence that acquired cue–outcome structure beliefs can affect the RIBC effect. In particular, we have shown that a particularly hard-to-explain interference effect, in this case the RIBC effect, might be the result of an inference process based on structure beliefs. These experiments highlight the importance of considering abstract structural meta-knowledge in order to understand HCL, particularly when learning takes place in very simple situations.

**Author note** This research was supported by Grant No. SEJ-7896 from Junta de Andalucía and Spanish Ministry of Science and Innovation Grant Nos. PSI2011-24662 and PSI2013-43516-R. The authors thank Tom Beesley for his comments on an earlier version of this manuscript.

## References

- Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, 49, 415–445. doi:10.1016/j.jml.2003.08.006
- Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin*, 114, 80–99. doi:10.1037/0033-2909.114.1.80
- Cobos, P. L., Gutiérrez-Cobo, M. J., Moris, J., & Luque, D. (2017). Dependent measure and time constraints modulate the competition between conflicting feature-based and rule-based generalization processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43, 515–527. doi:10.1037/xlm0000335
- Cobos, P. L., López, F. J., & Luque, D. (2007). Interference between cues of the same outcome depends on the causal interpretation of the events. *Quarterly Journal of Experimental Psychology*, 60, 369–386. doi:10.1080/17470210601000961
- Cochrane Collaboration. (2014). Review Manager (RevMan) Version 5.3 [Computer program]. Copenhagen, Denmark: The Nordic Cochrane Centre, The Cochrane Collaboration.
- Darlow, A. L., & Sloman, S. A. (2010). Two systems of reasoning: Architecture and relation to emotion. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1, 382–392. doi:10.1002/wcs.34
- De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *Spanish Journal of Psychology*, 10, 230–241. doi:10.1017/S1138741600006491
- De Houwer, J., & Beckers, T. (2002). A review of recent developments in research and theories on human contingency learning. *Quarterly Journal of Experimental Psychology*, 55B, 289–310. doi:10.1080/02724990244000034
- Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective reevaluation of causality judgements. *Quarterly Journal of Experimental Psychology*, 49B, 60–80. doi:10.1080/1713932614
- Escobar, M., Pineño, O., & Matute, H. (2002). A comparison between elemental and compound training of cues in retrospective

- reevaluation. *Animal Learning & Behavior*, 30, 228–238. doi:10.3758/BF03192832
- Keppel, G., Bonge, D., Strand, B. Z., & Parker, J. (1971). Direct and indirect interference in the recall of paired associates. *Journal of Experimental Psychology*, 88, 414–422. doi:10.1037/h0030910
- Lipp, O. V., & Dal Santo, L. A. (2002). Cue competition between elementary trained stimuli: US miscuing, interference, and US omission. *Learning and Motivation*, 33, 327–346. doi:10.1016/S0023-9690(02)00001-2
- Luque, D., Cobos, P. L., & López, F. J. (2008). Interference between cues requires a causal scenario: Favorable evidence for causal reasoning models in learning processes. *Learning and Motivation*, 39, 196–208. doi:10.1016/j.lmot.2007.10.001
- Luque, D., Morís, J., Cobos, P. L., & López, F. J. (2009). Interference between cues of the same outcome in a non-causally framed scenario. *Behavioural Processes*, 81, 328–332. doi:10.1016/j.beproc.2008.11.009
- Luque, D., Morís, J., Orgaz, C., Cobos, P. L., & Matute, H. (2011). Backward blocking and interference between cues are empirically equivalent in non-causally framed learning tasks. *Psychological Record*, 61, 141–152.
- Luque, D., Morís, J., & Cobos, P. L. (2010). Spontaneous recovery from interference between cues but not from backward blocking. *Behavioural Processes*, 84, 521–525. doi:10.1016/j.beproc.2009.12.016
- Matute, H., & Pineño, O. (1998a). Cue competition in the absence of compound training: Its relation to paradigms of interference between outcomes. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 38, pp. 45–81). San Diego: Academic Press.
- Matute, H., & Pineño, O. (1998b). Stimulus competition in the absence of compound conditioning. *Animal Learning & Behavior*, 26, 3–14. doi:10.3758/BF03199157
- Miller, R. R., & Escobar, M. (2002). Associative interference between cues and between outcomes presented together and presented apart: An integration. *Behavioural Processes*, 57, 163–185. doi:10.1016/S0376-6357(02)00012-8
- Morís, J., Cobos, P. L., Luque, D., & López, F. J. (2014). Associative repetition priming as a measure of human contingency learning: Evidence of forward and backward blocking. *Journal of Experimental Psychology: General*, 143, 77–93. doi:10.1037/a0030919
- Pineño, O., & Miller, R. R. (2007). Comparing associative, statistical, and inferential reasoning accounts of human contingency learning. *Quarterly Journal of Experimental Psychology*, 60, 310–329. doi:10.1080/17470210601000680
- Riding, R. J., & Rayner, S. (1998). *Cognitive styles and learning strategies: Understanding style differences in learning and behavior*. London: D. Fulton.
- Shanks, D. R. (2007). Associationism and cognition: Human contingency learning at 25. *Quarterly Journal of Experimental Psychology*, 60, 291–309. doi:10.1080/17470210601000581
- Shanks, D. R. (2010). Learning: From association to cognition. *Annual Review of Psychology*, 61, 273–301. doi:10.1146/annurev.psych.093008.100519
- Vadillo, M. A., Castro, L., Matute, H., & Wasserman, E. A. (2008). Backward blocking: The role of within-compound associations and interference between cues trained apart. *Quarterly Journal of Experimental Psychology*, 61, 185–193. doi:10.1080/17470210701557464
- Van Hamme, L. J., & Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation*, 25, 127–151. doi:10.1006/lmot.1994.1008
- Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 5–47). Hillsdale: Erlbaum.
- Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General*, 121, 222–236. doi:10.1037/0096-3445.121.2.222