Original Article

QJEP

Quarterly Journal of Experimental Psychology
2020, Vol. 73(5) 645–653

DOI: 10.1177/1747021819887725
qjep.sagepub.com
SAGE

The effect of conditioned inhibitors and preexposed cues on the outcome-specific Pavlovian-to-instrumental transfer effect in humans

Daniel E Alarcón[1,2] and Charlotte Bonardi[2]

Abstract

Using a human Pavlovian-to-instrumental transfer (PIT) task, Alarcón and Bonardi showed that the selective elevation of instrumental responding produced by excitatory transfer cues was reduced when these cues were presented with a conditioned inhibitor (CI), relative to a control cue that was simply preexposed. However, previous research has shown that preexposed cues might also acquire inhibitor-like properties. This study aimed to contrast the inhibitory properties of CIs and preexposed cues, using novel stimuli as controls, in summation and PIT tests. Participants were trained to perform two actions, each reinforced with a distinct outcome (O_1 or O_2). Two images were trained as CIs, each signalling the absence of one of the outcomes, by presenting them with a cue that was otherwise followed by that outcome (e.g., A→O_1, AI→no O_1). In contrast, the preexposed cues were simply presented in the absence of the outcomes. In the summation test, participants rated the likelihood of the outcomes in the presence of two independently trained excitatory cues, each presented with a CI, a preexposed cue, or a novel stimulus. Similarly, in the PIT test, participants performed both actions in the presence and absence of these compounds. In the summation test, the CIs and the preexposed cues reduced participants' expectations of the outcomes more than the novel stimuli. However, in the PIT test, only the CIs reduced the selective elevation of responding produced by the transfer cues. These results might reflect distinct properties of stimuli trained as CIs and those simply preexposed.

Keywords

Pavlovian-to-instrumental transfer; conditioned inhibition; decision making; incentive motivation; cue-reactivity

Received: 26 March 2019; revised: 15 October 2019; accepted: 16 October 2019

A conditioned inhibitor (CI) signals the *absence* of a significant event. For example, if a conditioned stimulus (CS) A is always paired with an unconditioned stimulus (US) except when accompanied by another cue I, that is, A→US, AI→no US, this unexpected omission of the US drives I to acquire inhibitory properties opposite to those of A. These are often evaluated in *summation* or *retardation* tests (Rescorla, 1969). In the summation test, the CI is presented in compound with a separately trained signal for the US. Inhibition is evident if the CI suppresses this cue's ability to predict the US more than a control stimulus. In the retardation test, the inhibitor and a suitable control stimulus are both separately paired with a new US and the rate of acquisition is compared; learning an excitatory relationship should be slower if the stimulus is an inhibitor.

Alarcón and Bonardi (2016) recently demonstrated conditioned inhibition in a different manner, using a human Pavlovian-to-instrumental transfer (PIT) task (for reviews on PIT, see Cartoni, Balleine, & Baldassarre, 2016; Holmes, Marchand, & Coutureau, 2010). Participants were trained to perform two responses (R_1, R_2), each reinforced by a distinct *outcome* (pictures of food or drinks O_1, O_2; that is, R_1→O_1; R_2→O_2). Two neutral images that were later used as *transfer cues*, T_1 and T_2, were paired with the same outcomes (i.e., T_1→O_1; T_2→O_2), and in a subsequent PIT test participants performed R_1 and R_2 while T_1 and T_2 were presented with one of two neutral *control* stimuli, C_1 and C_2

[1]Departamento de Psicología, Universidad de Chile, Santiago, Chile
[2]School of Psychology, University of Nottingham, Nottingham, UK

Corresponding author:
Daniel E Alarcón, Departamento de Psicología, Universidad de Chile, Av. Capitán Ignacio Carrera Pinto N°1045, Santiago, Chile.
Email: dealarcon@u.uchile.cl

Table 1. Experimental design.

Instrumental	Pavlovian phase			Summation test	PIT test
	Pre-training	Inhibitory			
$R_1 \rightarrow O_1$	$S_1 \rightarrow O_1$	$S_1A \rightarrow O_1$	S_1I_1-	T_1I_1/T_2I_2	T_1I_1/T_2I_2
$R_2 \rightarrow O_2$	$S_2 \rightarrow O_2$	$S_2B \rightarrow O_2$	S_2I_2-	T_1C_1/T_2C_2	T_1C_1/T_2C_2
		C_1C_2-			
		$T_1 \rightarrow O_1$	$T_2 \rightarrow O_2$	T_1N_1/T_2N_2	T_1N_1/T_2N_2
		C_1-	C_2-	T_1I_2/T_2I_1	
		I_1-	I_2-		

S_1, S_2, I_1, I_2, C_1, C_2, N_1, N_2, T_1, T_2: neutral fractal images; I_1, I_2: conditioned inhibitors; T_1, T_2: transfer cues; C_1, C_2: preexposed stimuli; N_1, N_2: novel stimuli; R_1 and R_2: keyboard responses; O_1 and O_2: food and drink images; -: no outcome; PIT: Pavlovian-to-instrumental transfer.

(T_1C_1/T_2C_2). Participants performed R_1 more than R_2 during T_1C_1, but the reverse during T_2C_2—a selective elevation of instrumental responding known as *outcome-specific PIT*. It has been argued that PIT results from the associations between both R_1 and T_1 with O_1: presentation of T_1 activates the O_1 representation, which elicits performance of R_1 (cf. Trapold & Overmier, 1972). This contrasts with the general form of PIT, in which a CS invigorates instrumental responding regardless of whether the CS and the instrumental responses are trained with the same or different outcomes, as long as they are both of the same motivational valence (cf. Rescorla & Solomon, 1967).

In addition, Alarcón and Bonardi (2016) paired a *stimulus* S_1 with O_1 unless accompanied by an *inhibitor* l_1 ($S_1 \rightarrow O_1$; $S_1l_1 \rightarrow$no O), making l_1 an inhibitor for O_1. In the PIT test, the selective elevation of responding during T_1C_1 and T_2C_2 was eliminated when T_1 and T_2 were presented with inhibitor l_1 (T_1l_1/T_2l_1). Pavlovian inhibitors are commonly thought to act by suppressing activation of the outcome representation, and so the authors attributed this effect to l_1's inhibitory properties interfering with the $T_1 \rightarrow O_1 \rightarrow R_1$ chain that mediates PIT. This conclusion was supported by the results of a *summation* test (cf. Rescorla, 1969), in which participants' expectation of O_1 and O_2 during T_1 and T_2, respectively, was significantly weaker during T_1l_1 and T_2l_1 than during T_1C_1 and T_2C_2.

In that study, the control stimuli C_1 and C_2 were preexposed in compound in the absence of any outcome, meaning they were treated identically to I *except* for never being presented with an excitatory CS—so they should not have become inhibitory. However, some have argued that, because the context in which the cue is presented also elicits expectancy of the US, this is sufficient for the stimulus to develop what is termed *differential inhibition* (see Miller, Hallam, Hong, & Dufore, 1991). For instance, Karazinov and Boakes (2004) reported that a CI and a preexposed stimulus both reduced expectation of the US relative to a novel cue—although the CI did this more effectively than the preexposed cue. However, in further experiments, the authors tried to enhance the "inhibitory" properties of the preexposed cue by increasing the excitatory strength of the context during training. Despite these manipulations, the degree to which the preexposed cues reduced participants' expectations of the US did not change—not what would be expected if differential inhibition were responsible. Nonetheless, these findings raise the possibility that the control cues in our PIT study might also appear inhibitory relative to a novel control stimulus. This study explored this possibility.

Alarcón and Bonardi (2016) also found that the effect of the CI was not outcome-specific: although I only signalled the absence of O_1, in the summation test, it reduced expectation of O_2 during T_2 *as effectively* as expectation of O_1 during T_1 and also suppressed the PIT effect produced by T_1 and T_2 to a similar extent. This suggests I predicted the absence of *both* outcomes, even though it had only been trained as a signal for omission of O_1. Although some previous studies have reported evidence consistent with these results (e.g., LoLordo, 1967; Nieto, 1984; Pearce, Montgomery, & Dickinson, 1981), there are also several studies showing that an inhibitor is more effective in reducing responses to a cue that signals the same outcome as that inhibited by the inhibitor (e.g., Delamater, Sosa, & LoLordo, 2003). Although the factors determining whether the inhibitors will act in a general or an outcome-specific manner is unclear, in the case of Alarcón and Bonardi (2016) it may have been because only one inhibitor was trained; perhaps if an inhibitor had also been trained for O_2, the inhibitors would act more selectively. This possibility was also explored.

In this study, we used an outcome-specific PIT design to assess the effect of two inhibitors, each signalling the absence of a distinct outcome, in summation and PIT tests. Initially, participants received instrumental training in which R_1 and R_2 produced outcomes O_1 and O_2 (Table 1). After this they were trained that images S_1 and S_2 (in compound with filler cues A and B) produced O_1 and O_2 *unless* accompanied by l_1 and l_2; thus, l_1 and l_2 predicted the absence of O_1 and O_2, respectively ($S_1A \rightarrow O_1$, S_1I_1-, $S_2B \rightarrow O_2$, S_2I_2-). Two transfer stimuli, T_1 and T_2, were also paired with O_1 and O_2, and the critical control cues C_1 and C_2 were presented in compound without reinforcement,

the same number of times as l_1 and l_2; l_1, l_2, C_1, and C_2 were also presented in isolation.

In the subsequent summation test, we measured participants' expectations of the outcomes by presenting CSs signalling those outcomes together with the inhibitors or with control stimuli that were either novel or preexposed during training. Thus, T_1 and T_2 were presented with the inhibitors for the outcomes they predicted (T_1l_1 and T_2l_2, respectively), the preexposed cues C_1 and C_2 (T_1C_1 and T_2C_2), and two *novel* cues N_1 and N_2 (T_1N_1 and T_2N_2). This novel control condition allowed us to evaluate the extent to which the preexposed cues C_1 and C_2 had become differential inhibitors and acquired inhibitory properties similar to those shown by l_1 and l_2. We expected a greater reduction of participants' expectations to be produced by the inhibitors than the novel control cue; the question was whether the preexposed control cue would also demonstrate inhibitory properties, reducing expectation of the outcome more than the novel stimulus. In addition, T_1 and T_2 were also presented with inhibitors for the alternative outcomes (T_1l_2 and T_2l_1), to establish whether training inhibitors for both outcomes would render them outcome-specific. If inhibition is outcome-specific, lower expectation of the outcomes should be evident when the CS and the inhibitor were both trained with the same outcome.

In the PIT test which followed, we compared the effect of the inhibitors and preexposed control cues on outcome-specific PIT, by presenting the CSs in compound with the CIs, the preexposed stimuli, or the novel control cues (T_1l_1 and T_2l_2, T_1C_1 and T_2C_2, and T_1N_1 and T_2N_2). We predicted a reduction of the selective elevation of instrumental performance produced by the CSs, that is, outcome-specific PIT, when these cues were presented with the inhibitors relative to the novel control cues; the question was again whether the preexposed control cues also displayed inhibitory properties, reducing outcome-specific PIT more than the novel stimuli.

Method

Participants

The participants were 27 students from the University of Nottingham, aged 18–28 (4 males, 23 females). One participant was excluded for only performing one of the two responses in the instrumental phase. Students from the School of Psychology received course credit and the rest received an inconvenience allowance of £4. The number of participants was chosen based on the number of counterbalancing conditions and on previous studies conducted by the authors (Alarcón & Bonardi, 2016).

Apparatus and materials

The task was programmed in PsychoPy (Peirce, 2007) on a standard computer with a 20-inch screen. Twelve neutral fractal images were used as CSs, and images of foods and drinks (4 of each) as outcomes. All images were 8 cm × 8 cm (see Alarcón, Bonardi, & Delamater, 2018). The neutral images were presented immediately to the left or right of the screen centre, and the outcomes were presented centrally. The components of each neutral stimulus compound were presented an equal number of times on the left and right. A fixation cross (10 mm × 10 mm) was located at the centre of the screen in the instrumental phase and PIT test and was replaced by a fixation dot (3 mm × 3 mm) in the Pavlovian phase. R_1 and R_2 were pressing the keys "z" and "m." For half of the participants, "z" was reinforced with food and "m" with drink, and the reverse for the remaining participants. In the summation test, participants used the mouse to click on the rating scale, and the "space" bar to advance through the phases. Instructions were presented on the screen before each phase.

Procedure

Participants received a brief explanation of the task and then completed a consent form before they were guided into a quiet room.

Instrumental training

Participants were instructed to learn the relationships between pressing the "z" and "m" keys and different rewards (images of foods or drinks) and to obtain as many rewards as they could. Each response was reinforced according to a variable ratio (VR) 5 schedule, such that an outcome was presented after an average of five responses. Each outcome was presented in the centre of the screen, either until 0.8 s had elapsed or until another key was pressed, at which point it was replaced by the fixation cross. This phase ended when participants had received 100 outcomes in total.

Pavlovian phase

Participants were informed that different images would appear on the screen, some of them followed by the rewards and others not, and that they would have to answer a series of questions about these relationships by pressing numbers on the keyboard. In each trial, the text "Which reward will appear now?" was presented at the top of the screen, one CS image (or compound) below it, and the text "1) Food 5) Drink 9) Nothing" at the bottom. When participants pressed one of the numbers on the keyboard (1, 5, or 9), these stimuli were removed from the screen and the correct outcome was presented; in addition, depending on the participant's response, the text "Correct!" in green font, or "Oops! That was wrong" in red font, was presented at the top or bottom of the screen, respectively. The intertrial interval (ITI) was 2 s, in which only the fixation dot was present.

The initial, pre-training stage was divided into two blocks, each comprising four trials of $S_1 \rightarrow O_1$ and four trials of $S_2 \rightarrow O_2$, to establish S_1 and S_2 as signals for their respective outcomes. The inhibitory stage that followed was divided into two blocks, each comprising four trials of $S_1A \rightarrow O_1$ and $S_2B \rightarrow O_2$, six trials each of $T_1 \rightarrow O_1$, $T_2 \rightarrow O_2$, S_1I_1-, S_2I_2-, and C_1C_2-, and three trials of I_1-, I_2-, C_1, and C_2 (Table 1). The filler cues A and B were added to S_1 and S_2 on reinforced trials to ensure that participants could not simply learn that when two cues were presented no outcome followed, without needing to pay attention to the individual cues in each compound. The order of the trial types was semi-random within each block. Participants could, if they wished, take a break between these blocks.

Summation test

Participants were informed that images would appear on the screen, after which they must rate the likelihood of an outcome's appearance using a rating scale. On each trial the text "In a scale from 1 to 100, how likely is it that this image will be followed by" together with the word "FOOD" or "DRINK" was presented at the top of the screen. For half of the trials, the questions ended with the word "FOOD" and the compounds consisted of the transfer cue signalling food with either the inhibitor trained with food, for example, T_1I_1, the inhibitor trained with drinks, for example, T_1I_2, one of the preexposed control cues, for example, T_1C_1, or one of the novel cues, T_1N_1. For the other half, the question ended with the word "DRINK," and the stimuli presented were corresponding compounds containing the transfer cue signalling drink, for example, T_2I_2, T_2I_1, T_2C_2, and T_2N_2. The compounds were presented, and below them a rating scale with "0" on the left and "100" on the right. The test consisted of two blocks, each comprising two trials of each compound (Table 1); the order of questions was randomised within each block.

Instrumental retraining

This was designed to serve as a reminder of the specific R-O associations learned at the beginning of the experiment, and was identical to the initial instrumental phase, except participants earned only 50 outcomes in total.

PIT test

Participants were instructed to press either "z" or "m" as much as they wanted in order to obtain the rewards, but no outcomes were presented. In each trial, the fixation dot was present for 2 s (pre-CS baseline period), followed by a 2-s CS presentation (CS period). The test was divided into three blocks, each comprising two presentations of each stimulus compound, presented in a semi-random order.

Data treatment

The data were analysed using analysis of variance (ANOVA). Significant two-way interactions were explored with simple main effects analysis using the pooled error term. Partial eta-squared (η_p^2) and its 90% confidence interval were given for significant effects and interactions. The responses made in the Pavlovian conditioning trials were grouped as "correct" if participants chose the outcome predicted by the cues to the question "Which reward will appear now?," and "incorrect" if they chose either the other outcome or the "nothing" option (see below). Responses in the PIT test were computed separately for CS and preCS periods, transformed to responses per minute (rpm) and grouped according to the outcome signalled by the transfer cue and the response. Responses trained with the same outcome as that signalled by the transfer cue were grouped as *Same*, for example, R_1 during T_1 compounds, otherwise as *Different*, for example, R_2 during T_1 compounds. Difference scores were calculated by subtracting responding during the pre-CS period from that during the CS period. Positive scores reflect an elevation of responding over background levels, and higher *Same* than *Different* scores reflect outcome-specific PIT.

Results

The Pavlovian and instrumental phases were completed uneventfully. The mean number of R_1 responses and O_1 presentations were 269.9 (SEM = 14.4) and 51.8 (SEM = 2.9), respectively, and for R_2 and O_2 251.2 (SEM = 13.4) and 54.4 (SEM = 2.9), respectively. The mean proportion of correct responses in the inhibitory training (0 = *no correct answers*, 1 = *all correct answers*), for blocks 1 and 2, respectively, were 0.58–0.58 (SEM = 0.05–0.06) for S_1A/S_2B; 0.55–0.76 (SEM = 0.04–0.05) for S_1I_1/S_2I_2; 0.82–0.92 (SEM = 0.05–0.04) for C_1C_2; 0.63–0.87 (SEM = 0.03–0.03) for T_1/T_2; 0.69–0.87 (SEM = 0.05–0.05) for I_1/I_2; and 0.65–0.9 (SEM = 0.05–0.05) for C_1/C_2. An ANOVA with Block (1, 2) and Cue (S_1A/S_2B, S_1I_1/S_2I_2, C_1C_2, T_1/T_2, I_1/I_2, C_1/C_2) showed a significant main effect of Block, $F(1, 25) = 61.18$, $MSe = .03$, $p < .001$, significant main effect of Cue, $F(5, 125) = 10.06$, $MSe = .05$, $p < .001$, and a significant Block × Cue interaction, $F(5, 125) = 4.73$, $MSe = .024$, $p = .001$. The analysis of the interaction confirmed more correct responses in Block 2 relative to Block 1 for compounds S_1I_1/S_2I_2 ($p < .001$), C_1C_2 ($p = .001$), T_1/T_2 ($p < .001$), I_1/I_2 ($p < .001$), C_1/C_2 ($p < .001$), but not for S_1A/S_2B ($p = .86$). In view of the fact that performance on S_1A/S_2B trials did not increase over blocks, we further explored the incorrect responses to these compounds and found that in 83% of the cases participants chose the option "nothing" (rather than the alternative outcome). This was likely caused by S_1 and S_2 also being paired with the

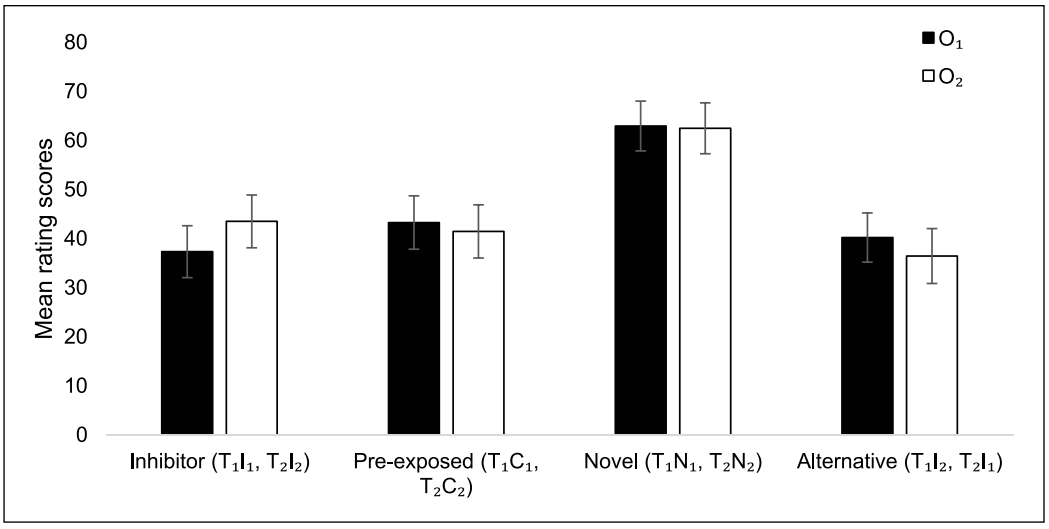

Figure 1. Mean rating scores (participants' expectancies) ± SEM to the questions about O_1 and O_2 for the transfer cues in compound with the inhibitors (T_1I_1, T_2I_2), the preexposed control cues (T_1C_1, T_2C_2), the novel control cues (T_1N_1, T_2N_2), and the alternative inhibitors (T_1I_2, T_2I_1) in the summation test.

absence of the outcomes when presented with the inhibitors, rendering them potentially ambiguous cues.

The rating scores in the summation test to T_1 and T_2, according to whether they were paired with the corresponding CI (T_1l_1 and T_2l_2), the preexposed control cue (T_1C_1 and T_2C_2), the novel control cue (T_1N_1 and T_2N_2), or the alternative inhibitor (T_1l_2 and T_2l_1), are plotted in Figure 1. An ANOVA with Outcome (O_1, O_2) and Cue (Inhibitor, Preexposed, Novel, Alternative) showed a significant main effect of Cue, $F(3, 75)=10.9$, $MSe=607.6$, $p<.001$, $\eta_p^2=.30$. Nothing else was significant, largest $F(3, 75)=1.3$, $MSe=184.1$, $p=.28$ for the Outcome × Cue interaction. Post hoc comparisons confirmed higher rating scores to the novel compounds than to the inhibitory, $p<.01$, preexposed, $p<.01$, and alternative compounds, $p<.001$. No other difference was statistically significant. The analysis confirmed that participants rated the likelihood of outcome presentation to be as low during the preexposed cues as during the Pavlovian inhibitors, and lower in the presence of all these compounds than the novel stimuli (Inhibitor = Alternative = Preexposed < Novel); in addition, ratings were unaffected by whether the CI was trained with the same outcome as the transfer cue or the alternative outcome (Inhibitor = Alternative), showing that, just as in Alarcón and Bonardi (2016), the inhibitors were not outcome-specific by this measure.

The results of the PIT test are presented in Figure 2, which shows that, as expected, the transfer cues (T_1 and T_2) elicited more Same than Different responses when presented with the novel control stimuli (T_1N_1 and T_2N_2). This outcome-specific PIT effect appeared equally substantial in the preexposed stimulus compounds T_1C_1 and T_2C_2, but was reduced with the inhibitory compounds (T_1I_1 and T_2I_2). ANOVA with Outcome (Same, Different) and Compound (Preexposed, Inhibitory, Novel) as factors showed a significant main effect of Outcome, $F(1, 25)=9.7$, $MSe=15,389.1$, $p<.01$, $\eta_p^2=.28$ and a significant Outcome × Compound interaction, $F(2, 50)=3.27$, $MSe=4,936.2$, $p=.046$, $\eta_p^2=.12$. The effect of Compound was not significant, $F<1$. Exploration of the interaction confirmed higher Same than Different responses for the Preexposed, $F(2, 50)=13.92$, $MSe=4,936.2$, $p<.001$, and Novel compounds, $F(2, 50)=21.52$, $MSe=4,936.2$, $p<.001$, but not for the Inhibitory compounds, $F(1, 50)=1.33$, $MSe=4,936.2$, $p=.27$. Equivalent analysis of preCS responding revealed nothing significant, largest $F(2, 50)=1.13$, $MSe=5,327.89$, $p=.33$ for the Outcome × Compound interaction. For the preexposed (T_1C_1, T_2C_2), inhibitory (T_1I_1, T_2I_2), and novel (T_1N_1, T_2N_2) compounds, respectively, the mean preCS Same responses were 185.2 (SEM = 31.5), 192.9 (SEM = 33.1), and 173.8 (SEM = 30.0), and the mean Different responses 176.5 (SEM = 28.6), 167.9 (SEM = 30.3), and 191.5 (SEM = 33.9).

Discussion

This experiment aimed to establish whether, in our procedure, preexposed stimuli would behave like CIs in summation and PIT tests. When compounded with the novel control stimuli N_1 and N_2, the transfer cues T_1 and T_2 elicited expectation of their respective outcomes; moreover, this effect was significantly reduced when they were compounded with the inhibitors l_1 and l_2, and *also* the preexposed control stimuli C_1 and C_2. This indicates that preexposure of C_1 and C_2 in training had made them as inhibitory as "true" Pavlovian inhibitors. Karazinov and Boakes (2004) also found that preexposed cues were more inhibitory than novel cues in a summation test,

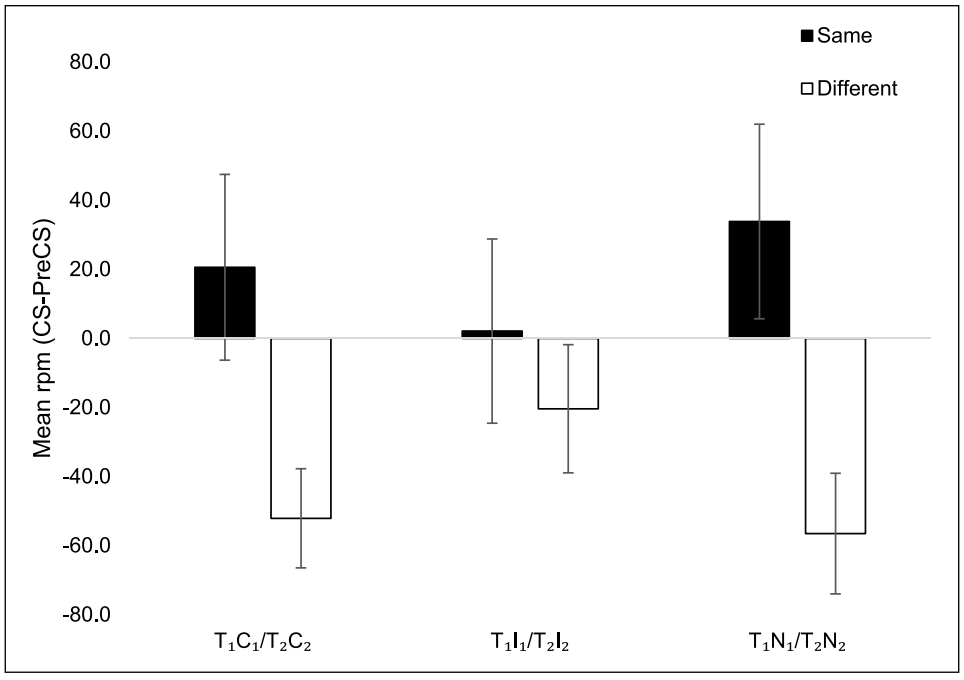

Figure 2. Same and different responding to the transfer cues in compound with the preexposed control stimuli (T_1C_1/T_2C_2), the inhibitors (T_1I_1/T_2I_2), and novel control cues (T_1N_1/T_2N_2) in the PIT test.

but—in contrast to our findings—less so than Pavlovian inhibitors, a difference we also observed in previous work (Alarcón & Bonardi, 2016). Karazinov and Boakes speculated that the difference between the Pavlovian and differential inhibitors stemmed from differences in how they were trained. While the Pavlovian CI was trained both as a single cue and in a compound, the preexposed differential CI was only trained alone. Seeing the preexposed cue in a compound for the first time at test might have produced uncertainty in the participants, making their ratings to this compound more conservative than to the other cues (the ratings to this compound were closer to the centre of a rating scale marked as "Don't know"). However, in both Alarcón and Bonardi (2016) and the experiment reported here the Pavlovian CIs and the preexposed stimuli were presented the same number of times as single cues and compounds across the task, making it unlikely this factor was responsible. One potentially critical difference between our two studies is that Alarcón and Bonardi (2016)—which was primarily aimed to explore the effect of Pavlovian inhibition on PIT—excluded participants who had not learned the Pavlovian inhibitory discrimination in training. It is possible that considering only the participants that showed inhibition in the summation test might have increased the differences between the preexposed control and the inhibitor in the PIT test; however, we reanalysed the data of the present experiment using the same response criterion as Alarcón and Bonardi (2016), but still found no significant differences between these cues.

In summary, although the summation technique indicates that preexposure produces conditioned inhibition, it is unclear whether or not this is as strong as that produced by Pavlovian training. Moreover—in view of the failure of Karazinov and Boakes to manipulate the degree of differential inhibition as theory would predict—it is also worth considering alternative explanations. For example, perhaps the preexposed cue is becoming associated with the idea that "nothing" is going to happen, and expectation of this neutral event interferes with expectation of the outcome, reducing response ratings. This could have a similar effect in the summation test as the active expectation of outcome omission produced by the Pavlovian inhibitor (cf. Alarcón & Bonardi, 2016).

Consistent with this suggestion, although the results of the summation test suggested the preexposed cues acquired inhibitory properties, the PIT test did not. Here, the selective elevation of instrumental responding produced by T_1 and T_2—outcome-specific PIT—was eliminated only by the CIs, and by neither the preexposed or novel cues, which did not differ (cf. Alarcón & Bonardi, 2016). These results suggest that—in contrast to the results of the summation test—the preexposed cues were *not* inhibitory. This discrepancy between summation and PIT tests may be due to their differential sensitivity in detecting, perhaps subtle, differences in inhibition. Moreover, our sample size may not have been large enough to detect small differences between the inhibitors and the preexposed controls in the summation test, and between the preexposed controls and the novel cues in the PIT test. For instance, a larger sample

size might have revealed that the preexposed controls were inhibitory relative to the novel cues, but less so than the Pavlovian inhibitors, that is, inhibitors < preexposed < novel. Alternatively, it may be that the nature of the inhibitory properties endowed by Pavlovian inhibition training and nonreinforced preexposure is qualitatively different and thus manifests differently depending on how it is tested. This would suggest that simple preexposure does not result in true inhibition, summation tests are sensitive to factors other than pure Pavlovian inhibition, and the PIT test is a purer measure of this property. These alternative explanations make it necessary to continue studying the effect of conditioned inhibition on PIT.

Another feature of Alarcón and Bonardi's (2016) results was that the inhibitor for O_1 reduced summation and PIT effects just as much for T_2 as for T_1, suggesting that its inhibition was not outcome-specific. We suggested that this might be because only one CI-no US relationship was trained, and so in this study we trained two inhibitors, one for each of the outcomes. However, this had no effect on the result: in the summation test, there was no difference in expectation of the outcomes during the inhibitory compounds, regardless of whether the inhibitor predicted omission of the outcome predicted by the excitor with which it was paired, or the alternative outcome. In fact, there is already some controversy over whether CIs are specific to the sensory properties of the outcomes whose omission they signal, or whether they act via a solely motivational mechanism. Initial studies suggested that CIs trained with a specific US, such as a shock to the left eye, would inhibit responding equally to any CS signalling shock, regardless of the eye to which it was delivered (Nieto, 1984; Pearce et al., 1981; see LoLordo & Fairless, 1985, for a review). In contrast, more recent research, using PIT measures of inhibition, suggests a different story. Quail, Laurent, and Balleine (2017; see also Laurent, Wong, & Balleine, 2015) used a human task to train two inhibitors, one for each of two outcomes ($S_1 \rightarrow O_1$, $S_2 \rightarrow O_2$, S_1 $l_1 \rightarrow$nothing, S_2 $l_2 \rightarrow$nothing). The effects of l_1 and l_2 were evaluated in a PIT test, both when presented alone, when paired with the CSs with which they were trained (S_1l_1, S_2l_2)—and with the alternative CSs (S_1l_2, S_2l_1). If l_1, for example, were a specific inhibitor for O_1, it should reduce the PIT effect produced by S_1—which predicts the precise outcome whose omission l_1 signalled—but *not* that produced by S_2, which predicted the alternative outcome. This is what the authors observed.

Unfortunately, there were many differences between this study that demonstrated outcome-specific inhibition and our experiments which did not. One is that we evaluated learning in each trial of the Pavlovian conditioning phase by asking participants to press a key to predict the outcome to be delivered. Because of this, participants had to choose the option "Nothing" for the CIs and the control cues. It is possible that this consequence, being common to both inhibitors, hindered the development of outcome-specific inhibition. Quail et al. (2017), in contrast, evaluated learning at the end of the Pavlovian conditioning phase, and participants could answer if the inhibitor was associated with the absence of a particular outcome, for example, I_1 followed by no food, or with the general absence of the outcomes, for example, I_1 followed by no foods or drinks, which might have encouraged a discrimination between the CIs. In addition, Quail et al. separated their participants based on these answers (Experiment 2), finding that the "specific learners," who identified specific inhibitor→no outcome relationships, showed high levels of outcome-specificity in the PIT test, whereas the "general learners," those who identified the inhibitors as signalling the absence of any outcome, did not show any signs of such specificity. It is possible that we might show a similar effect by separating the participants similarly; however, our measures of Pavlovian learning do not allow us to divide our participants into specific or general learners.

Another difference is that Quail et al. did not evaluate their inhibitors' properties against separately trained test excitors, and only used a PIT test (Rescorla, 1969). Conversely, our first experiment that demonstrated non-outcome-specific inhibition in the PIT test only trained one inhibitor, while in the present experiment we did not evaluate the alternative compounds in the PIT test. Our evidence for non-outcome-specific inhibition came only from a summation test, which Quail et al. did not include. Thus, there is no real conflict of results here—but a key need to bridge the gap between these various studies in order to establish whether or not inhibition is outcome-specific in these kinds of tasks.

Another point worth mentioning is the use of a brief instrumental retraining before the PIT test, the purpose of which was to serve as a reminder of the instrumental associations learned at the beginning of the task (see Alarcón & Delamater, 2019, for the same procedure in rodent research). It might be argued that this is similar to an "overtraining" procedure, resulting in habitual responding rather than goal-directed behaviour, which is less dependent on the outcome representation. However, we do not consider this training to be sufficient to produce habits. There is one example in which habits have been achieved by overtraining instrumental responses in humans, but it required several separate sessions of instrumental conditioning (Tricomi, Balleine, & O'Doherty, 2009), and recently a series of failed attempts to reproduce those findings have been published (de Wit et al., 2018). Related to overtraining, Garofalo and Robbins (2017) provided evidence that extended training in humans enhances the outcome-specific PIT effect and it would be interesting to assess if the ability of the CIs to reduce selective PIT remains after instrumental overtraining.

A final consideration is that, although it was designed to be similar to the tasks used in animal research, we did not obtain a direct measure of Pavlovian conditioning but rather asked participants about explicit expectations of the outcomes, as in contingency learning tasks (e.g., Shanks, 2007). Some researchers have argued that the processes involved in

this type of task depend on the generation of propositions about the relationships between the events, rather than the traditional associative view, and there have been recent efforts to explain outcome-specific PIT with a propositional account (e.g., Hogarth et al., 2014; Seabrooke, Hogarth, & Mitchell, 2016). However, the outcome-specific PIT effect is consistently found in both animal and human studies, suggesting that the measure of Pavlovian conditioning might not be critical. For instance, the inhibition of PIT results found by Quail et al. (2017) was remarkably similar to previous results found in rodents by the same group of researchers (Laurent et al., 2015). Nevertheless, adding direct measures of Pavlovian conditioning might provide further information about the underlying processes in this type of task.

In summary, we have demonstrated that simple preexposure can make a stimulus behave like a CI in a summation test, relative to a novel control; however, this was not evident in a PIT test of inhibition. Of several possible interpretations, one possibility is that summation tests are sensitive to effects other than Pavlovian inhibition and that the PIT test is a better measure of a stimulus' inhibitory properties. We also failed to produce evidence of outcome-specific inhibition in a summation test, despite training two inhibitors, one for each outcome. Although this ostensibly contradicts findings from other studies (e.g., Quail et al., 2017), a more likely explanation is that this discrepancy arises from procedural differences. More generally, we believe that these findings expand our understanding of the mechanisms underlying PIT and that this task might be helpful to study the effect of conditioned inhibition in maladaptive behaviour, such as drug-seeking behaviours and overeating.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The elaboration of this manuscript was supported by a postdoctoral grant awarded by Conicyt (Fondecyt # 3170166) to D.E.A.

ORCID iDs

Daniel E Alarcón https://orcid.org/0000-0001-8609-3479
Charlotte Bonardi https://orcid.org/0000-0001-8073-8184

References

Alarcón, D. E., & Bonardi, C. (2016). The effect of conditioned inhibition on the specific Pavlovian-instrumental transfer effect. *Journal of Experimental Psychology: Animal Learning and Cognition*, *42*, 82–94. doi:10.1037/xan0000087

Alarcón, D. E., Bonardi, C., & Delamater, A. R. (2018). Associative mechanisms involved in specific Pavlovian-to-instrumental transfer in human learning tasks. *Quarterly Journal of Experimental Psychology*, *71*, 1607–1625. doi:10.1080/17470218.2017.1342671

Alarcón, D. E., & Delamater, A. R. (2019). Outcome-specific Pavlovian-to-instrumental transfer (PIT) with alcohol cues and its extinction. *Alcohol*, *76*, 131–146.

Cartoni, E., Balleine, B., & Baldassarre, G. (2016). Appetitive Pavlovian-instrumental transfer: A review. *Neuroscience & Biobehavioral Reviews*, *71*, 829–848. doi:10.1016/j.neubiorev.2016.09.020

Delamater, A. R., Sosa, W., & LoLordo, V. M. (2003). Outcome-specific conditioned inhibition in Pavlovian backward conditioning. *Animal Learning & Behavior*, *31*, 393–402. doi:10.3758/BF03196000

de Wit, S., Kindt, M., Knot, S. L., Verhoeven, A. A., Robbins, T. W., Gasull-Camos, J., . . . Gillan, C. M. (2018). Shifting the balance between goals and habits: Five failures in experimental habit induction. *Journal of Experimental Psychology: General*, *147*, 1043–1065.

Garofalo, S., & Robbins, T. W. (2017). Triggering avoidance: Dissociable influences of aversive Pavlovian conditioned stimuli on human instrumental behavior. *Frontiers in Behavioral Neuroscience*, *11*, 63.

Hogarth, L., Retzler, C., Munafo, M. R., Tran, D. M., Troisi, J. R., Rose, A. K., . . . Field, M. (2014). Extinction of cue-evoked drug-seeking relies on degrading hierarchical instrumental expectancies. *Behaviour Research and Therapy*, *59*, 61–70.

Holmes, N. M., Marchand, A. R., & Coutureau, E. (2010). Pavlovian to instrumental transfer: A neurobehavioural perspective. *Neuroscience & Biobehavioral Reviews*, *34*, 1277–1295. doi:10.1016/j.neubiorev.2010.03.007

Karazinov, D. M., & Boakes, R. A. (2004). Learning about cues that prevent an outcome : Conditioned inhibition and differential inhibition in human predictive learning. *Quarterly Journal of Experimental Psychology*, *57*, 153–178. doi:10.1080/02724990344000033

Laurent, V., Wong, F. L., & Balleine, B. W. (2015). δ-Opioid receptors in the accumbens shell mediate the influence of both excitatory and inhibitory predictions on choice. *British Journal of Pharmacology*, *172*, 562–570. doi:10.1111/bph.12731

LoLordo, V. M. (1967). Similarity of conditioned fear responses based upon different aversive events. *Journal of Comparative and Physiological Psychology*, *64*, 154–158.

LoLordo, V. M., & Fairless, J. L. (1985). Pavlovian conditioned inhibition: The literature since 1969. In R. R. Miller & N. E. Spear (Eds.), *Information processing in animals: Conditioned inhibition* (pp. 1–49). Hillsdale, NJ: Lawrence Erlbaum.

Miller, R. R., Hallam, S. C., Hong, J. Y., & Dufore, D. S. (1991). Associative structure of differential inhibition: Implications for models of conditioned inhibition. *Journal of Experimental Psychology. Animal Behavior Processes*, *17*, 141–150. doi:10.1037/0097-7403.17.2.141

Nieto, J. (1984). Transfer of conditioned inhibition across different aversive reinforcers in the rat. *Learning and Motivation*, *15*(1), 37–57. doi:10.1016/0023-9690(84)90015-8

Pearce, J. M., Montgomery, A., & Dickinson, A. (1981). Contralateral transfer of inhibitory and excitatory eyelid conditioning in the rabbit. *Quarterly Journal of Experimental Psychology: Comparative & Physiological Psychology*, *33*(1), 45–61. doi:10.1080/14640748108400828

Peirce, J. W. (2007). PsychoPy-psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. doi:10.1016/j.jneumeth.2006.11.017

Quail, S. L., Laurent, V., & Balleine, B. W. (2017). Inhibitory Pavlovian-instrumental transfer in humans. *Journal of Experimental Psychology: Animal Learning and Cognition*, *43*, 315–324. doi:10.1037/xan0000148

Rescorla, R. A. (1969). Pavlovian conditioned inhibition. *Psychological Bulletin*, *72*, 77–94. doi:10.1037/h0027760

Rescorla, R. A., & Solomon, R. L. (1967). Two-process learning theory: Relationships between Pavlovian conditioning and instrumental learning. *Psychological Review*, *74*, 151–182.

Seabrooke, T., Hogarth, L., & Mitchell, C. J. (2016). The propositional basis of cue-controlled reward seeking. *Quarterly Journal of Experimental Psychology*, *69*, 2452–2470. doi:10.1080/17470218.2015.1115885

Shanks, D. R. (2007). Associationism and cognition: Human contingency learning at 25. *Quarterly Journal of Experimental Psychology*, *60*, 291–309.

Trapold, M. A., & Overmier, J. B. (1972). The second learning process in instrumental learning. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 427–452). New York, NY: Appleton-Century-Crofts.

Tricomi, E., Balleine, B. W., & O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *European Journal of Neuroscience*, *29*, 2225–2232.