The University of
Nottingham

UNITED KINGDOM · CHINA · MALAYSIA

Thurston, Meghan Dory and Cassaday, Helen J. (2015)
Conditioned inhibition of emotional responses:
retardation and summation with cues for IAPS
outcomes. Learning and Motivation, 52 . pp. 69-82.
ISSN 0023-9690

Access from the University of Nottingham repository:
http://eprints.nottingham.ac.uk/30587/2/Thurston%20and%20Cassaday%20L%26M%20final
%20author%20version.pdf

Copyright and reuse:

A note on versions:

Conditioned inhibition of emotional responses:
Retardation and summation with cues for IAPS outcomes

Thurston, M.D.[a&b]* & Cassaday, H.J.[a]

[a] School of Psychology, University of Nottingham, University Park, Nottingham, NG7 2RD
[b]*Correspondence: <u>Mt303@le.ac.uk</u>, Clinical Psychology, University of Leicester, 104 Regent Road, Leicester, LE1 7LT

Running head: Conditioned inhibition of emotional responses

Key words: Inhibitory learning; Retardation test; Summation test; International Affective Picture System

Abstract

Conditioned inhibition occurs when a stimulus inhibits the responses that would normally occur to a conditioned stimulus that previously predicted an outcome of interest (the unconditioned stimulus, which elicits responding unconditionally). The present study tested inhibitory learning using emotionally salient cues provided by the use of pictures from the International Affective Picture System (IAPS). The procedure in use was adapted to confirm the demonstration of conditioned inhibition using two key transfer tests, retardation and summation. Experiment 1 showed the development of the predicted discrimination learning for negative outcomes but not for positive outcomes. Experiment 2 found evidence for retardation. Furthermore, this reduced learning was clearly related to the conditioned emotional response to the US images; individuals rated transfer images as positive if they had previously signalled the absence of a negative outcome. Experiment 3 showed that the conditioned inhibition was confirmed by summation test. Thus, inhibitory learning was confirmed by both retardation and summation tests, which between them control for alternative explanations of apparent conditioned inhibition, conducted on different participants but using the same discrimination learning procedure. Moreover, the use of emotionally salient cues as the unconditioned stimuli more closely resembles the traditional Pavlovian paradigm.

[195/200 words]

1. INTRODUCTION

Associative learning is an essential cognitive ability that allows us to understand relationships between environmental events, through the development of stimulus-stimulus associations. Such associations reflect apparent contingencies between signal (conditioned stimulus, CS) and outcome (unconditioned stimulus, US). Thus, the underlying mechanisms of associative learning are of fundamental theoretical importance. One particularly important aspect of associative learning is the ability to inhibit associations under circumstances which indicate that earlier established contingencies are no longer operative. An environmental cue which reliably indicates that a CS (e.g., A) will not be followed by the otherwise expected US is known as a conditioned inhibitor (CI, e.g., B). One method of establishing conditioned inhibition is via a feature negative discrimination procedure. For example, in training trials an excitatory CS is paired with a US (e.g., A→US); interspersed with reinforced training trials, the CS is paired within an alternative non-reinforced compound (AB-) and the participant learns that the CI indicates the absence of the otherwise expected US ([CS+CI]→'no US'). In common with other aspects of associative learning, inhibitory learning can be studied in humans as well as other animals.

Based on the idea that CIs acquire negative associative strength, it is widely accepted that there are two key tests for conditioned inhibition, the retardation test and the summation test (Hearst, 1972; Papini & Bitterman, 1993; Rescorla, 1969; Wasserman, Franklin & Hearst, 1974; Williams, Overmier & LoLordo, 1992). In a retardation test, a true CI should take longer to be converted into an excitatory CS. In the summation test (initially used by Pavlov to demonstrate conditioned inhibition, Pavlov, 1927) a true inhibitor should inhibit responding to a new CS (with which it has not previously been paired). It has been argued that to conclusively demonstrate conditioned inhibition, ideally both of these tests must be passed to rule out other alternative explanations of the apparent inhibition (Rescorla, 1969). For example, in a summation test too much attention may be paid to the designated CI at the cost of the accompanying CS, thus the notional CI may rather distract from the CS and reduce responding to it. In a retardation test, the opposite case could be true in that too little attention may be paid to the CI, because its prior training history involves only non-reinforced exposures. In this case any reduction in learning about the CI would simply be an artefact of reduced attention at the discrimination learning stage, because of latent inhibition. Both of these alternative explanations rely on attention being imperative to responding

(Mackintosh, 1975; Pearce & Hall, 1980) and hence it was proposed that both tests are required to discount attentional explanations of performance in inhibitory learning tests (Rescorla, 1969). However, the role of attention is inconclusive in the absence of any direct evidence of its effects in conditioned inhibition procedures and not always plausible. For example, retardation could in theory result from latent rather than conditioned inhibition but this seems unlikely given that the discrimination is learned during the 'pre-exposure' provided by the training phase, which is not optimal for the production of latent inhibition as the CI is presented together with the CS (albeit without further reinforcement, i.e. presented with the image denoting no US). Provided appropriate control conditions are used, a single test may be sufficient to provide credible evidence of conditioned inhibition (Papini & Bitterman, 1993; Williams et al., 1992). However, the two-test strategy adopted in the present study allows for the use of simpler designs to demonstrate conditioned inhibition with a reasonable effect size, while at the same time eliminating alternative attentional explanations. The longer term objective is to devise task variants which are suitable for use in clinical populations, for whom the complexity of the discriminations to be learned can be a limiting factor.

An additional consideration arises in that previous human conditioned inhibition studies have not typically used stimuli likely to elicit particularly strong emotional responses in participants. For example, using a food migraine task in which participants were required to predict which foods prevented the incidence of a migraine for an experimental character (Karazinov & Boakes, 2004), or in a 'Mission to Mars' task requiring participants to watch planets appear on the screen and predict whether an intact or exploded rocket would appear (Kantini, Cassaday, Batty, Hollis, & Jackson, 2011a, Kantini, Cassaday, Hollis, & Jackson, 2011b; Migo, Corbett, Graham, Smith, Tate, Moran & Cassaday, 2006). Participants in these studies were motivated to complete the task successfully but the stimuli used by way of US outcomes were unlikely to elicit strong emotional responses directly. Animal studies in contrast use emotionally salient USs which may be either appetitive (e.g., food) or aversive (e.g., foot shock) which are likely to elicit strong emotional responses directly (Fernando et al., 2014). These animal studies are not directly comparable with studies with human participants which are motivated indirectly (e.g., by the reward of successfully completing the task). The International Affective Picture System (IAPS) images used in the present studies are an improvement in this regard in that valence and arousal ratings have been used to categorise images as positive, negative or neutral (Lang, Bradley, & Cuthbert, 2005). As such, the

positive and negative IAPS pictures should elicit unconditioned responses directly and in this sense provide more suitable stimuli for conditioning and lend themselves to human conditioning variants more comparable with those used in animals. Previous studies of inhibitory learning using the IAPS only selected positive images (He, Cassaday, Howard, Khalifa, & Bonardi, 2011; He, Cassaday, Park, & Bonardi, 2012). Matched aversive and appetitively-motivated procedural variants would help to extend the generality of these findings.

Inhibitory learning has been examined in relation to psychological and psychiatric conditions (He et al., 2011, 2012; Kantini et al., 2011a, 2011b) as well as individual differences within the normal range (He, Cassaday, Bonardi, & Bibby, 2013, Heym, Kantini, Checkley, & Cassaday, 2014; Migo et al., 2006). Yet, there is a paucity of workable experimental procedures which pass both the retardation and summation tests for inhibitory learning. Therefore, the present study was developed based on the standard inhibitory learning design (Pavlov, 1927; Rescorla, 1973) using a series of procedural variants of a computer-based task which presented street scene images as CSs, followed by positive or negative IAPS images as US outcomes (or a neutral screen on trials when the CS was presented together with a CI). Experiment 1 showed that, for the particular learning task adopted in the present study, the necessary discrimination between CS and CS plus CI was not learned when the US outcomes were positive IAPS images. Therefore, Experiments 2 and 3 went on to develop retardation and summation test variants using negative IAPS images.

Both Experiments 2 and 3 had four stages: pre-discrimination, discrimination, test (retardation or summation) and extinction. Participants were trained on the inhibitory learning discrimination and conditioned inhibition was confirmed by either retardation or summation test. As far as possible, identical instructions were used for both the retardation and summation test variants demonstrated in Experiments 2 and 3. Both variants involved evaluative conditioning (EC) or changes in liking of the CS (measured as positive versus negative ratings at the discrimination learning stage) as well as predictive learning ('what would come next' at the test stage). This switch in task demands was made in order to determine trial-by-trial awareness ratings, suitable for statistical analyses, in addition to the qualitative measure of awareness provided by the post-task question used in Experiment 2.

To date, inhibitory learning has received very little attention in the EC literature (Hofmann, Houwer, Perugini, Baeyens, & Grombez, 2010). In the present experiments, although the test stages introduced a predictive learning component, the inhibitory learning discrimination was established using EC procedures.

2. EXPERIMENT 1

2.1. Methods

2.1.1. Participants

A total of 24 participants via opportunity sampling from the University of Nottingham and local community volunteered to take part in this experiment. There were 5 males and 19 females with a mean age of 20 (range from 18-28). All participants had normal or corrected to normal vision, were not colour-blind and were naïve to the current task and hypothesis. The study was approved by The University of Nottingham, School of Psychology Ethics Committee. Participants received an inconvenience allowance of £5 to cover their travel expenses.

2.1.2. Materials

Twenty colour pictures, 10 positive, and 10 negative were selected from the IAPS[1] and used interchangeably as the USs: on each trial one of the IAPS images (from the pool of 10) was randomly selected as the US. IAPS pictures with any sexual nature were excluded from use as there is typically a gender bias in the ratings. An off-white screen was used to signal the absence of any US ('no US'). Two black and white street scenes were used as CSs. Each of the street scenes showed a road edged by pavement and buildings. The street scenes provided neutral images, clearly distinguishable from the coloured IAPS images in use (none of which depicted any street scenes).Two coloured images of street furniture (e.g. post box and car) provided the CIs (and matched stimuli from this same category allocated in a counterbalanced fashion would later provide comparison novel CSs in the retardation stage).

---

[1] Negative images numbers: 2800, 2095, 3102, 3170, 3301, 3350, 9040, 9410, 9570, 9635. Positive images numbers: 1440, 1750, 1811, 1920, 2040, 2160, 2395, 7330, 8380, 8496.

They were photo-shopped into the CS image and were each consistently paired with the same compatible CS, so that the CI did not look out of place within the scene (see Figure 1).

Learning was measured as changes in CS valence ratings using a rating scale presented at the bottom of the computer screen: 9 (positive); 5 (neutral); 1 (negative).

---- Figure 1 about here ----

All stimuli were presented on the screen of a personal computer using E-Prime (version 1.1) software. The computer was positioned at eye level, approximately 0.5m away from the participant. The computer comprised a PC with a screen, keyboard and mouse.

2.1.3. Procedure

Table 1 details the pre-discrimination and discrimination training stages of the conditioned inhibition task. The identities of the CS stimuli were counterbalanced across both positive and negative US outcomes. However, each CS was consistently paired with the same CI in order to ensure that the CI was appropriately sized and embedded at an appropriate location and orientation (Figure 1).

Table 1. *The design of the task*

Pre-Discrimination		Discrimination Training	
CS	US	CS	US
CS1	Positive image	CS1	Positive image
CS2	Negative image	CS2	Negative image
		[CS1+CI1]	Off-white screen
		[CS2+CI2]	Off-white screen

Legend: CS1 and CS2 were the two street scene images used in a counterbalanced fashion to predict the US outcomes. In total 20 IAPS images from the IAPS database were used to provide the positive and negative USs; 10 of each category of outcome. CI1 and CI2 were the images of street furniture, which were photoshopped into street scenes CS1 and CS2 respectively and used on non-reinforced trials in which IAPS image presentations were replaced with a 'no US' off-white screen.

All instructions were presented on a white background in black text (font Courier New, point size 17) in bold and positioned in the centre of the screen. Instructions remained in place until the participant pressed the mouse. Each trial was separated by an inter-trial-interval of 1000 ms during which a grey screen was presented. The inter-stimulus-interval was represented by a brief white screen.

Pre-discrimination

Instructions advised the participants that they would be presented with a series of images and that they were required to rate them using the rating scale (1-9) which was presented simultaneously at the bottom of the screen. All stimuli were presented on a white screen with the image aligned in the centre of the screen. A CS would appear on the screen and remain on until the participant had rated it. A US would then appear on the screen and remain on until the participants had rated it. There were 20 CS→US trials in total, 10 for each category of outcome.

Discrimination training

A CS or [CS+CI] image would appear on the screen and remain on until the participant had completed the rating scale. Instructions informed the participants to use the rating scale to predict the positive to negative valence level of the image they thought would appear next. If a CS was presented, an image from the corresponding IAPS category would follow. On inhibited [CS+CI] trials, the absence of a US was indicted by the 'no US' off-white screen. The corresponding US or No US screen remained in place until the participants had rated it. Instructions informed the participants to rate this image based on how they felt about it (the positive to negative valence level). The ratio of reinforced to non-reinforced presentations was 2:3; specifically there were 8 CS→US trials and 12 [CS+CI] →'no US' trials. This 2:3 ratio was used to balance the total number of trials for CS and [CS + CI] over the discrimination and pre-discrimination stages. Instructions at this stage directed participants to guess what would follow when uncertain.

2.1.4. Design

All data were analysed using SPSS (version 15.0) with an alpha of $p < 0.05$. Paired samples t-tests used a 95% confidence interval. The dependent variables were the ratings of the images provided by the participants. Both the CS and US ratings were analysed in Experiment 1, to check that the US images were consistently rated as positive versus negative.

The pre-discrimination data were entered into a 2 x 10 within subjects ANOVA with factors valence (positive and negative) and trials (1-10). For the discrimination learning analyses, the CS and [CS + CI] ratings data were entered into a 2 x 2 x 8 within subjects ANOVA with factors inhibition (CS vs. [CS+CI]), valence (positive vs. negative) and trials (1-8). As explained above, there was an uneven number of CS and [CS+CI] trials during discrimination learning, 8 and 12 respectively. Therefore, for comparison by ANOVA, the last 8 [CS + CI] trials (trials 4-12) were compared with the 8 CS trials. This procedure of excluding the early trials on which the CI was first introduced also minimised the likelihood that responding on these trials was reduced because the novel configuration including the CI was distracting for the participants. For the US ratings, the data were entered into a 3 x 8 within subjects ANOVA with factors valence (positive, negative and the neutral 'no US' screen) and trials. As per the [CS + CI] ratings the corresponding last 8 trials were used for analysis, to confirm that the positive versus negative US ratings were maintained for the duration of the task.

Due to the design of the experiment, (changes in) the ratings given to the images presented only reflect learning if they interact with trial type. In other words, for discrimination learning to be demonstrated a significant interaction between inhibition and valence is required. If the discrimination is not learned, further analyses are superfluous.

2.2. Results

US ratings

ANOVA confirmed that the US images were rated according to the IAPS valence categories from which they were selected. There was a significant main effect of valence at both the pre-discrimination, $F_{(1,23)} = 180.068$, MSE = 3718.533, $\eta^2 = .887$, $p < .001$, and discrimination learning stage of the experiment, $F_{(2,46)} = 35.307$, MSE = 1084.751, $\eta^2 = .692$, $p < .001$. As would be expected, the positive IAPS US pictures were rated more positively than the negative IAPS US pictures, minimum t(23) = 13.419, p < .001 (pre-discrimination: positive = 7.371 ± .219, negative = 1.804 ± .234, discrimination training: positive = 6.802 ± .433, negative = 2.081 ± .345). Confirming that US ratings were stable over time, there were no significant effects involving trials.

CS ratings: Pre-discrimination

There was a significant main effect of valence, $F_{(1,23)} = 7.854$, MSE = 117.513, $\eta^2 = .255$, $p = .010$. The CS associated with a positive IAPS US was rated overall higher (5.975 ± .210) than

9

the CS associated with a negative IAPS US (4.985 ± .297). There were no other significant main effects or interactions, maximum $F_{(9,207)} = 1.551$, for the main effect of trials. The fact that there was no valence by trials interaction means that there was no significant learning over the pre-discrimination stage which presented participants with 10 trials per category of outcome.

CS ratings: Discrimination stage

Importantly, there was a significant interaction between valence and inhibition, $F_{(1,23)} = 9.224$, MSE = 58.521, $\eta^2 = .286$, $p = .006$. Figure 2 shows that, as expected, based on what participants predicted would appear next, they rated the CS negative lower than the corresponding [CS+CI] compound, $t_{(23)} = 3.470$, $p = .002$. However, the difference in the ratings of the CS positive and the corresponding [CS+CI] was not significant. There were no other significant main effects or interactions, maximum, $F_{(7,161)} = 1.261$, for the interaction between valence and trials. Because the discrimination was not fully learned, the data were not further analysed.

---- Figure 2 about here ----

2.3. Discussion

Analysis of the discrimination learning trials provided evidence that participants learned the discrimination between the CS and compound [CS+CI] presentations but only for the negative US outcomes. Although the IAPS stimuli have been categorised by valence on the basis of a very large sample of ratings (Lang et al., 2005), the positive IAPS pictures are generally viewed as more subjective and less arousing. In the present experiment, a relatively high proportion of participants (approximately 15) commented that they found some of the 'positive' IAPS US pictures less salient than the 'negative' IAPS US pictures; for example, an ice cream cone may not be rated as positive by someone who is dieting or who does not like ice cream. Therefore, the positive IAPS pictures were removed from the discrimination training stage of Experiment 2, to simplify the design and so that participants would focus on the discrimination based on the negative images.

3. EXPERIMENT 2

3.1. Methods

This was run using the same procedures as Experiment 1. Additionally, there was a retardation stage. The further training given at the retardation test stage was 'incongruent' in the sense that the stimulus previously established as CI now predicted the opposite of what had been the case during training. To the extent the discrimination was learned, at the retardation stage the CI under test now predicted the very outcome for which it had previously signalled the omission. Thus retardation testing would be expected to elicit the opposite affective response (conditioned aversion for the negative images) to that which would have been generated by the inhibitory training (conditioned relief, as on inhibitory trials no aversive outcome followed). An extinction stage provided a further measure of the strength of the previously learned associations.

Table 2. *The stages of the conditioned inhibition task*

Pre-Discrimination		Discrimination Training		Retardation Test		Extinction Test
CS	US	CS	US	CS	US	CS
CS1	Negative	CS1	Negative			
		[CS1+CI]	no US	CI	Negative	CI
				CS2	Negative	CS2
				CS3	Positive	CS3
				CS4	Positive	CS4

Legend: The design across the four stages of the retardation task variant. The numbers refer to the different CS or CI stimuli used in the task – street scene images and street furniture respectively. Ten negative IAPS images provided the US at the discrimination learning stage. There was one CI image, plus three novel CS images introduced at the retardation stage. Additionally ten positive US IAPS images were introduced at the retardation stage to act as filler trials.

3.1.1. Participants

A total of 60 participants were recruited as for Experiment 1. There were 19 males and 41 females with a mean age of 19 (range from 18-55 years).

3.1.2. Materials

The stimuli were the same as in previous experiments. For the discrimination learning stage, all positive stimuli were removed from the task so only the 10 negative IAPS images were used. Conditioned inhibition was tested via retardation test with 10 negative IAPS images; 10 positive IAPS images were introduced at this stage to act as filler trials.

3.1.3. Procedure

Pre-discrimination

Similar to Experiment 1, participants were presented with 10 CS→US trials and were required to rate the images on a scale of 1-9.

Discrimination

As in Experiment 1, participants were presented with 12 CS→US trials and 8 [CS+CI] →'no US' trials. They were required to rate the images on a scale of 1-9.

Retardation

Instructions informed the participants that they would be presented with a series of images and that they needed to rate them using the rating scale (1-9) that would appear at the bottom of the screen simultaneously. All stimuli were presented on a white screen with the image aligned in the centre of the screen. Three novel CSs were provided by the same style of street furniture stimuli as the CIs. Thus retardation could be assessed relative to the rate of acquisition with a matched stimulus predicting the same US outcomes. The novel CS or disembodied CI would remain on the screen until participants predicted, by clicking on the rating scale, what image they thought would next be presented. Since the retardation test examines the rate of acquisition with a stimulus formerly trained as an inhibitor, both the CS and CI were now followed by a US (selected from the same pool of 10 images used in the discrimination training stage). After participants predicted what would come next, a US would appear on the screen for 1000 ms. There were in total 20 trials for each stimulus type: the previously trained CI now being presented as a CS; the novel CS paired with a negative US. Additionally two novel CS were paired with positive USs; these were introduced as filler trials to ensure participants were actively engaged with the task demands. Thus in total this stage comprised of 80 trials.

Extinction

Instructions and presentation of the stimuli on the screen were presented as per the discrimination training and retardation test stages. The novel CS (introduced at the retardation stage) or the disembodied CI were presented individually on the screen and remained on the screen until the participants had rated the stimulus. Participants were asked to use the CS or disembodied CI to predict (using the rating scale) what would come next. However, no US was presented at this stage. Participants completed 20 trials using each of the stimuli: the previously trained CI incongruently transferred and now being presented as a CS and the novel CS, a total of 40 trials at this stage. In total, the full version of the task took approximately 25 minutes to complete.

Awareness check

Finally, in addition to the trial-by-trial expectancy measures of the test stage, participants were asked, at the end of the task, if they could explain to the experimenter what predicted the appearance of the negative images on the computer screen.

3.1.4. Design

All data were analysed using SPSS (version 15.0) with an alpha of $p < 0.05$. Paired samples t-tests used a 95% confidence interval. Data were analysed for the pre-discrimination, discrimination, retardation and extinction stages. Due to the design of the experiment, valence ratings only reflect learning if they are different on inhibited and non-inhibited trials. Specifically, for conditioned inhibition to be confirmed by the retardation test would require a significant interaction between inhibition and blocks. In each case, the dependent variable was the ratings of the images provided by the participants.

The pre-discrimination CS ratings were entered into a one-way ANOVA with the repeated measures factor of trials (1-10). For the discrimination learning analyses, the CS data were entered into a 2 x 8 within subjects ANOVA with factors inhibition (CS vs. [CS+CI]) and trials (4-12). The last eight trials were used for the analyses, as per the rationale outlined in Experiment 1.

The data from the retardation and extinction phase were blocked into five blocks of four trials. These data were entered into a 2 x 5 within subjects ANOVA with factors, inhibition (CI vs. CS) and blocks (1-5). Data from the first 8 trials were further analysed using a 2 x 8 within subjects ANOVA with factors, inhibition (CI vs. CS) and trials (1-8). Paired samples

t-tests were used to examine trial-by-trial differences between CI vs. CS ratings over the first 8 trials.

3.2. Results

Pre-discrimination

<u>CS ratings:</u> There was a significant main effect of trials, $F_{(9,531)} = 3.460$, MSE = 0.582, $\eta^2 = .055$, $p < .001$. Over the ten trials there were non-systematic fluctuations but generally the participants rated the CS images positively (trial 1 = 6.200 ± .216, trial 10 = 6.433 ± .194) and these positive ratings were not depressed by the negative US images, which followed. A one sample t-test showed that the ratings values were overall significantly different from 5 – the value assigned to reflect neutral ratings, $t_{(59)} = 7.083$, $p < .001$.

Discrimination training

<u>CS versus [CS+CI] ratings:</u> There was a significant main effect of inhibition, $F_{(1,59)} = 97.647$, MSE = 954.009, $\eta^2 = .623$, $p < .001$. The CS images associated with the negative US pictures were rated lower (2.942 ± .217) than the [CS+CI] compound which was not reinforced (4.935 ± .150).

There was a significant interaction between inhibition and trials, $F_{(7,413)} = 2.853$, MSE = 4.376, $\eta^2 = .046$, $p = .033$. There were non-systematic fluctuations over the 10 trials but generally the CS images were rated as negative (trial 4 = 3.483 ± .322, trial 12 = 2.817 ± .263) and the [CS+CI] compound images were rated as neutral (trial 4 = 4.750 ± .203, trial 12 = 5.033 ± .176). There were no other significant effects maximum $F_{(7,413)} = 1.098$, for the main effect of trials.

Retardation stage

<u>CS versus CI ratings:</u> There was a significant main effect of inhibition, $F_{(1,59)} = 7.877$, MSE = 15.925, $\eta^2 = .118$, $p < .001$. This arose because the previously trained CI was rated overall more positive (2.636 ± .129) than the novel CS (2.310 ± .153), t(59) = 2.807, p<.05. There was also a significant main effect of blocks, $F_{(4,236)} = 89.207$, MSE = 96.375, $\eta^2 = .730$, $p < .001$. Figure 3A shows that both the previously trained CI and the novel CS were rated more negatively as training progressed (block 1 = 4.05 ± .119, block 5 = 1.931 ± .143). Importantly, there was a significant interaction between inhibition and blocks, $F_{(4,236)} = 16.741$, MSE = 12.061, $\eta^2 = .226$, $p < .001$. Inspection of Figure 3A suggests that this interaction arose because the ratings of the CS and the previously trained CI occurred at

different rates. Indeed, consistent with a difference in learning rates, the interaction between inhibition and blocks was also significant in the linear trend $F_{(1,59)} = 20.478$, MSE $= 28.137$, $\eta^2 = 0.258$, $p < .001$. Furthermore, consistent with the view that inhibitors acquire emotional properties, the initial ratings were different in that the stimuli previously trained as inhibitors for negative outcomes were rated more positively than matched neutral stimuli at the start of acquisition.

---- Figure 3 about here ----

This observation was confirmed statistically in that the initial block one ratings for the previously trained CI and the novel CS were significantly different, $t_{(59)} = 6.014$, $p < .001$. For both the CS and the previously trained CI, the drop in the ratings reached significance only between blocks one and two, $t_{(59)} = 5.835$, $p < .001$, and $t_{(59)} = 9.440$, $p < .001$, respectively. However, as might be expected given the difference in baseline, Figure 3B shows that the drop from block one to two was greater for the previously trained CI. Therefore, a more focused analysis was carried out on the first eight trials (first two blocks) of the retardation stage.

On the trial-by-trial analysis, there was a significant main effect of inhibition, $F_{(1,59)} = 16.616$, MSE $= 114.817$, $\eta^2 = .220$, $p < .001$. The previously trained CI images were rated as overall more neutral ($3.506 \pm .114$) compared to the novel CS ($2.773 \pm .175$). There was a significant main effect of trials, $F_{(7,413)} = 55.094$, MSE $= 240.354$, $\eta^2 = .743$, $p < .001$, as both the previously trained CI and the novel CS were rated progressively more negatively over the first eight trials (trial 1 $= 6.275 \pm .231$, trial 8 $= 2.142 \pm .185$). Importantly, there was a significant interaction between inhibition and trials, $F_{(7,413)} = 10.642$, MSE $= 27.386$, $\eta^2 = .523$, $p < .001$. This suggests a difference in the profile of ratings of the CS and the previously trained CI over the first eight trials (see Figure 3B). Consistent with a difference in learning rates on a trial-by-trial basis, the interaction between inhibition and trials was also significant in the linear trend $F_{(1,59)} = 38.423$, MSE $= 140.667$, $\eta^2 = 0.394$, $p < .001$. For the previously trained CI now presented as a CS, there was a significant difference in the ratings between trial one and two ($t_{59} = 6.78$, $p < .001$), trial two and three, ($t_{59} = 5.37$, $p < .001$) and trial four and five ($t_{59} = 2.12$, $p = .038$). For the novel CS there was a significant difference in the ratings between trial one and two ($t_{59} = 5.237$, $p < .001$) and trial two and three, ($t_{59} = 2.248$, $p = .028$) but there was no difference for any later trials. There were no other significant differences by t-test. Participants were still rating the previously trained CI - now

presented as a CS - differently, specifically more negatively, by trials five and six. This suggests that they were still learning about the previously trained CI, at a point where the ratings of the novel CS were showing no further significant change. These differences in the ratings suggest that the rate of acquisition was different for the two stimuli. More specifically, participants were slower to learn about a previously trained CI now presented as a CS compared to the rate of acquisition seen with a novel CS.

Extinction stage

<u>CS and CI ratings:</u> There was a significant main effect of inhibition, $F_{(1,59)} = 13.789$, MSE = 18.375, $\eta^2 = .194$, $p < .001$. The previously trained CI images (both CI images) were rated higher (2.245 ± .156) than the novel CS images (both CS images) (1.895 ± .157). There was also a significant main effect of blocks, $F_{(4,236)} = 29.592$, MSE = 10.891, $\eta^2 = .383$, $p < .001$. There were non-systematic fluctuations but overall both the previously trained CI and the novel CS images were progressively rated as negative over the blocks (block 1 = 2.594 ± .110, block 5 = 2.023 ± .177). Importantly, there was a significant interaction between inhibition and blocks, $F_{(4,236)} = 48.993$, MSE = 15.751, $\eta^2 = .524$, $p < .001$. The previously trained CI was rated less positively (block 1 = 3.412 ± .13, block 5 = 2.042 ± .194) whereas the novel CS was consistently rated as negative over the blocks (block 1 = 1.775 ± .132, block 5 = 2.004 ± .187).

Awareness check

Out of the 60 participants tested, 51 reported that they were aware of the contingencies. These participants correctly identified which street furniture image was associated with a negative US at the third stage of the task (retardation stage). Out of the other 9 participants, four reported that they had no awareness at all, two reported they were not aware of the contingencies but thought the task was about the stimuli getting progressively more unpleasant, one thought there was a 50/50 chance of a negative IAPS picture appearing on the screen and that it was completely random, and two participants thought there was a pattern to the sequence of IAPS pictures (e.g., the same negative image presented twice followed by another negative image presented twice).

3.3. Discussion

Participants rated the CS associated with a negative IAPS image outcome lower (more negatively) than the compound presentations of the same CS with its respective CI. Thus the discrimination between CS and [CS+CI] was clearly learned in Experiment 2. Moreover, the demonstration of conditioned inhibition was confirmed by the retardation test method in that participants rated the previously trained CI higher (more positively) than the novel CS at the retardation stage. Furthermore, when the results were further analysed it was shown that there was some evidence of significant further learning by the fourth to fifth trial in the case of the previously trained CI. Learning about the novel CS had reached asymptote by trial three in that there was no evidence of any further learning on later trials. This result - that learning was slower for the previously trained CI compared to a novel CS - demonstrates that the inhibitor was a true inhibitor in that its previously acquired inhibitory properties transferred over into the retardation test stage (Rescorla, 1969). These data suggest that this retardation of new learning was at least in part attributable to the affective responses generated at the discrimination learning stage of the experiment. More specifically, training participants with the contingency CS $\rightarrow$ US Negative, [CS+CI] $\rightarrow$ 'no US', resulted in the development of more positive ratings of the CI which signalled the absence of a negative outcome. This was revealed by the higher (more positive) ratings given to the previously trained CI at the start of the retardation stage.

However, since the use of the retardation test cannot exclude the possibility that reduced learning about the CI may also result from reduced attention at the discrimination learning stage, Experiment 3 was conducted using the Experiment 2 task protocol, developed to examine inhibitory learning using a summation test design based on that used previously (Kantini et al., 2011a, 2011b; Migo et al., 2006). The only difference in the experimental training procedures of Experiments 2 versus Experiment 3 of the present study was the addition of an alternative CS to the discrimination stage (CS$_t$, the transfer stimulus, not paired with the CI during training) and the adaptation of the test procedure. At the summation test, the CS$_t$ was presented alone and also together with the inhibitor [CS$_t$+CI]. A further stimulus introduced at the summation test, to which participants would be expected to generalise their excitatory learning (S$_g$) had similar features to the pre-trained CS (Kantini et al., 2011a, 2011b; Migo et al., 2006). The S$_g$ was also paired with the inhibitor at the summation test stage [S$_g$+CI]. If inhibition has truly been demonstrated with the present procedures,

responding to CS_t and S_g should be higher (reflecting reduced expectation of a negative outcome) on trials when these stimuli are compounded with the inhibitor.

4. EXPERIMENT 3

4.1. Methods

4.1.1. Participants

A total of 12 participants were recruited using the same procedures. This number of participants was based on the effect sizes seen in Experiments 1 and 2. There were five males and seven females with a mean age of 35 (range 25-58 years).

4.1.2. Materials

The images used for the stimuli were the same as in the previous experiments, with the two following exceptions. An additional stimulus was introduced at the discrimination stage, CS_t, the transfer stimulus which was paired with the negative IAPS image but not presented together with the CI. A second additional stimulus, S_g, the generalised stimulus, was introduced at the summation stage. S_g was similar to the earlier trained stimuli but had not been presented in the discrimination stage.

4.1.3. Procedure

The pre-discrimination stage used identical stimuli and instructions to those used in Experiments 1 and 2. The discrimination stage was adapted to include the presentation of an additional stimulus (CS_t); this was paired with negative IAPS images as the US but never compounded with the CI at the discrimination learning stage. Minus trials, in which the off white 'No US' screen was presented alone, were also included at the discrimination stage in order to weaken any direct association between the CI and the absence of the US outcome (Migo et al., 2006) (see Table 3). At the summation stage the participants were presented with the CS_t, and S_g followed by a US negative, and the [CS_t+CI], and [S_g+CI] followed by the 'No US' off white screen. At the extinction stage the CS_t, [CS_t+CI], S_g, and [S_g+CI] images were presented without the US. Table 3 shows the four stages of the Experiment 3 summation test variant.

Table 3. *The stages of the Experiment 3 summation test variant*

Pre-Discrimination		Discrimination Training		Summation Test		Extinction Test
CS	**US**	**CS**	**US**	**CS**	**US**	**CS**
CS	Negative	CS	Negative	CS_t	US Negative	CS_t
		CS_t	Negative	S_g	US Negative	S_g
		[CS + CI]	No US, off white screen	$[CS_t+CI]$	No US, off white screen	$[CS_t +CI]$
			Minus trial, No US, off white screen	$[S_g+CI]$	No US, off white screen	$[S_g +CI]$

Legend: The design across the four stages of the summation task variant. Ten negative US IAPS images provided the US. Minus trials were introduced at the discrimination training stage to weaken any direct association between the CI and the absence of a US image. The transfer stimulus (CS_t) was not paired with the CI during training; at the summation test, the CS_t was presented alone and also together with the inhibitor $[CS_t+CI]$. An additional stimulus with similar features to the pre-trained CS, to which participants would be expected to generalise their excitatory learning (S_g) was also introduced at the summation test stage, and was also presented (in this case unreinforced) together with the inhibitor at the summation test stage $[S_g+CI]$.

At the summation stage the task instructions explained that participants should use the rating scale presented on the screen with the designated cues to predict the valence of the images to follow . They were told that the images presented would be followed by negative images or their absence (the 'No US').They were asked to rate all of the images (including the 'No US'). The instructions and procedural details were otherwise the same as those used in Experiments 1 and 2. At the pre-discrimination and discrimination stages there were 10 trials for each of the CS and CS_t, followed by the US. At the summation test and extinction stages there were 10 trials involving each stimulus type and US; one US was presented from a collection of US images. This version of the task took approximately 15 minutes to complete.

4.1.4. Design

All data were analysed using SPSS (version 15.0) and used an alpha of $p < 0.05$ and paired samples t-tests used a 95% confidence interval.

The pre-discrimination CS ratings were analysed by one-way ANOVA with the repeated measures factor of trials (1-10). For the discrimination learning analyses, the CS ratings were entered into a 2 x 10 within subjects ANOVA with factors, inhibition (CS, [CS+CI]) and trials (1-10). The CS data for the summation and extinction stages were entered into a 2 x 2 x 10 within subjects ANOVA with factors, inhibition (presence or absence of CI), stimulus type (CS_t, transfer, S_g, generalised) and trials (1-10).

4.2. Results

Pre-discrimination

<u>CS ratings:</u> There was no significant effect of trials $F(9,99) = 1.355$, MSE = .501, $\eta^2 = .500$, $p = .219$. Thus, as in Experiments 1 and 2, there was no evidence of any learning over the ten conditioning trials prior to discrimination training.

Discrimination training

<u>CS versus [CS+CI] ratings:</u> There was a significant main effect of inhibition, $F(1,11) = 12.911$, MSE = 256.267, $\eta^2 = .540$, $p = .007$. This arose because the CS image was being rated lower (more negatively than the [CS+CI] compound which was being rated as neutral (not different from 5) (Figure 4). There were no other significant main effects or interactions, maximum $F(9,99) = 1.921$, MSE =1.609, $\eta^2 = .750$, $p = .057$, for the interaction between inhibition and trials.

---- Figure 4 about here ----

Summation test

<u>CS_t /S_g versus [CS_t+CI]/[S_g+CI] ratings:</u> There was a significant main effect of inhibition, $F(1,11) = 33.274$, MSE = 529.200, $\eta^2 = .752$, $p < .001$. Importantly, there were no significant effects involving trials, which would suggest further learning at this stage, nor were there any significant effects involving stimulus type, maximum $F(1,11) = 3.498$, MSE = 5.633, $\eta^2 = .241$, $p = .199$, for the interaction between inhibition and stimulus type indicating both the transfer and generalised stimulus passed the summation test. Thus both the CS_t /S_g were rated

overall lower (more negatively) than the corresponding [CS_t+CI]/[S_g+CI] compounds which were rated neutral. (Figure 4).

Extinction stage

CS_t /S_g versus [CS_t+CI/[S_g+CI]] ratings: There was a significant main effect of inhibition, $F(1,11) = 30.449$, MSE = 596.302, $\eta^2 = .735$, $p < .001$. As above, there were no significant effects involving stimulus type or trials, maximum $F(1,11) = 4.160$, MSE = 5.852, $\eta^2 = .274$, $p = .066$, for the interaction between inhibition and stimulus type. Thus both the CS_t /S_g were being rated lower (more negatively) than the [CS_t+CI]/[S_g+CI] compounds which were being rated as neutral (Figure 4).

4.3 Discussion

Experiment 3 showed that conditioned inhibition was also confirmed by the summation test. As expected, participants learned the discrimination between the CS and [CS+CI] compound. At the summation stage, participants rated both the transfer (CS_t) and generalised (S_g) stimuli as negative in comparison to the accompanying [CS_t+CI] and [S_g+CI] compounds, which were rated around neutral. This result - that the inhibitory properties of the CI had transferred over to the CS_t and S_g at the summation stage of the test - further demonstrates that conditioned inhibition has occurred.

5. GENERAL DISCUSSION

The present study tested inhibitory learning using emotionally salient IAPS images. To have matched positive and aversively motivated procedural variants would have been ideal. However, Experiment 1 showed the development of the predicted discrimination learning for negative outcomes but not for positive outcomes. Clear discrimination between CS versus [CS+CI] trials is necessary (although not sufficient) for the demonstration of conditioned inhibition, therefore the negative IAPS images only were used to provide the USs in Experiments 2 and 3. Experiment 2 showed that the predicted conditioned inhibition discrimination was again reliably learned with negative images and there was some evidence of retardation. Furthermore, this reduced learning was clearly related to the conditioned emotional response to the US images; individuals rated transfer images as positive if they had previously signalled the absence of a negative outcome. Experiment 3 showed the conditioned inhibition discrimination was again reliably learned with negative images and in this case conditioned inhibition was confirmed by a summation test. Statistically, the effect

sizes ranged from modest (in Experiment 1) to strong (in Experiments 2 and 3) for the discrimination learning aspect of the task. The Experiment 2 retardation effect was reflected in the interactions between inhibition and blocks as well as between inhibition and trials, which were of modest and moderate effect sizes respectively. The Experiment 3 summation test effect was strong.

Despite uncertainty as to the precise role of attention in excitatory and inhibitory learning, there is reasonable consensus that both retardation and summation tests together provide strong evidence for 'true inhibition', because between them they rule out the most obvious competing explanations as to how a notional CI might detract from a CS. Thus, the two test strategy for testing conditioned inhibition is widely adopted in animal studies (Cole, Barnet, & Miller 1997; Horne & Pearce, 2010; Rescorla & Holland, 1977; Sansa, Rodrigo, Santamaria, Manteiga & Chamizo, 2009; Urcelay, Perelmuter, & Miller, 2008). However, many studies using human participants have used only a summation test (Grillon & Ameli, 2001; He et al., 2011, 2012; Kantini et al., 2011a, 2011b; Karazinov & Boakes, 2004; Migo et al., 2006; Neumann, Lipp, & Siddle, 1997). To our knowledge, there has been only one successful demonstration of conditioned inhibition via both summation and retardation in humans, and this was in a non-standard backward conditioned inhibition procedure (Urcelay et al., 2008). Retardation tests may present a particular challenge in that inhibitors are known to generate opponent processes (Dickinson & Dearing, 1979; Konorski, 1948; 1967; Solomon & Corbit, 1978). Thus, stimuli used as inhibitors in experimental studies start neutral but over time an inhibitor for a salient negative outcome should acquire positivity (Konorski, 1967).

Indeed, in the present study, the fact that participants rated the previously trained CI more positively at the start of the retardation test in Experiment 2 suggests that its affective properties contributed to the retardation of learning. Thus, findings were consistent with the hypothesis that participants should treat the previously trained CI as a safety signal; over time previously neutral stimuli acquire positive properties because they signal the absence of a negative outcome (Cándido, Maldonado, & Vila, 1991; Cicala & Owen, 1976; Dickinson, 1980; Fernando, Urcelay, Mar, Dickinson, & Robbins, 2014; Konorski, 1967; Morris 1975). Participants demonstrated their emotional responses to the stimuli via the ratings scale and the results of these ratings were consistent with the mechanisms proposed to underlie retardation theoretically. The results confirmed that participants had attached emotional relevance to the previously trained CI, this contributed to the way they rated stimuli and in

consequence how they learned about the stimuli subsequently in comparison to novel stimuli (Dickinson & Pearce, 1977; Konorski, 1948; 1967; Konorski & Szwejkowska, 1956).

The inhibitory modulation of affective reactions demonstrated in the present study is consistent with opponent processes theories of inhibitory learning (Dickinson & Dearing, 1979; Konorski, 1948; 1967; Solomon & Corbit, 1978) based on the observation of approach-withdrawal reactions (Hearst et al., 1980; Hearst & Franklin, 1977; Wasserman et al., 1974). These direct behavioural tests of conditioned inhibition are unlikely to be confounded by attentional processes and yield results consistent with opponent processes theories of inhibitory learning (Dickinson & Dearing, 1979; Konorski, 1948; 1967; Solomon & Corbit, 1978). For example, a subject will approach a CS+ for an appetitive outcome such as food. Conversely, presentation of a CI- for an appetitive outcome elicits withdrawal or even avoidance responses (Hearst et al., 1980; Hearst & Franklin, 1977; Wasserman et al., 1974). Such opposing reactions are also consistent with the idea that CIs acquire negative associative strength with respect to a particular category of outcome but introduce additional considerations in relation to the conditioned emotional reaction elicited. In the case of human participants, for the same category of outcome CSs versus CIs would be predicted to be rated differently for emotional valence (Konorski, 1967). Similarly, there is evidence showing that CIs provide safety signals in avoidance learning. Safety signal stimuli generated by the animal's actions, provide feedback confirming the successful execution of the avoidance response and can thus act as secondary reinforcers of this behaviour (Cándido et al., 1991; Cicala & Owen, 1976; Dickinson, 1980; Dinsmoor, 2001; Galvany & Twitty, 1978; Morris 1975). In the case of human participants, CSs and CIs would be predicted to be rated differently for emotional valence (Konorski, 1967). However, studies of human participants have to date focused on measuring associative strength rather than valence ratings of the pre-trained stimuli as here. The valence ratings used in the present study provide direct measures of the reactions elicited by stimuli pre-trained as CIs and compared with (in other respects similar) novel stimuli. At the discrimination stage, as required, participants learned that the CS signalled a negative outcome while the [CS+CI] compound signalled the absence of any such outcome. At the summation stage, participants rated both CS_t and S_g as negative, which reflects the expectation of a negative outcome, whereas they rated the $[CS_t+CI]$ and $[S_g+CI]$ presentations as neutral, indicating an expectation of the absence of any such outcome. The stimulus type (CS_t vs. S_g) had no significant effect on summation test performance. This

result means that the inhibitory properties of the CI had transferred over to both CS_t and S_g, confirming that the CI was a true inhibitor.

Thus, Experiment 3 is consistent with previous research that has demonstrated conditioned inhibition via a summation test not only in human studies (Grillon & Ameli, 2001; He et al., 2011, 2012; Kantini et al., 2011a, 2011b; Karazinov & Boakes, 2004; McNally & Reiss, 1984; Migo et al., 2006; Neumann et al., 1997, Wilkinson, 1989) but also in animal studies (Cole et al., 1997; Horne & Pearce, 2010; Pineño, 2010; Rescorla & Holland, 1977; Sansa et al., 2009; Urcelay et al., 2008). Moreover, conditioned inhibition via summation test was relatively simple to demonstrate in the present study (in comparison with the retardation test method which a required trial-by-trial examination of the rate of learning). Summation was shown both with conventional transfer stimulus (CS_t) and novel matched stimulus to which participants would be expected to generalise their excitatory responding (S_g, not previously paired with the inhibitor). It could be argued that in a summation test too much attention is paid to the CI which therefore distracts from the CS and reduces responding. This is why the two tests are ideally both needed to confirm conditioned inhibition (although some summation tests control for external inhibition by including distractors Kantini et al., 2011a; Kantini et al., 2011b)

To conclude, the task developed over the course of the present study has successfully demonstrated discrimination learning, followed by conditioned inhibition, as confirmed by the retardation and summation tests. To our knowledge these tests have been successfully applied in human studies of inhibitory learning only infrequently, and very seldom have procedures been tested by both methods. The inhibitory properties of an established CI showed the transfer which is held to be typical of a true inhibitor (Grillon & Ameli, 2001; Hearst, 1972; Kantini et al., 2011a; Kantini et al., 2011b; Karazinov & Boakes, 2004; McNally & Reiss, 1984; Migo et al., 2006; Neumann et al., 1997; Rescorla, 1969), passing both retardation and summation tests. This procedure, developed with negative IAPS image outcomes, represents an improvement on earlier inhibitory learning procedures which have been developed for use with human participants. Firstly, because inhibitory learning was demonstrated using a relatively simple discrimination learning procedure, and confirmed by both retardation and summation tests which between them control for alternative explanations of apparent conditioned inhibition. Secondly, because the use of emotionally salient cues as the US more closely resembles the traditional Pavlovian paradigm. Moreover, to our

knowledge the present study provides the first direct test of the inhibitory modulation of affective reactions in humans using procedures suitable for use in a clinical population. Although the present procedures were not purely evaluative, they are consistent with associative accounts of EC and its susceptibility to inhibitory learning, which to our knowledge has not previously been demonstrated (Hofmann et al., 2010).

[8,394 words]

REFERENCES

Cándido, A., Maldonado, A. & Vila, J. (1991). Effects of duration of feedback in signalled avoidance. *Animal Learning Behavior*, *19*, 81-87.

Cicala, G.A. & Owen, J.W. (1976). Warning signal termination and a feedback signal may not serve the same function. *Learning and Motivation*, *7*, 356-367.

Cole, R.P., Barnet, R.C. & Miller, R.R. (1997). An evaluation of conditioned inhibition as defined by Rescorla's two-test strategy. *Learning and Motivation*, *28*, 323-341. doi: 10.1006/lmot.1997.0971.

Dickinson, A. (1980). *Contemporary Animal Learning Theory*. New York: Cambridge University Press.

Dickinson, A. & Dearing, M.F. (1979). Appetitive-aversive interactions and inhibitory processes. In A. Dickinson & R.A. Boakes (Eds.) *Mechanisms of Learning and Motivation*, 203-231. Erlbaum, Hillsdale, New Jersey: Lawrence Erlbaum Associates Inc Publishers.

Dickinson, A. & Pearce, J.M. (1977). Inhibitory interactions between appetitive and aversive stimuli. *Psychological Bulletin, 84*, 690-711.

Dinsmoor, J.A. (2001). Stimuli inevitably generated by behavior that avoids electric shock are inherently reinforcing. *Journal of Experimental Analysis of Behavior*, *75*, 311-333.

Fernando, A.B.P., Urcelay, G.P., Mar, A.C., Dickinson, A. & Robbins, T.W. (2014). Safety signals as instrumental reinforcers during free-operant avoidance. *Learning & Memory*, *21*, 488-497, doi: 10.1101/lm.034603.114

Galvany, P.F. & Twitty, M.T. (1978). Effects of inter-trial interval and exteroceptive feedback duration on discriminative avoidance acquisition in the gerbil. *Animal Learning & Behaviour*, *6*, 166-173.

Grillon, C. & Ameli, R. (2001). Conditioned inhibition of fear-potentiated startle and skin conductance in humans. *Psychophysiology*, *38*, 807-815.

Grings, W.W., Carey, C.A. & Schell, A.M. (1974). Comparison of two methods for producing response inhibition in electrodermal conditioning. *Journal of Experimental Psychology*, *103*, 658-662.

He, Z., Cassaday, H.J., Howard, R.C., Khalifa, N. & Bonardi, C.M. (2011). Impaired Pavlovian conditioned inhibition in offenders with personality disorders. *Quarterly Journal of Experimental Psychology, 64*, 2334-2351.

He, Z., Cassaday, H.J., Park, S.B.G. & Bonardi, C. (2012). When to hold that thought: An experimental study showing reduced inhibition of pre-trained associations in schizophrenia. *PLoS ONE*, *7*(7), e42175.

He, Z., Cassaday, H.J., Bonardi, C. & Bibby, P.A. (2013). Do personality traits predict individual differences in excitatory and inhibitory learning? *Frontiers in Psychology*, 4, 245.

Hearst, E. (1972). Some persistent problems in the analysis of conditioned inhibition. In R.A. Boakes and M.S. Halliday (Eds.), *Inhibition and* learning (pp. 5-39). London: Academic Press.

Hearst, E., Bottjer, W.S. & Walker, E. (1980). Conditioned approach-withdrawal behaviour and some signal-food relations in pigeons: Performance and positive negative "associative strength". *Bulletin of the Psychonomic Society, 16*, 183-186.

Hearst, E. & Franklin, S.R. (1977). Positive and negative relations between a signal and food: Approach-withdrawal behaviour to the signal. *Journal of Experimental Psychology: Animal Behaviour Processes*, *3*(1)*,* 37-52.

Heym, N., Kantini, E., Checkley, H.L.R. & Cassaday, H.J. (2014). Tourette-like behaviors in the normal population are associated with hyperactive/impulsive ADHD-like behaviors but do not relate to deficits in conditioned inhibition or response inhibition. *Frontiers in Psychology,* 5, 946.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens F. & Grombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*, *136*, 390-421.

Horne, M.R. & Pearce, J.M. (2010). Conditioned inhibition and superconditioning in an environment with a distinctive shape. *Journal of Experimental Psychology: Animal Behavior Processes, 36*, 381-394.

Kantini, E., Cassaday, H.J., Batty, M.J., Hollis, C. & Jackson, G.M. (2011a). Associative learning in ADHD: improved expression under methylphenidate. *Open Journal of Psychiatry*, *1*, 20-29.

Kantini, E., Cassaday, H.J., Hollis, C.P. & Jackson, G.M. (2011b). The normal inhibition of associations is impaired by clonidine in Tourette Syndrome. *Journal of the Canadian Academy of Child and Adolescent Psychiatry, 20*, 96-106.

Karazinov, D.M. & Boakes, R.A. (2004). The effectiveness of inhibitors in human predictive judgments depends on the strength of the positive predictor. *Learning and Behavior*, *32*, 348-359.

Konorski, J. (1948). *Conditioned reflexes and neuron organization*. Cambridge Massachusetts: Cambridge University Press.

Konorski, J. (1967). *Integrative activity of the brain*. Chicago: Chicago University Press.

Konorski, J. & Szwejkowska, G. (1956). Reciprocal transformations of heterogeneous conditioned reflexes. *Acta Biologica Experimentalis*, *11*, 141-165.

Lang, P. J., Bradley, M. M. & Cuthbert, B. N. (2005). *International Affective Picture System (IAPS): Instruction Manual and Affective Ratings.* Technical Report A-6. University of Florida, Gainesville, FL.

Mackintosh, N.J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*, 276-298.

McNally, R.J. & Reiss, S. (1984). The preparedness theory of phobias: the effects of initial fear level on safety-signal conditioning to fear-relevant stimuli. *Psychophysiology*, *21*, 647 652.

Migo, E.M., Corbett, K., Graham, J., Smith, S., Tate, S., Moran, P.M. & Cassaday, H.J. (2006). A novel test of conditioned inhibition correlates with personality measures of schizotypy and reward sensitivity. *Brain Research*, *27*, 299-306.

Morris, M.G.M. (1975). Preconditioning of reinforcement properties to an exteroceptive feedback stimulus. *Learning and Motivation*, *6*, 289-298.

Neumann, D.L., Lipp, O.V. & Siddle, D.A.T. (1997). Conditioned inhibition of autonomic Pavlovian conditioning in humans. *Biological Psychology*, *46*, 10, 223-233.

Pavlov, I.P. (1927). *Conditioned reflexes.* Oxford: Oxford University Press.

Papini, M.R. & Bitterman, M.E. (1993). The two strategy in the study of inhibitory conditioning. *Journal of Experimental Psychology: Animal Behavior. Processes*, *19*, 342-352.

Pineño, O. (2010). Enhanced consumption of an aversively conditioned taste following the presentation of a "medicine" taste. *Behavioural Processes*, *84*, 625-628.

Pearce, J.M. & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not unconditioned stimuli. *Psychological Review, 87,* 532-552.

Rescorla, R.A. (1969). Pavlovian conditioned inhibition. *Psychological Bulletin, 72*, 77-94.

Rescorla, R.A. (1973). Effect of US habituation following conditioning, *Journal of Comparative and Physiological Psychology*, *82*, 137-143.

Rescorla, R.A. & Holland, P.C. (1977). Associations in Pavlovian conditioned inhibition. *Learning and Motivation*, *8*, 429-447.

Rescorla, R.A. & Wagner, A.R. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and non-reinforcement. In A.H. Black and W.F. Prokasy (Eds.), *Classical conditioning: Current research and theory,* (pp. 64-99). New York: Appleton-Century-Crofts.

Sansa, J., Rodrigo, T., Santamaria, J.J., Manteiga, R.D. & Chamizo, V.D. (2009). Conditioned inhibition in the spatial domain. *Journal of Experimental Psychology Animal Behaviour Processes*, *35*, 566-577.

Solomon, R.L. & Corbit, J.D. (1978). An opponent process theory of motivation. *The American Economic Review*, *68*, 12-24.

Urcelay, G.P., Perelmuter, O. & Miller, R.R. (2008). Pavlovian backward conditioned inhibition in humans: summation and retardation tests. *Behavioural Processes*, *77*, 299-305.

Wasserman, E.A., Franklin, S.R. & Hearst, E. (1974). Pavlovian appetitive contingencies and approach versus withdrawal to conditioned stimuli in pigeons. *Journal of Comparative and Physiological Psychology*, *86*, 616-627.

Wilkinson, G.M., Lovibond, P.F., Siddle, D.A.T. & Bond, N. (1989). Effects of fear-relevance on electrodermal safety signal learning. *Biological Psychology, 28*, 89-104.

Williams, D.A., Overmier J.B. & LoLordo, V.M. (1992). A re-evaluation of Rescorla's dictums about Pavlovian conditioned inhibition. *Psychological Bulletin*, *111*, 275-290.

FIGURE LEGENDS

Figure 1. (A) Example street scene (CS); photoshopped scene to include street furniture [CS+CI]; off-white screen used to signal the absence of any US outcome ('no US'). **(B)** Example of the novel CS stimuli used at the retardation stage (approximately to scale). Stimuli such as benches and letterbox were selected as suitable for insertion in the street scenes to act as inhibitors (CIs) during the discrimination learning phase. The CIs versus novel CSs were counterbalanced and all similar stimuli in the same category.

Figure 2. Discrimination stage valence ratings in Experiment 1, shown for [CS] images followed by positive versus negative outcomes or their absence [CS+CI] shown as the off-white screen. Error bars represent the standard error of the mean.

Figure 3. (A) Valence ratings over successive blocks of presentation of the previously trained CI versus CS images in the Experiment 2 retardation test. The required transfer was incongruent with prior learning in that the CI trained as a signal of the omission of a negative outcome now predicted the presentation of a negative IAPS picture. Error bars represent the standard error of the mean. **(B)** Experiment 2 valence ratings over the first eight trials ratings of the CI and CS images for incongruent transfer of the CI stimulus at the retardation stage of the conditioned inhibition task. Error bars represent the standard error of the mean.

Figure 4. Valence ratings for the discrimination, summation and extinction stages of Experiment 3. At every stage of the procedure, the CS images were rated as more negative compared to the [CS+CI] images which were rated a neutral. Thus the discrimination between the the different stimulus configurations was successfully learned, inhibition transferred in the summation test, and the same difference was preserved at the extinction stage of testing. Importantly, there was no difference between the previously trained excitor (CS_t) and the novel CS (S_g) at the summation stage of the procedure (so these data are shown collapsed across the two CS types). Error bars represent the standard error of the mean.
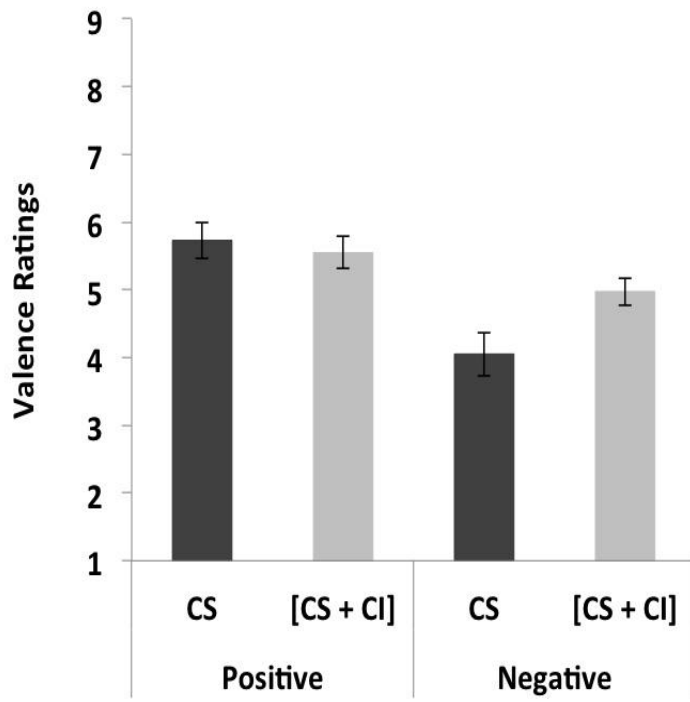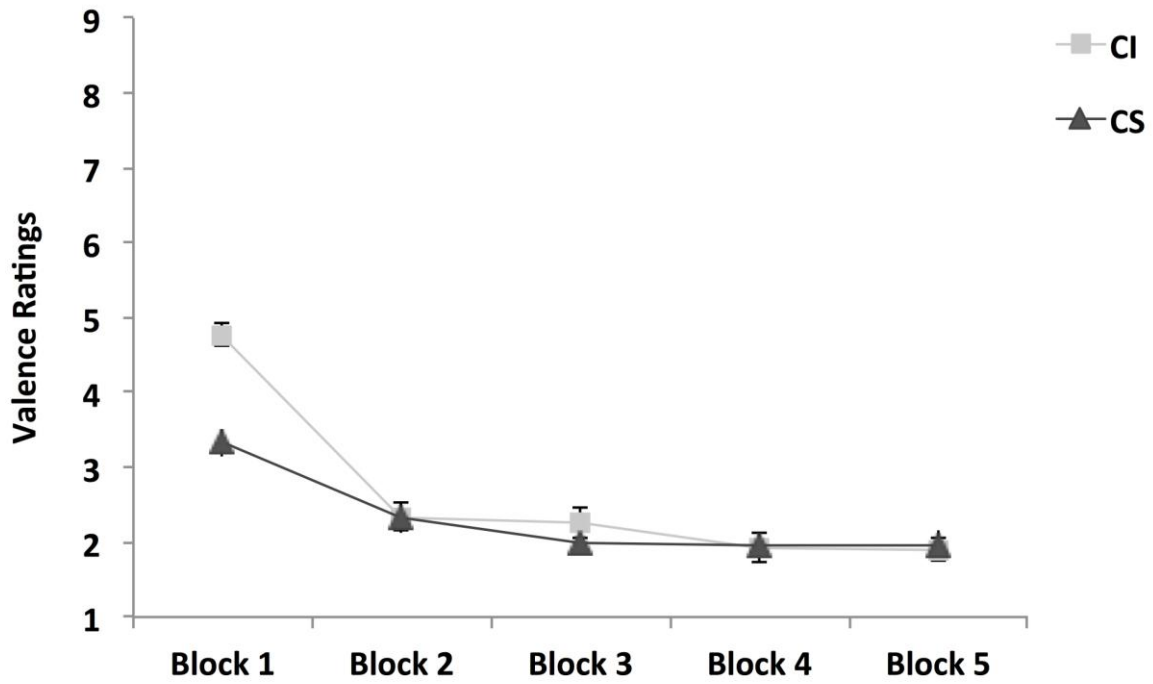
Figure 1

A)

B)

Figure 2
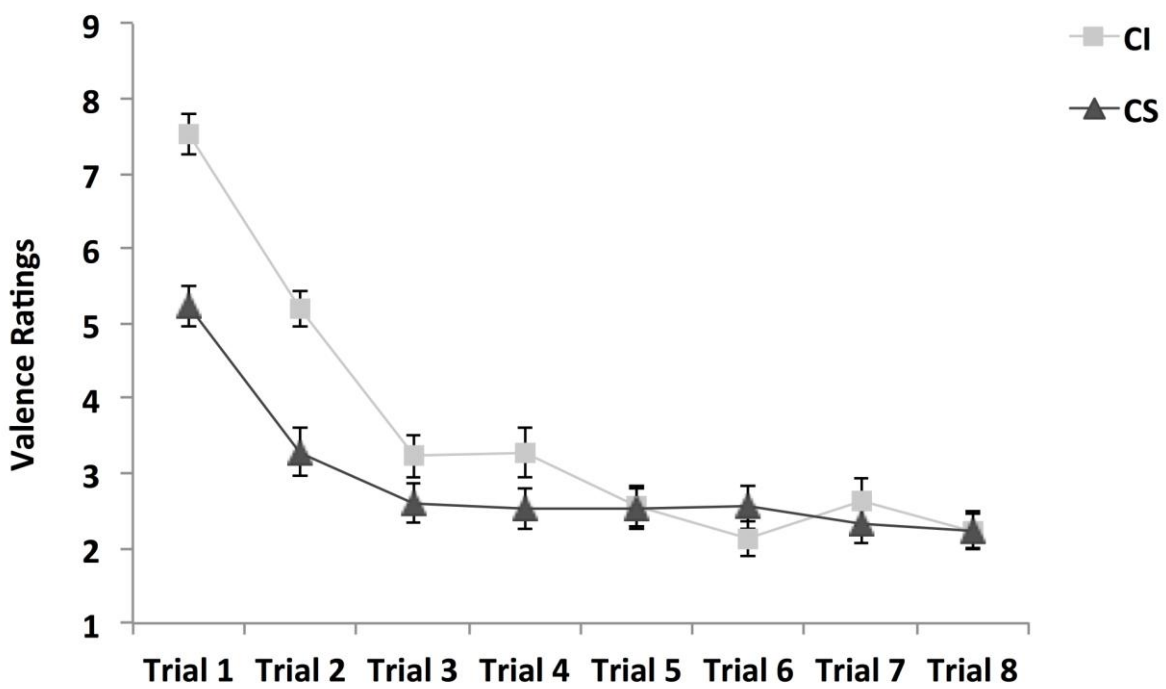
Figure 3A

Figure 3B

Figure 4