University of Southampton

Faculty of Life and Environmental Sciences

School of Psychology

Relationships between Associative and Non-associative Inhibition in Feature Negative and Extinction Preparations

by

Ovidiu Ionuț Brudan

https://orcid.org/0000-0002-6678-1749

Thesis for the degree of Doctor of Philosophy

June 2024

University of Southampton

Abstract

Faculty of Life and Environmental Sciences

School of Psychology

Doctor of Philosophy

Relationships between Associative and Non-associative Inhibition in Feature Negative and Extinction Preparations

by

Ovidiu Ionuț Brudan

Associative learning comprises of a wide range of mechanisms through which an organism can gain an evolutionary advantage by learning about the surrounding environment and adapting to changes. Associative learning is highly flexible allowing for learnt associations to be changed if there are no longer advantageous or become redundant, and two ways of changing the meaning of previously learnt associations is through conditioned inhibition and extinction. The latter is of particular interest for changing the maladaptive mechanisms developed as a result of addiction for example. Conditioned inhibition and extinction rely on inhibition, individual differences in inhibition having major potential implications, however inhibition is not a purely associative construct and has been defined in many different ways in the wider field of Psychology. The current thesis aimed to assess the link between associative and non-associative inhibition, which are assumed to be independent subtypes of inhibition, by using various inhibition measures. For associative inhibition, the speed of feature negative discrimination learning, conditioned inhibition, speed of extinction, and context inhibition were used. For non-associative inhibition, following the structure of inhibition proposed by Bari and Robins (2013), measures of cognitive inhibition, delay discounting, and response inhibition were employed. It was also aimed to assess the effectiveness of using compound extinction techniques in the forms of super-extinction and deepened extinction compared to cue alone extinction. The final aim was to compare which of three formal models of associative learning Rescorla-Wagner, configural Rescorla-Wagner, and Pearce configural model, is best at predicting the observed data in the extinction study where cue alone, super-extinction, and deepened extinction were compared. For the first aim no evidence of a relationship between associative and non-associative inhibition was found, with the potential exception of a link with the Behavioural Inhibition System of the Behavioural Inhibition System/Behavioural Activation System. As a result, it was concluded that associative inhibition is an independent inhibitory construct, unrelated to non-associative inhibition, and therefore should be included as such in future inhibitory models. Regarding the extinction methodology comparisons, it was found that neither super-extinction or deepened extinction resulted in a more stable extinction, in fact super-extinction was found to be more unstable compared to cue alone and deepened extinction. It was concluded that based on the current results, compound extinction was not a reliable method of enhancing extinction. When assessing the predictions of the three formal models of associative learning, the Pearce configural model was found to be the best overall model, however this model was not the best for every participant.

Table of Contents

Research Thesis: Declaration of Authorship

Print name: Ovidiu Ionuț Brudan

Title of thesis: Relationships between Associative and Non-associative Inhibition in Feature Negative and Extinction Preparations

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signature: Date: 06.06.2024

Acknowledgements

I would like to thank my supervisors for all their help and advice throughout the PhD, especially during the uncertain times of the pandemic/lockdown. I am very grateful for all the valuable things they taught me during my studies and I am looking forward to applying this knowledge in my future work. I would also like to thank my friends and family for always supporting me and encouraging me to reach my goals.

Chapter 1 Introduction

The current thesis focuses on addressing three main aims across a series of five studies. The first aim is to assess whether associative inhibition is related to the wider construct of inhibition, referred to as non-associative inhibition throughout the thesis. For this purpose associative inhibition has been measured in various ways including: the speed of feature negative discrimination learning, the speed of extinction, conditioned inhibition, and context inhibition. For non-associative inhibition the focus was on the most widely spread measures of inhibition used in the wider field of Psychology and included: a self-reported measure of impulsivity, a measure of the behaviour inhibition system/behaviour activation system, delay discounting, and the stop signal reaction task.

The second aim focused on investigating different ways of obtaining a faster and more stable extinction. Extinction is one of the associative inhibition phenomena where inhibition plays a large part, and given the importance of extinction in clinical contexts such as addiction, understanding the techniques which lead to a more lasting extinction as well as the role inhibition plays is essential. For this purpose, cue alone extinction was compared with two compound extinction techniques, super-extinction initially and then super-extinction and deepened extinction.

The final aim of the thesis was to further assess the interactions between cues during the above-mentioned extinction techniques using three formal models of associative learning: Rescorla-Wagner, configural Rescorla-Wagner, and Pearce configural model. These models had different a priori predictions regarding the speed of extinction and robustness of the procedures, and their predictions have been evaluated against the observed data obtained in the last study of the series.

1.1 Associative and Non-associative Inhibition

Inhibition can be defined in many different ways depending on the field of study, however without being too specific it can be defined as a construct that employs a set of mechanisms in order to supress or stop certain processes. Given this general definition it is apparent that many forms of inhibition can be proposed as a result of the variety of processes that can be supressed. For example in the field of associative learning, inhibition is defined on the basis of cues, associations, and responses. In the wider field of Psychology inhibition can be defined based on a wide range of behavioural or cognitive processes resulting numerous related or independent inhibition constructs. Associative inhibition is rarely considered and studied alongside non-associative inhibition despite being very similar in definition, the current thesis aims to explore the relationship between the two types of inhibition to determine whether they are independent or related subtypes of inhibition.

1.1.1 Associative Inhibition

Associative learning encompasses a wide variety of mechanisms that allow organisms to learn about their environment and continuously adapt to change, gaining therefore an evolutionary advantage. As a result, a large number of invertebrates display behaviour consistent with associative learning (see Loy et al., 2021 for a full review of organisms which have been observed to be capable of associative learning). Associative learning mechanisms develop through association of stimuli over time as follows: if a neutral stimulus (NS) is repeatedly paired with an unconditioned stimulus (US) which elicits an unconditioned response (UR), the NS is expected to become a conditioned stimulus (CS) which elicits a conditioned response (CR). The CR is often similar (or at least related) to the UR elicited by the US. Therefore, the NS can become a CS capable of evoking new responses which are indicative of US expectation due to the newly formed NS-US association. To put it more simply, take the example of a rat, which is given food (US) after every time a light (NS) turns on. Following multiple repetitions of this pairing, the rat comes to expect food every time the light turns on. In this case, the light, initially a NS, becomes a CS which elicits a CR indicative of food expectation (e.g. approaching the food dispenser) which was previously observed only in the presence of the US.

The key feature that makes associative learning so successful is its flexibility. Once organisms learn about their environment and establish associations based on the stimuli relationships encountered, they are also able to adapt to change, if the associations were to become redundant, maladaptive, or simply sub-optimal. The current thesis focuses on two such mechanisms: conditioned inhibition and extinction. Conditioned inhibition refers to the associative learning phenomenon observed when a stimulus, a conditioned inhibitor, signals the absence of an otherwise expected outcome. Therefore, through conditioned inhibition an organism can behave differently in certain situations. Extinction refers to the process of presenting the CS without the expected outcome, which causes the organism to behave differently than expected. Both phenomena involve inhibitory processes and are discussed in more detail throughout the thesis. In addition, the concept of inhibition is not unique to associative learning and has been discussed at length in other literatures. As a result, the current thesis aims to both examine associative inhibition and look for relationships between associative and non-associative inhibition in an attempt to understand the structure of the construct of inhibition. Associative and non-associative inhibition could be independent subtypes of inhibition or they could rely on common inhibitory mechanisms.

1.1.1.1 Conditioned Inhibition

Conditioned inhibition (CI) is an associative learning mechanism that allows organisms to modify previously learnt responding in order to adapt to changes in the environment. CI relies on the interaction between cues to change the behaviour generated by an existing association. For CI to be established an organism needs to learn that a certain stimulus, a conditioned inhibitor, indicates that an outcome, which would normally be expected, is actually omitted. Going back to the previous example of the rat, let's assume that the rat learnt to expect food every time a light was turned on (A+ trials), making the light a CS (CS A). If on some trials CS A was presented alongside (in compound with) a novel stimulus X, a sound for instance, and the compound was non-reinforced (meaning that the rat did not receive any food when the two stimuli were presented together, an AX− trial) then following multiple presentations of A+ and AX- trials, cue X is expected to become a conditioned inhibitor. As a result, the rat should no longer show signs of food expectancy on AX trials, because X indicates the absence of the otherwise expected food, but the rat is expected to continue to expect food when CS A is presented by itself.

One early theory of conditioned inhibition proposed by Konorski (1948) suggests that the conditioned inhibitor acquires the ability to signal the absence of an event. In the same manner the conditioned stimulus comes to be associated with an unconditioned stimulus in a traditional CS-US association, according to Konorski a CS-no US association can be formed which predicts the non-occurrence of the US. Later models, such as the Rescorla-Wagner model which is discussed in more detail below, focus on inhibition and excitation as polar opposites and learning is driven by expectancy. In the context of conditioned inhibition, the inhibitor would be driven to acquire a negative valence becoming inhibitory as a result of the expectancy violation when the putative conditioned inhibitor is presented in compound with an excitatory cues which is reinforced when presented alone but now in the compound (Rescorla & Wagner, 1972).

To demonstrate the inhibitory properties of a putative conditioned inhibitor following the training described above, a two-test strategy was suggested (Rescorla, 1969). According to this approach, the CS should pass both a summation test and a retardation test. In order to conduct a summation test, the putative conditioned inhibitor is presented in compound with another CS which received reinforced training but which was never paired with the conditioned inhibitor. Building up on the previous example, we can assume that the CS used in the summation test is CS C, and C+ trials were trained intermixed with the A+ and AX− trials. Therefore, a summation test would consist of presenting cues C and X in a CX compound with the expectation that the negative associative value of cue X and the positive associative value of cue C would mathematically summate (Cole et al., 1997; Rescorla & Wagner, 1972). The consequence of this summation would be reduced responding to the CX compound relative to a control stimulus, for example a CN compound in which N had not received inhibitory training. A review of formal associative learning models which describe cue interactions as exemplified in summation tests is provided in the formal associative learning models section.

The second test, the retardation test, consists of multiple reinforced presentations of the conditioned inhibitor (X+) with the aim of assessing the acquisition of the conditioned inhibitor against the acquisition of a novel and neutral stimulus. The acquisition of the X is expected to require more trials to reach an asymptotic level or a previously determined level compared to a novel stimulus, which received no previous training, if X did indeed became a conditioned inhibitor previously. This is because, according to the Rescorla-Wagner model, the conditioned inhibitor is expected to have a negative valence (associative strength) which first needs to reach zero in order for it to continue to develop into an excitatory, positive stimulus, whereas the neutral stimulus would start from zero (Cole et al., 1997).

Rescorla (1969) argued that both summation and retardation tests should be used to determine if a cues became a conditioned inhibitor following training. Using both tests would allow for a clear distinction between conditioned inhibition and behaviour that is controlled by other extraneous factors such as attention. For example, a stimulus which is more salient and attracts more attention has the potential of reducing responding in a summation test, but would consequently be conditioned more rapidly in a retardation test. Conversely, a stimulus that is less salient and attracts less attention might have a slow conditioning during the retardation test but would not reduce responding during a summation test (Rescorla 1969).

In an experiment (Experiment 1) using rats Rescorla and Holland (1977) explored the properties of a conditioned inhibitor trained with a simple A+, AX− procedure, known as feature negative (FN) discrimination. The rats were placed in Skinner boxes and received A+ trials with a low tone followed by an electric shock. During the AX− trials, the low tone was paired with a light and the compound was not followed by a shock. After training, the putative conditioned inhibitor X was tested in compound with cue B, a high tone which had received excitatory (B+) training alongside the feature negative discrimination. The rats were trained to press a bar in the cage in order to receive food. Given that an electric shock was used to reinforce cues, fear was interpreted as shock expectancy, fear being measured as a change in bar pressing behaviour during stimulus presentation (put simply the rat would stop its bar pressing behaviour when cues indicative of shock were presented, the rat showing therefore fear for the expected shock). The rat's bar pressing behaviour prior and during stimuli presentations were used to compute a suppression ratio (during/(during+prior), accordingly a value of 0 indicated no responding while the stimulus was present and a value of .5 was reflective of no changes in responding during the CS presentation compared to prior stimulus presentation levels. During a summation test consisting of B and BX trials, the rats showed less bar pressing during the presentation of B (high suppression ratio) compared to BX which suggested that the inhibitory properties of X had transferred to a different CS – X reduced fear to A as well as to B. Despite the successful transfer effect reported in the first experiment, in the second experiment no transfer between USs was observed, suggesting that the learning that occurs as a result of the feature negative discrimination training could be different from a simple CS-US association (Rescorla & Holland, 1977).

Conditioned inhibition is not a mechanism that is restricted to the discrete cues that are being used in a learning procedure and it can extend to the context in which the learning takes place. If the context in which the learning takes place is salient enough, an organism could encode the context as a cue present in the background. In a series of two experiments using human participants and a game-like learning task Glautier, Elgueta, and Nelson, (2013) showed that a context can become inhibitory. A review of the context-cue interaction in the context of extinction is provided in the extinction section below.

As previously mentioned a feature negative discrimination is generally used to train conditioned inhibition, however slight variations to the training procedures have been shown to lead to different outcomes. The other possible outcome that could result from a feature negative discrimination is discussed next.

1.1.1.2 Occasion Setting

A traditional FN discrimination consists of a series of A+/XA− trials, however multiple variations of this procedure have been used. For example, Holland and Lamarre (1984) tested the effect of including a delay to the presentation of the XA compound in a series of three experiments using a simultaneous (A+, XA−) and sequential (A+, X→A−) feature negative procedure to train conditioned inhibition in rats. In Experiment 1, rats were divided into two groups, one group received simultaneous feature negative discrimination training while the other received serial training where the feature was presented first (X→A−). For the simultaneous training group substantial transfer of inhibition was observed at test, however for the serial group little to no transfer was observed. Experiment 2 aimed to validate these results by reinforcing the excitors during the testing phase to clarify whether the lack of transfer was due to the properties of the occasion setter or simply due to a loss of excitation. Additionally, the second experiment used a within-subjects design meaning that each rat received both serial and sequential training for two using four separate cues (A+, XA−, B+, Y→B−). The results supported the findings of the first experiment and the transfer of inhibition was greater for the cue that received simultaneous training compared to the cue that received serial training. Experiment 3 sought to assess whether the lack of inhibition transfer was due a potential association formed between the two cues used in the feature negative discrimination. The feature negative could gain the ability to evoke an excitatory representation of the excitatory cue it was paired with which could interfere with the transfer test where a different excitatory cue is used. To account for this, in experiment 3 the excitatory cue (A) was extinguished after the feature negative training, however the results were in line with those if the previous to experiments and the serial presentation training group showed less transfer compared to the simultaneous group (Holland & Lamarre, 1984).

The results suggested that the two procedures led to two different learning outcomes. The conditioned inhibitor that resulted from the simultaneous training easily transferred to and supressed responding to another excitatory CS (B); however, the transfer of the “conditioned inhibitor” that resulted from the sequential training was limited. As a result, simultaneous feature negative training was concluded to result in conditioned inhibition, while sequential feature negative training was suggested to result in a new type of learning mechanism: occasion setting (Holland & Lamarre, 1984). This type of stimulus was also called a modulator or a remote initiating stimulus, the main theories that explained the assumed mechanisms through which these operate were the modulation and hierarchical theories. The modulation theory assumes that the occasion setter influences the activation threshold of the US representation, increasing or decreasing sensitivity to excitatory cues. As a result, the CS’s ability to activate the US representation is greater or lower in when the occasion setter is present than when it is absent. The hierarchical theory assumes that the occasion setter influences the association formed between the CS and US and operates on this association (see Bonardi et al., 2017 for full review).

The main difference between a stimulus becoming a conditioned inhibitor or an occasion setter is the mechanism though these are assumed to operate. While a conditioned inhibitor is assumed to be a simple cue which gained inhibitory strengths and stops responding though its direct association with the US, occasion setters are thought to work through a more complex mechanism. An occasion setter is assumed to affect responding independently of its associative properties and the resulting behaviour cannot be elicited by the occasion setter in isolation (Holland, 1989; Ross & Holland, 1981).

Although conditioned inhibition and occasion setting can be trained by almost identical procedures, the two mechanisms are fundamentally different. For the current thesis, the most important distinction between X becoming a conditioned inhibitor and X becoming an occasion setter, is that the inhibitory properties of a conditioned inhibitor are general and can therefore transfer between CSs while the inhibitory properties of an occasion setter are specific to the CS it was trained with. If X is a conditioned inhibitor trained using a feature negative discrimination (A+/AX−), and is presented in compound with B (which was previously reinforced, B+) as part of a summation test, X would be able to supress responding to B. On the other hand, if X is an occasion setter and is presented in compound with B it would not supress responding to B even if it successfully supresses responding to A during training (Holland, 1989, 1992). Several studies however showed that occasion setter transfers can occur, but are usually incomplete (see Bonardi et al., 2017 for full review). An important factor that could contribute to the transfer of occasion setters is thought to be generalisation, transfer between similar cues is assumed to be more likely compared to cues that are different. In the case of occasion setters, transfers between cues that were occasion-set through training and acquired a second order association as a result were more often reported within the literature. Accordingly, through generalisation due to their similar properties, transfers between two occasion-setters might be more likely compared to transfers between an occasion-setter and a cue that is novel, neutral, or that has only acquired first order (direct) associations (Bonardi & Hall, 1994; Bonardi et al., 2017; Honey & Hall, 1989).

Occasion setters are associative learning mechanisms defined as having a modulatory function and are assumed to act upon other associations/relationships rather than acquiring the ability to influence behaviour directly. In other words an occasion setter can strengthen or weaken the ability of a CS to elicit a response, rather than eliciting the response themselves. Holland (1992) provided an account of the properties of occasion setters distinguishing them from conditioned inhibitors. According to this account, when using a feature negative or feature positive discrimination, depending on the circumstances, the feature could gain the capacity to influence the expression of the already established CS-US association. As a result, rather than indicating the occurrence of the US, the feature would indicate if the CS would be followed by the US and would therefore, set the occasion for reinforcement/non-reinforcement of the CS.

Based on this distinction between conditioned inhibitors and occasion setters, conditioned inhibitors are assumed to work using a first order association, while occasion setters are assumed to work using a second order association. A first order association is characterised by a direct link between the conditioned inhibitor and the US, the conditioned inhibitor supressing the US through its own inhibitory association with the US. A second order association on the other hand is characterised by a link between the occasion setter and a CS-US association. In this scenario the occasion setter would not supress the US directly, but rather it would supress the CS-US association (Holland, 1992).

Although a number of studies showed that adding a delay to a feature negative/positive discrimination is the most reliable way to train occasion setting, a recent series of studies provided evidence that the delay is not necessary and participants might develop condition inhibition or occasion setting strategies based on some underlying individual differences (Glautier & Brudan, 2019; Lee & Lovibond, 2021). Glautier and Brudan (2019) classified participants into inhibitors or non-inhibitors based on the level of context inhibition shown to a context that was used to extinguish a previously trained response. In Experiment 1, participants learnt to respond to a cue (D) in context A over a series of training trials. Next cue D was extinguished in a new context: B using a series of unreinforced trials. Then participants were presented with a summation test in which they were asked to predict whether or not cue G would be reinforced in context B, cue G previously received the same training as cue D in context A, but was never shown in context B. Based on this summation test participants were classified as inhibitors and non-inhibitors. The inhibitors' group was assumed to have learnt that context B was a conditioned inhibitor forming therefore a first order association, while the non-inhibitors were assumed to have learnt that the context was an occasion setter, developing a second order association as a result. In Experiment 2 participants received feature negative discrimination training (I+/IJ−) followed by the reversal of the feature (J+ trials) and were then tested to assess whether the feature negative survived the reversal training. The test consisted of IJ trials to check if J still suppressed responding to I after the J+ trials. For non-inhibitors the feature negative discrimination survived the reversal training of the feature to a greater extent compared to the participants who were classified as inhibitors. These differences were statistically significant and suggest that the two groups did develop different learning mechanisms after being exposed to an identical training procedure. If inhibitors developed conditioned inhibition and used a first order association to solve the feature negative discrimination then the reinforced trials of the feature would disrupt its ability to inhibit responding. On the other hand for the non-inhibitors, if they developed a second order association (occasion setting) to solve the feature negative discrimination the reinforced trials of the feature would not impact its ability to supress the CS-US association since it was supressing the association itself and it was not in direct association with the US (Bonardi et. al, 2017; Bouton, 1994; Nelson, 2002, for further analysis). These findings suggests that conditioned inhibition and occasion setting could be reliant on some underlying individual differences and participants could develop either of the mechanisms.

Given that the concept of inhibition is not unique to associative learning, these individual differences could be rooted in the concept wider of inhibition. The mechanisms used in associative and non-associative inhibition to control behaviour could be shared and associative and non-associative inhibition could influence one another. As a result, the current thesis focuses on assessing the link between associative inhibition and non-associative inhibition (a review of the wider concept of non-associative inhibition is provided below).

1.1.2 Non-associative Inhibition

Outside the field of associative learning, inhibition is most often studied in the context of impulsivity, as impulsive behaviour is assumed to be caused by impaired underlying inhibitory processes (Hofmann et al., 2009). Impulsivity however, is often considered a multifaceted personality concept lacking an agreed upon definition as a result of the large number of underlying factors (Bari et al., 2011). Some researchers have even argued that a unitary concept of impulsivity/a single type of impulsive behaviour does not exist, but rather impulsivity should be regarded as an umbrella term which comprises of several related phenomena (Evenden, 1999). Evenden (1999) titled this “varieties of impulsivity”, these varieties being assumed to cluster together and create the concept of impulsivity. In the current review multiple definitions of impulsivity were found in the literature some examples including: attentional, lack of persistence, motor, cognitive, non-planning, novelty seeking, hyperactivity, reward dependence, disinhibition. The fact that impulsive behaviour was defined in many, very distinct ways suggests that there are also multiple inhibitory mechanisms responsible for controlling these types of impulsivity given the substantial difference between the resulting observed behaviour.

For example, actions that appear to lack adequate forethought are sometimes regarded as examples of impulsive behaviour (Broos et al., 2012). However, "actions that lack forethought" can be considered a higher order construct (along with others e.g. susceptibility to boredom) each of which, although apparently different, is brought about by variation on a small number of dimensions. For instance, one of the types of inhibition discussed below refers to a specific situation when a response is initiated by a signal but a second signal flags that the response initiated by the first signal should not be executed. In this situation, if the inhibitory mechanisms of the organism are sub-optimal the response is executed and the resulting behaviour would be described as impulsive. Another type of inhibition considered below would be vital in a situation when people are faced with a choice between two rewards varying in value delivered at different times. If the immediate but smaller reward is preferred over the delayed but larger one, this behaviour would also be described as impulsive, although the type of inhibitory mechanisms which are sub-optimal in this example are different from the ones in the previous example.

Bari and Robbins (2013) presented a possible subdivision of the construct of inhibition in an attempt to map the sub-types that form the construct. The first distinction made was between cognitive and behavioural inhibition. Cognitive inhibition was defined as a mental process which could be responsible for allowing individuals to focus on certain stimuli/situations by inhibiting irrelevant stimuli for example. With this first distinction also comes the first differentiation in the way in which the two types of inhibition are measured. Cognitive inhibition is measured using self-reported measures of impulsivity such as the Barratt Impulsivity Scale-11 (BIS-11) which is discussed in more detail below. The BIS-11 uses items such as "I don't pay attention." and "I have racing thoughts." which capture cognitive inhibition, or the lack thereof. On the other hand, behavioural inhibition is usually measured using behavioural tasks. While the cognitive inhibition does not divide any further, behavioural inhibition was proposed to comprise of three smaller, more specific types of inhibition: response inhibition, delayed gratification, and reversal learning (Bari & Robbins, 2013).

The link between non-associative inhibition subtypes has been repeatedly assessed, the results usually suggesting that the subtypes are not correlated, however within the literature some associations were reported, but these were not consistently replicated. For example, Logan et al. (1997) used an extraversion scale from the Eysenck Personality Inventory (cognitive inhibition) and a stop signal reaction task (response inhibition) to assess the link between impulsivity and the inhibition of proponent responses. The results suggested that impulsivity was not associative with the response to the primary go signal, but it was found to be associated with the participant's ability to stop their ongoing responses (Logan et al., 1997). Most similar studies failed to replicate these results and no significant relationships between self-reported measures of impulsivity and behavioural inhibition were reported. For example Enticott et al. (2006) reported mixed results when using similar measures of cognitive inhibition (Barratt Impulsiveness Scale) and behavioural inhibition (motor inhibition, stop signal reaction task, stroop task, negative priming). The results suggested that the stop signal reaction task performance was not significantly correlated with cognitive inhibition, furthermore the remaining measures of behavioural inhibition were found to be associated only with subscales of the Barratt Impulsiveness Questionnaire; with the exception of the stroop task (Enticott et al., 2006). Aichert et al. (2012) used a very similar battery of tests (antisaccade, stroop, stop-signal, and go/no-go tasks) and assessed their relationship with the Barratt Impulsiveness Scale. In this study they reported that the go/no-go and antisaccade tasks were significantly associated with the BIS-11 scale, while the stroop and stop signal tasks were not. Together, these studies highlight the uncertainty surrounding the concept of impulsivity and inhibition, the correlation between different measures being inconsistent across the literature.

Paulsen and Johnson (1980) investigated the relationship between six different measures of impulsivity (Delay of Gratification, Walk-the-Line-Slowly, Matching Familiar Figures Test, Schenectady Kindergarten Rating Scales, a teacher rating scale, and the Porteus Maze Test) in a sample of preschool children while taking into account age, sex, and IQ. The results indicated that the correlations between the measures of impulsivity were low and mostly not significant. Using a principal components factor analysis with varimax rotation they also noted that three main factors emerged and these factors were centred around age, sex, and IQ, with each measure of impulsivity loading onto one of the factors (Paulsen & Johnson, 1980). This suggests that impulsivity is a multifaceted concept with uncorrelated sub factors which could be influenced/controlled by different mechanisms. Using a similar methodology Caswell et al. (2015) reported four factors: the was represented by the stop signal reaction task, the second by Information Sampling and Matching Familiar Figures Tasks, the third by Immediate Memory Task, and the fourth by the Delay Discounting and the Monetary Choice Questionnaires. The lack of clear, consistent associations between the subtypes of impulsivity/inhibition found in the literature strengthen the hypothesis that impulsivity and therefore inhibition are multifaceted constructs with complex substructures formed of uncorrelated factors.

For the purpose of the current thesis, following the proposed subdivision of inhibition by Bari and Robbins (2013), cognitive inhibition measured using self-reported questionnaires, response inhibition measured by a Stop Signal Task (SST), and delay discounting measured by a monetary choice questionnaire, along with measures of associative inhibition (defined as conditioned inhibition, occasion setting, and extinction separately) have been used in order to assess whether associative and non-associative inhibition are dependent upon a single inhibition construct. The factors identified by Bari and Robbins (2013) were most commonly reported in the literature, for example Reynolds et al. (2006) used three self-reported measures of impulsivity in the forms of personality measures and three behavioural measures of inhibition covering both response inhibition (two tasks) and delay discounting (one task). The results showed that the personality measures were not correlated to the behaviour measures which supports the hypothesis that the two main underlying inhibition subtypes proposed by Bari and Robbins (2013) are independent subtypes of inhibition. Furthermore, the response inhibition tasks and the delay discounting task were also not significantly correlated supporting the above proposed distinction between the two subtypes of behavioural inhibition (see also Broos et al., 2012).

1.1.2.1 Cognitive Inhibition

As previously mentioned, cognitive inhibition refers to the mechanisms responsible for inhibiting irrelevant information or stimuli in order to allow an organism to focus on relevant information/stimuli or tasks and is generally captured by self-reported measures of impulsivity. These self-reported measures allow the respondent to consider a wide range of contexts and reflect upon their behaviour in order to report on their cognitive processes across a variety of changing scenarios (Aichert et al., 2012). Simultaneously, there is a potential trade-off in accuracy when it comes to using self-report measures. When asked to reflect back on their behaviour, respondents could recall situations which are not characteristic of their general behaviour, therefore the accuracy of self-report measures depend on the respondent's ability to evaluate their behaviour over time and contexts (Reynolds et al., 2006).

Cognitive inhibition has been included in a multitude of personality models that consider impulsivity a personality trait for this reason cognitive inhibition is captured in a large number of self-report measures. Usually personality models focus on dysfunctional impulsivity as it is presumed to be linked to a failure to inhibit responses deemed to be inappropriate, and ineffective processing of information. For example in Eysenck's (1990) model of personality there were three main personality traits which were assumed to govern the individual differences in personality. These traits are extraversion, neuroticism, and psychoticism, out of which psychoticism is of importance to impulsivity as individuals high in psychoticism are expected to behave impulsively (Sato, 2005). In the resulting self-report measure of personality, cognitive inhibition was measured with items such as: ” Do you stop to think things over before doing anything?” and “Do you generally ‘look before you leap’?” (Eysenck et al. , 1985). Buss and Plomin (1975) proposed a personality measure formed of: activity, sociability, impulsivity, and emotionality, impulsivity being one of the main personality factors. These are just a couple of examples illustrating the importance of impulsivity and therefore inhibition beyond associative learning, in personality theories.

Due to the importance of impulsivity, many self-report measures were design to assess the construct, the Barratt Impulsivity Scale (BIS-11) being the most comprehensive and widely used measure of impulsivity (Patton et al., 1995). In the development of the scale four different models have been employed: social, behavioural, psychological, and medical. A total of three underlying factors are assessed using this scale, namely: motor impulsivity, attentional impulsivity, and non-planning. These three factors refer to acting without thinking, the ability to focus attention on a given stimulus, and lack of forethought or a focus on the present in this respective order. As part of the questionnaire cognitive inhibition is assessed using items such as: “I often have extraneous thoughts when thinking.”, “I concentrate easily.”, and “I have “racing” thoughts.”.

1.1.2.2 Delayed Gratification

The next subtype of inhibition proposed by Bari and Robbins (2013) is a behavioural subtype and refers the tendency to favour small short-term gains in favour of larger delayed ones (Ainslie, 1975; Ho et al.,1999). Likewise, this also translates to a preference for larger yet delayed losses over immediate but smaller loses. These counterintuitive preferences are inferred to arise from impaired inhibitory mechanisms involved in delay discounting. In order to measure this type of inhibition, delay discounting tasks are used, the concept of delay discounting being defined as the systematic measure of the tendency for the perceived value of a given reward to decrease depending on the delivery time (Rachlin & Green, 1972).

As part of a delay discounting task participants are presented with a choice between two rewards one reward is defined as small and available after a short delay S. The other reward is described as larger than the previous one but available after a delay of S + C, where S is the same delay used for the first reward and C is a constant delay (Ainslie, 2001). When presented with a series of such choices where the delay values are varied, the participant is expected to make a choice between the smaller and larger reward and continue to choose the same reward, up to a point when the other reward is perceived as more advantageous. As an example, a participant might perceive the smaller reward to be more advantageous when they consider S to be too long, however if S is reduced in subsequent trials, a change in preference is expected (Ainslie, 2001). This change is thought to be the result of the delayed reward becoming more attractive as the delay is shortened, and is referred to as the indifference point (Richards et al., 1997). When this indifference point is reached it is assumed that the subject rates the small but immediate reward as equal in value to the large but delayed reward.

Indifference points and the discounting curves have been extensively studied in an attempt to understand individual differences in discounting rates and the reasons why some participants might choose the least advantageous options. Some early economics models have assumed that the discounting curves are exponential in nature, and as a result the discounting of the value of a reward is a constant for every time unit that makes up the delay between the delivery of the reward and the moment it was chosen (Kirby, 1997). Accordingly, exponential curves due to their steady/constant predicted discounting assume that organisms act logically and treat units of time equally. As a result, high discounting rates were presumed to be an indication of a failure of some underlying inhibitory mechanisms (Bickel & Marsch, 2001). Furthermore, non-exponential discounting has been argued to result in maladaptive behaviour since it implies that organisms do not perceive the units of time to be equal regardless of the total delay, which disregards logical reasoning (Ainslie, 2001). Exponential curves, however, are unable to explain the changes in preference and discounting rates observed in empirical studies, because the exponential curves assume that an organisms' preference for a reward remains constant (Ainslie, 1975). Based on the observed discounting rates it was determined that these are not exponential, but rather hyperbolic (Kirby, 1997; Kirby & Herrnstein, 1995; Richards et al., 1997; Richards et al.,1999). Richards et al. (1997) first investigated discounting rates using rats that were offered 100 µl (microliters) of water from two water dispensers which had their delays adjusted using 0, 2, 4, 8, or 16 seconds. Once the rats reached their individual indifference points and the discounting rates were examined it was shown that a hyperbolic function explained the results best. Richards et al. (1999) later reproduced these results with human participants using monetary rewards and delays varying from 0 to 365 days.

These hyperbolic rates of discounting can be estimated using Mazur's (1987) equation shown below:

$$V = \frac{A}{1 + kD}$$

Equation 1

In

Equation 1 the V parameter represents the perceived value of the reward, while the A parameter represents the amount of this reward. The k parameter is the delay-discounting rate individual to each organism, and D parameter represents the time delay. From

Equation 1 the k

parameter is of upmost importance as it represents the slope of the hyperbolic discounting rate unique to each organism, making it a measure of intertemporal inhibition (Kaplan et al., 2016). Consequently, organisms with large k values discount delayed rewards more heavily than participants with small k values (Peters & Büchel, 2011).

Table 1 shows the rate at which a £1000 reward would be discounted by participants with four different values of k, over multiple time intervals using Mazur's equation. According to these predictions, a participant with a discounting value of .05 would value the £1000 reward as £833.3 if the reward was to be delivered after 4 weeks. Therefore, if these 4 hypothetical participants were offered a choice between £500 now or £1000 in 104 weeks (2 years), the participants with k values of .05 and .01 would choose the £500 reward now, while the participants with k values of .005 and .001 would choose the £1000 reward in 2 years. This is because the participants with higher k values perceive the delayed reward as less valuable than then immediate one (£161.29 for k = .05 and £490.2 for k = .01), while the participants with lower k values would perceive the delayed reward as more valuable (£657.89 for k = .005 and £905.8 for k = .001).

Delay	k = .05	k = .01	k = .005	k = .001
1 week	952.38	990.1	995.02	999
4 weeks	833.3	961.54	980.39	996.02
24 weeks	454.54	806.45	892.86	976.56
52 weeks	277.7	657.89	793.65	950.57
104 weeks	161.29	490.2	657.89	905.8

Table 1

Devaluation of a Reward of £1000 Over Time According to Mazur's Equation

Although discounting rates have been found to be stable across time, cultural differences in discounting rates were reported. For example Du et al. (2002) reported that Americans and Chinese participants discounted delayed rewards more heavily compared to Japanese participants. Another concern regarding the reliability of measuring discounting rates relates to the fact that the procedure described above used hypothetical rewards, nevertheless several studies have shown that the discounting rates calculated based hypothetical rewards and real rewards are highly correlated, meaning that hypothetical rewards can be used to approximate delay discounting (Johnson & Bickel, 2002; Lagorio & Madden, 2005). Additionally, high delay discounting rates were shown to be associated with substance-abuse and gambling (Bauer, 2001; Bickel et al., 2010; Rodriguez-Jimenez et al., 2006; Rubio et al., 2008).

1.1.2.3 Response Inhibition

Response inhibition was the last sub-type of behavioural inhibition proposed by Bari and Robbins (2013) that was used in the current thesis and it refers to the inhibition of ongoing or pre-potent motor responses. A failure in the underlying response inhibition mechanisms is thought to result in inhibitory dyscontrol, the individual being therefore unable to delay or terminate an ongoing response (Enticott et al., 2006).

Response inhibition can be measured using multiple tasks some of the most widely used tasks are the stop signal reaction task (SST) and the go/no-go task. For the purpose of the current thesis the discussion below focuses on the SST only. SST measures action cancellation which is an organism's ability to inhibit a response that was already initiated. In order to assess action cancellation during the SST, participants are asked to carry out an ongoing task which requires participants to respond to a stimulus that is repeatedly presented over a series of trials. Participants are also instructed that on a number of trials they should withhold responding to the stimulus, and that these trials are marked by the presence of a stop signal (Logan & Cowan, 1984). Participants therefore require fast inhibitory control mechanisms in order to successfully stop their ongoing response when the stop signal is presented (Logan, 1994).

The SST can be understood using a "horse-race model", which assumes that the behaviour observed on each trial is dependent on a race between two separate responses. One of the responses is generated by the go stimulus, the action the participants were instructed to perform on the series of trials. The second is the stop response generated by the stop signal that is presented only on some of the trials in the series. As a result, if the initial response is the first to reach completion the participant will complete the action (response) they were asked to perform on the trial, but if the second response reaches completion first, no action is observed. According to the horse-race model the response inhibition is dependent on the time it requires the two responses to reach completion as shown in Figure 1 (Logan & Cowan, 1984).

Figure 1

Representation of the Horse-race Model Recreated from (Logan & Cowan, 1984)

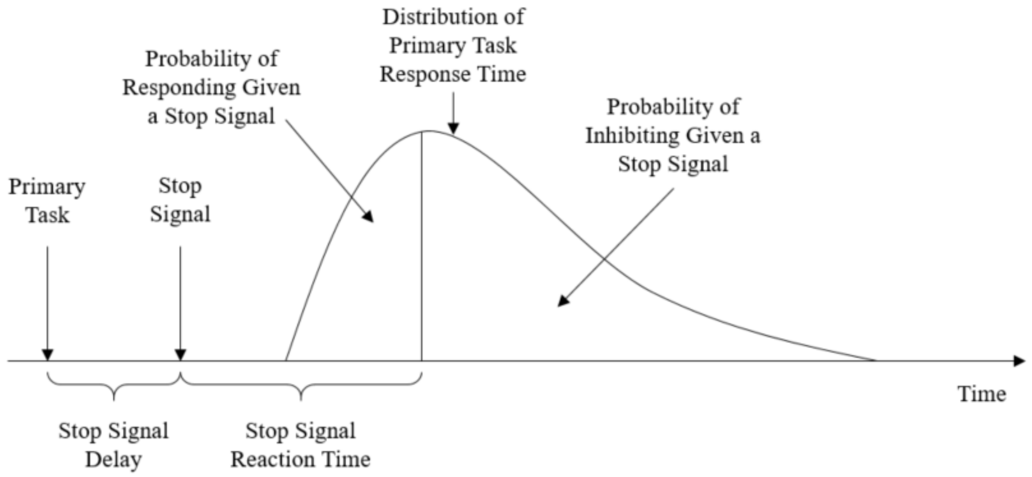


Figure 1 shows a basic illustration of the horse-race model recreated based on Figure 2 in Logan & Cowan (1984). Horizontally, time is represented as an X axis, and the distribution represents the primary task response time. The point at which the primary task starts and the point at which the stop signal is presented are also represented in the figure. The vertical line dividing the distribution of primary task reaction time represents the point at which the stop response reaches completion. Correspondingly, the distance from the point when the stop signal is presented and the point when the stop signal reaches completion is the stop signal response time (SSRT). The space between the primary task and the stop signal is the stop signal delay (SSD). Assuming that the SSD and SSRT are fixed by the task and individual differences respectively, then the probability of responding or inhibiting a response when a stop signal was presented is the integral under the corresponding parts of the distribution curve to the left and right of the line respectively. If we imagine that some participants have a fast (short) stop signal reaction time, the line dividing the primary task response time distribution would move to the left, reducing the probability of the primary response reaching completion before the stop signal. Similarly, if participants have a slow (long) stop signal response time the line would be moved to the right, having the opposite effect. To accurately estimate the SSRT for each participant the delay between the primary task offset and the stop signal offset is continually varied during the SST. The SSRT is the focal point of the SST as it represents a measure of inhibitory control each participant is assumed to have a unique value (Aron & Poldrack, 2005).

Similar to delay discounting, high SSRT scores and therefore dysfunctional response inhibition, was found to be associated with ADHD and substance abuse (Chamberlain et al., 2007; Fillmore & Rush, 2002; Schachar et al., 1993). Furthermore, the design of the SST is very similar to a feature negative discrimination used to train conditioned inhibition, and even to extinction if we assume that after extinction there are two opposite meanings available for the same stimulus. Given these similarities it could also be assumed that associative inhibition and response inhibition rely at least partly on common underlying mechanisms.

Reversal learning from the original subdivision proposed by Bari and Robbins (2013) was not included due to the considerable overlap with the measures of associative learning used in the studies of the current thesis, however previous studies focusing on the relationship between reversal learning and impulsivity/inhibition offer an insights into the possible link between associative and non-associative inhibition. These studies showed mixed results, similar to the literature on the relationship between the different subtypes of impulsivity/inhibition. For example, Zou et al. (2022) aimed to assess the relationship between impulsivity and probabilistic reversal learning using a sample of student participants. Impulsivity was measured using the Short Version of the Urgency Premeditation Perseverance Sensation Seeking and Positive Urgency (S-UPPS-P), and reversal learning was assessed using a custom built task. The results showed no significant relationship between the two although the analysis revealed that an effect of impulsivity on switching behaviour following as a result of consecutive non-reinforced trials (Zou et al., 2022). This suggests that associative and non-associative inhibition are individual sub components of inhibition but they could rely on common underlying mechanisms, or in influence each other indirectly. O'Donnell (2021) assessed the relationship between belay discounting, impulsivity measured using the Barratt Impulsiveness Scale, reversal learning, and cannabis use. They reported that delay discounting and impulsivity were higher in young adult cannabis users, while reversal learning was not significantly different between the two groups (users and control). This can be interpreted as further evidence for independent underlying mechanisms of associative and non-associative inhibition. Within the literature there are however studies which found evidence of a link between the two types of inhibition such as Franken et al. (2008) who reported weaker performance in a reversal learning tasks for participants who scored high on impulsivity. Impulsivity was measures using the Impulsiveness Scale of Eysenk's Impulsiveness Questionnaire and participants who scored high on this measure were observed to have impaired performance in the reversal learning task, having difficulties with behavioural adaptation based on the changes in the reinforcement of the target (Franken et al., 2008). Gullo at el. (2010) reported similar results using the Eysenck Personality Questionnaire – Revised to measure impulsivity and a reversal learning task, participants who scored higher on the impulsivity scale also making more mistakes in the reversal learning task. Together the mixed results reported within the literature further highlight the need to understand the underlying structure of impulsivity/inhibition which seems to comprise of multiple, independent constructs.

The first set of studies presented in the current thesis focus on assessing the link between associative and non-associative inhibition, where the former is defined as conditioned inhibition and the latter is measured using four different measures mapping onto the subdivision of inhibition proposed by Bari and Robbins (2013). The four measures of non-associative inhibition were the Barratt Impulsiveness Scale and the Behaviour Inhibition System/Behaviour Activation System for cognitive inhibition, a monetary choice questionnaire for delay discounting, and the Stop Signal Reaction Task for response inhibition. The relationship between associative and non-associative inhibition is also assessed in the second series of experiments where associative inhibition is defined as extinction rate and context inhibition. By assessing the relationship between associative and non-associative learning the current thesis aims to better understand the structure of inhibition and determine whether the two inhibition subtypes are independent or rely on common underlying mechanisms.

1.2 Extinction

Another way of changing a previously acquired association is through extinction. Compared to conditioned inhibition and occasion setting which change an existing association when the conditioned inhibitor/occasion setter is present, extinction refers to the serial presentation of a CS in a succession of non-reinforced trials. As a result, the CS stops eliciting a response giving the impression that the CS-US association is erased. Referring back to the initial example of the rat which learnt to expect food every time a light is turned on, if the rat is presented with the same light on multiple occasions and no food is delivered, the rat would come to learn that the light no longer predicts the delivery of food. Accordingly, the rat would stop displaying food expectancy when the light is turned on. Early models of associative learning treat extinction as unlearning, assuming that the CS-US association is destroyed due to the non-reinforcement of the CS. For example the Rescorla-Wagner model treats extinction as unlearning, because inhibition and excitation are viewed as opposites that sit on an axis at the minimum and maximum points of –1 and 1 respectively. In this context extinction would mean bringing a cue's associative strength to 0 where it doesn't have the ability to influence behaviour. Effects such as the spontaneous recovery and renewal however, show that although responding to a cue stops following extinction, the initially learnt association is not destroyed/erased.

1.2.1 Spontaneous Recovery

The spontaneous recovery effect refers to the restoration of a response which was extinguished following a time interval subsequent to the extinction and is one of the most basic associative learning phenomena. If the CS-US association was completely erased following extinction the original response should not be able to recover. In a more recent review of the phenomenon, Quirk (2002) examined the impact the passage of time had on the learning that took place in the extinction phase. Rats were trained using a tone CS and an electric shock as the US (T+), followed by 20 non-reinforced (extinction) trials (T−). The rats were assigned to one of the following groups: 30 minutes, 1, 2, 4, 6, 10, or 14 days, which indicated the interval of time following extinction after which the recovery effect was tested. Two control groups which received no extinction trials were also tested either 1 or 14 days after receiving the training. The results showed that the recovery effect increased gradually with the passage of time, reaching 100% by the 10th day, however the control groups did not significantly differ when compared to one another (1 day vs 14 days), suggesting that the passage of time only made a difference for the association that underwent extinction and not for the one that did not (see also Brooks & Bouton, 1993).

1.2.2 Renewal

The renewal effect refers to the resumption of an extinguished response following a change in the extinction context. This effects also supports the view that an extinguished association is not completely erased; two types of designs have been used to show this effect: ABA and ABC (Bouton & Bolles, 1979). The notations refer to how the context was varied across the experiment therefore, in an ABA design, training took place in context A then, extinction took place in context B, after which the context was changed again and a renewal test was presented in the original training context, A. The ABC design follows the same pattern, with the exception of the renewal test takes place in a novel/neutral context. Harris et al. (2000) showed that the extinction context plays an important role in the learning that takes place during extinction training. Rats were presented with two reinforced CSs, CS−A and CS−B in context C (ctxC), then each of the CSs were extinguished in separate contexts, different from the initial training context. The renewal test for each of the CSs was then performed in the context the other CS was extinguished in, and a renewal effect was observed for both. The results indicate that the extinction learning depends on the context, but also the fact that the context could control specific CS-US associations.

1.2.3 The role of context in extinction

Bouton (1993) proposed an explanation for the underlying mechanisms involved in extinction using the interference paradigm. This account is based upon four principles of animal long-term memory. The first and most important assumption is that the primary determinants of memory retrieval are the contextual stimuli experienced by an organism. Retrieval of memory representations is thus reliant on the similarity between the environment of conditioning and the environment at the moment of memory retrieval (McGeoch, 1932). Secondly, time can act as a context, more specifically the passage of time is assumed to cause a change in the context as the external and internal factors experienced by an organism inevitably change. Thirdly, specific memories are considered to be highly dependent on the context, conditioning (excitatory learning) is speculated to be less sensitive to contextual change than extinction (inhibitory learning). The last assumption is that interference between memories does not occur at encoding/input but rather at retrieval/output, Bouton (1993) suggesting that conditioning and extinction each store at least one representation in memory. As a result, when opposing representations are retrieved from the long-term memory they compete for the limited space available in the working memory (Bouton, 1993). This is very similar to Konorski's (1948) early theory of conditioning who proposed that a CS-no US association can be formed resulting in two meanings being available for a CS: CS-US and CS-no US.

According to this account an organism is assumed to store data about the stimuli it encounters: CS, US, as well as the context in which the pairing of the CS and US takes place and code this information into a representation. If the organism is then exposed to an extinction treatment, the organism would start to gradually encode a novel representation that marks the absence of the US when the CS is presented. This new representation can be regarded as a representation of inhibition. This interpretation of extinction contradicts the one presumed the early learning models by speculating that the learning that occurs during the acquisition stage is not erased during extinction. It is believed that this learning remains intact and is retrieved on every extinction trial dependent on the similarity between the extinction and acquisition context. Similarly, once the extinction learning is complete it would also be retrieved on every subsequent trial, therefore this newly encoded representation interferes with the representation of the initial training stage. The trial-by-trial reduction in responding during extinction is then the result of the incremental growth of the extinction representation. At the end of the extinction training, the organisms has two contradictory representations of the CS, therefore making the CS ambiguous (Bouton, 1993). As a result, the performance following extinction is contingent on the retrieval of the extinction representation, therefore unstable performance is to be expected. A key assumption in the theory is that the unstable performance is due to the extinction retrieval being dependent on the context. Bouton (1993) suggested that extinction is more dependent on the context than conditioning, therefore changing the context in which an organism finds itself following extinction is expected to result in a reduction in its ability to retrieve extinction, while the retrieval of conditioning is left unchanged. Accordingly, when the context is changed and the extinction retrieval is affected the two phenomena discussed above, spontaneous recovery and renewal occur. This explanation is straight-forward for renewal as the context is already the factor triggering the renewal effect, however the role of the context might be less apparent for recovery. Bouton (1993) proposed that the renewal effect is a consequence of the organism being removed from the temporal context of extinction, which happens naturally with the passage of time. The results from Quirk (2002) and Harris et al. (2000) can then be attributed to a change in context from the end of the extinction phase to the test phase, which in turn caused a failure in the retrieval of the extinction learning.

Bouton and Nelson (1994) further investigated the hypothesis according to which inhibition is context specific. Over a series of four experiments with rats using a feature negative discrimination procedure they have shown that a target cue was more difficult to inhibit in a context in which it was never inhibited before. In Experiment 1 the properties acquired by the feature as a result of feature negative discrimination training were examined. Two groups were used, the feature negative group received feature negative training (T+, LT−), and the control group was exposed to T and L alone trials (T+, L−). The results showed that for the feature negative group L became inhibitory and the cue supressed responding to a new cue N, while for the control group no such transfer was observed. This was interpreted by assuming that the inhibitor is responsible for at least partly supressing the memory representation of the US that is elicited by the CS. In Experiment 2 it was aimed to assess the impact of context change on conditioned inhibition. Once again two groups were used, each received two feature negative discrimination training with different features, one in context A and the second one in context B. The transfer of inhibition of the feature trained in context A was tested for the two groups in either context A or context B, and the transfer was successful indicating that inhibition is not affected by context change. Experiment 3 aimed to test the hypothesis according to which inhibition to the target and not the feature is affected by a context change. Once again two groups were used and the design mirrored the design of Experiment 2, however the same feature was used in the two contexts, but the targets were unique to the contexts. In the test stage the inhibitory properties of the feature were tested with the original target used in context A, and the test was performed in context A for one group and in context B for the other. The data indicated that the feature reduced responding in the original context of the target, context A more than in context B for which the target was novel. Because the feature was used in both contexts this effect was attributed the target following the context change. Additionally, this effect was only observed for the compound which was inhibitory, and not the cue alone which was excitatory, suggesting the inhibition was disrupted by the context change and not excitation. Finally, in Experiment 4 it was aimed to compare the effects studies in experiments 2 and 3: the effect of a context which on the feature and target. All rats received the same training: context A (T+, N+, LT−), context B (T+, N+, KT−), and the test was carried out in both context A and B. The data showed that the context switch impacted responding to the compound, but because the feature was not new to the test context, therefore this was attributed to the target and it was concluded the context switch affected the inhibition to the target. These results were interpreted by assuming that the context switch reduced the "inhibitability" of the target. In order to explain the effect Bouton and Nelson (1994) proposed that the feature could become either a first (conditioned inhibitor) or a second order (occasion setter) mechanism to inhibit reposing to the target cue. These first and second order associations are consistent with conditioned inhibition and occasion setting, however the occasion setting mechanism was proposed to be reliant on the presence of the context which would justify why a conditioned inhibitor can readily transfer and inhibit other cues while the occasion setter is dependent on the context. Similarly, when an organism undergoes a series of extinction trials, a new association is learnt, but as opposed to the initial learning from the conditioning training, the extinction learning has the context encoded into it, requiring the presence of the context in order to stop responding successfully.

1.2.3.1 Protection from extinction

An alternative explanation for extinction being reliant on the context is the protection from extinction theory. Described first by Chorazyna (1962) the protection from extinction phenomenon refers to the property of an inhibitor to shield the excitatory properties of a target cue that undergoes extinction. If for example, a cue A, which previously received reinforced training, is extinguished in the presence of an inhibitory cue I, the normal expected reduction in responding to cue A can be observed during the extinction trials. This reduction in responding is partly due to the non-reinforcement of the cue, however it is assumed to be also partly due to the inhibitory cue I being present. As a result, some of the original conditioning of cue A remains intact and once it is presented by itself the cue still elicits a response. Rescorla (2003) showed this effect in a series of four experiments using rats and pigeons where a conditioned inhibitor protected a cue from extinction when the cue was exposed to non-reinforced trials and the conditioned inhibitor was present. In Experiment 1 pigeons received reinforced training with cues A, B, and O, in the next stage reinforced trials of O were continued but the cue was also paired with T and the compound was nonreinforced. In the last stage of training cues A and B were presented in nonreinforced trials, A was nonreinforced in compound with T while B was nonreinforced by itself. The results showed that the presence of the conditioned inhibitor in compound with A protected the cue from extinction, pigeons responding more to cue A than B. Experiment 2 aimed to compare the protection from extinction effect of a conditioned inhibitor to the one produced by a neutral stimulus using a similar design to Experiment 1, with the addition of O+, OX−, OY+ trials. The results were similar to the ones of Experiment 1, the cue that was extinguished in compound with the conditioned inhibitor elicited more responding compared to the cue that was extinguished with the neutral cue. Experiment 3 used the same design as the previous experiment with the difference that cue Y, the control cue, was only presented by itself in nonreinforced trials. The resulting data followed the same pattern as the one showed in Experiment 2. The final experiment was a replication of Experiment 1 using a magazine approach and rats rather than autoshaping with pigeons, and the results once again showed that the conditioned inhibitor produced a protection from extinction effect.

Despite the potential of a conditioned inhibitor to protect a cue from extinction, such an inhibitory cue is rarely or never present during extinction, and the context is new to the organism in the case of an ABA/ABC design or expected to be neutral. In order for the neutral context to be able to protect a cue from extinction it is assumed that during the non-reinforced extinction trials the context gains inhibitory strength as the context can be regarded as a CS present in the background. Subsequently the inhibition gained by the context during the extinction trials could be sufficient for the context to act as an inhibitor and protect the target cue from extinction. Consequently, when removing this newly inhibitory context the renewal effect is observed, or recovery if the context is a temporal one. Bouton and Bolles (1979) argued against this assumption claiming that the context could not become a conditioned inhibitor as a result of the extinction trials. Larrauri and Schmajuk (2008) proposed that the properties of extinction can be explained by a model that considers only attentional and associative mechanisms. Using simulated data from the model they proposed that even when a context – US inhibitory association is formed these associations are difficult to detect because the attention paid to the inhibitory context is small.

This presumption that the context can become inhibitory during extinction and protect the target from extinction is crucial for the Rescorla-Wagner model in order to account for the renewal and recovery effect. Although this model can predict acquisition/extinction, overshadowing, blocking, conditioned inhibition, and other phenomena it relies on the context gaining inhibitory strength to be able to also explain renewal and recovery. This is because the model assumes that during extinction the strength of an association is permanently weakened and lost. As a result, no recovery/renewal effect should be observed following extinction. In contrast, Pearce's configural model assumes that an organism continuously forms new configural representations based on its experience, hence it does not treat extinction as unlearning. The recovery/renewal of a previously extinguished CS does not pose a threat to the model since the original associations are thought to be stored in the organism's memory. Each model is discussed in more detail below.

1.2.4 Applications

Understanding extinction and the underlying mechanisms by which it operates has important clinical implications. Although associative learning offers organisms an evolutionary advantage, it can also be maladaptive. One example where associative learning could become maladaptive is substance abuse. Using the same principles outlined above humans can develop associations between various substances such as drugs or alcohol, and a desired internal state. In this case the substance would be the CS and the US would be the desired internal state. Reinforcing these associations could lead to effortless and automatic drug use/alcohol consumption or craving which in turn causes addiction (Everitt & Robbins, 2016). Addiction can be treated using exposure therapy and cue exposure therapy among others, and these two treatments rely on training the patient to not respond to substances or cues related to substances (CSs). This learning can be regarded as extinction, as such we can imagine the example of a patient who developed a maladaptive association which prompted the patient to consume alcohol when exposed to certain cues such as the sight or smell of alcohol. Accordingly, the goal of cue exposure therapy would be to extinguish this association, and encourage the patient to not engage in alcohol consumption while in the presence of the alcohol cues. Considering the previous discussion about extinction, we know that it does not erase an association, therefore there is always the risk of the response returning leading to relapse (Conklin & Tiffany, 2002; Crombag et al., 2008; Taylor et al., 2009; Vervliet et al., 2013). Furthermore, if extinction is influenced by the context, this could translate to the rehabilitation environment influencing the learning that occurs while the patient is in this environment, heightening the risk of relapse once the environment is changed. Even if the context in which the learning takes place doesn't influence the extinction of the maladaptive association, there is the risk that with a passage of time the association will recover as the learning could be dependent on a temporal context. As an added complication if the context was to become inhibitory during extinction, the initial discussion about conditioned inhibition and occasion setting becomes relevant as the context could develop different properties (first order or second order associations).

Given the importance of extinction and the complexity of the presumed underlying mechanisms by which it operates, consolidating the learning that takes place during extinction and reducing the recovery and renewal are topics of high interest in the field of associative learning. Below two possible ways of obtaining a lasting extinction are discussed.

1.2.4.1 Super-extinction

Super-extinction refers to the extinction procedure where a target cue undergoes extinction in compound with another cue that received previous reinforced training, identical to the target cue prior to extinction (Rescorla, 2000). If two excitatory cues are paired in compound, then according to Rescorla and Wagner (1972) their associative strengths should mathematically summate. Consequently the organism's expectation for the newly formed compound to be reinforced should be higher than the separate reinforcement expectation for the individual cues that form the compound. Due to this difference in expectation the compound is expected to have a more rapid extinction compared to a cue extinguished in isolation. This prediction relies on the assumption of the Rescorla-Wagner model that learning is driven by the discrepancy between the organism's expectations and the events experienced (Rescorla & Wagner, 1972). Contrary to the Rescorla and Wagner's (1972) theory, Pearce's configural view of associative learning would predict that such extinction training should lead to a more unstable extinction and therefore more recovery. This is because the organism is assumed to encode various configures during extinction which would be reliant on the compound of the two cues. After the extinction training the organism would have two available meanings for the target cue: one which indicates that the cue was reinforced when presented alone and one which indicates that the cue was non-reinforced when presented in compound with another cue. Under these circumstances when the organism is presented with the recovery test which comprises the target cue alone, the learning that occurred before extinction would win as it resembles the test the most (Pearce, 1987). These two theories and resulting models are discussed in more detail below.

In a series of four experiments Rescorla (2000) investigated the difference between cue alone extinction (X+, X−) and super-extinction (A+, X+, AX−). In Experiment 1 a cue undergoing extinction in compound with another excitatory cue was observed to lead to less recovery compared to a cue that received the same training by itself. Although super-extinction showed less recovery, the extinction observed was not more rapid, but slower than cue alone extinction. Experiment 2 used an instrumental learning paradigm for a systematic replication of the first experiment and the results were similar, showing that compound extinction with an excitatory cue resulted in less responding to the target cue at test. No difference in the speed of extinction was observed for this experiment. Experiments 1 and 2 used a between-subjects design meaning that an alternative solution could be proposed for the observed data: for the compound extinction group because cue A was present during the extinction trials it could be assumed that the extinction of A could generalize to X leading to less responding to X at test. Consequently, Experiments 3-4 aimed to replace the results of the first two experiments using within-subjects design. The results supported the findings of the first experiments, compound extinction leading to less recovery when using a within-subject design. These results support the hypothesis that compound extinction leads to less recovery but did not show a faster extinction. Thomas and Ayres (2004) however, did observe a faster extinction when two excitatory cues received extinction training in compound. In Experiments 4a and 4b three excitatory cues were extinguished in compound and faster extinction rates were observed to the compound compared to cue alone extinction. In addition, the group that received compound extinction also showed less recovery.

Despite the success of training a more rapid and stable extinction using super-extinction shown by animal studies, several human studies have failed to replicate these results (Lovibond et al., 2000; Vervliet et al., 2007). For example, Vervliet et al. (2007) conducted a fear conditioning experiment using a sample of student participants; during the experiment two neutral stimuli A and B were trained as shock predictors. Following the initial training, A and B were extinguished in compound, after which cue A was presented by itself in a test stage. Both shock expectancy and electrodermal responding indicated a strong recovery effect for cue A following super-extinction, bringing into question the effectiveness of the procedure. Similarly, McConnell et al. (2013) aimed to explore the effect of multiple cue extinction using compounds of three stimuli across a series of conditioned suppression experiments with rats. In Experiment 1 of the paper three distinct groups received one of the following extinction treatments: target extinction in compound with two additional neutral cues, target extinction in compound with one excitatory and one neutral cue, or target extinction in compound with two additional excitatory cues. The results showed that the first group (compound extinction with two neutral cues) showed less recovery following extinction followed by the third (compound extinction with two excitatory cues) and then the second groups (compound extinction with one neutral and one excitatory cue). Overall, these results pose a problem for Rescorla's initial demonstration, suggesting that the total error reduction might not be driving a faster more rapid extinction as initially suggested.

Furthermore, the protection from extinction assumption becomes problematic in the case of a super-extinction design. This assumption essential for some theories to predict recovery and renewal, but in the case of super-extinction because of the elevated expectation for reinforcement of the compound, the context of extinction should become even more inhibitory, which should lead to more recovery and renewal. To date there is no consensus as to whether super-extinction is effective or not as it was not been extensively studied following Rescorla's initial demonstration of the phenomenon.

1.2.4.2 Deepened Extinction

The second method of training a more stable extinction is deepened extinction, for which the extinction training is separated into two stages. In the first stage, a target cue is extinguished by itself following a traditional extinction procedure (single cue non-reinforced trials). In the second part of the extinction the partially extinguished target cue is paired with a cue that received excitatory training during the previous training stage. For the rest of the extinction training the two cues are extinguished in compound. This procedure resembles conditioned inhibition where the partially extinguished target is the conditioned inhibitor, therefore this is meant to drive the extinction of the target cue further than a traditional extinction procedure. Following this training it is possible for the target cue to go below the value of 0 which is usually obtained through traditional extinction, as a result the cue could become inhibitory which in turn would result is less recovery (Reberg, 1972; Rescorla, 2006).

Rescorla (2006) examine the properties of deepened extinction in a series of five experiments and across all experiments a spike in responding was reported on the first trial of the compound extinction. It is this spike in responding that is assumed to be responsible for the additional extinction. Most importantly Rescorla showed that deepened extinction resulted in less recovery compared to a traditional target alone extinction. These results were attributed to the enhanced excitation brought on by the pairing of the partially extinguished target cue and the excitatory cue. The results were replicated in both further animal and human studies (Coelho et al., 2015; Culver et al., 2015; Janak et al., 2012; Janak & Corbit, 2011; Kearns et al., 2012; Leung et al., 2012). One such study was conducted by Coelho et al. (2015) and employed a fear conditioning experiment. To compare extinction procedures two groups were used, a control and a deepened extinction group, which received cue alone or cue alone followed by compound extinction of the target with an excitatory cue respectively. Acquisition and extinction were trained on day one and on day two participants took part in a spontaneous recovery and reinstatement tests. The results showed that deepened extinction led to less recovery compared to the control group, however there was no difference between the two groups in the reinstatement test.

Although the results of Rescorla (2006) were replicated across the literature, several studies failed to obtain the same effect leading to a mix of evidence. Kearns et al. (2012) presented a study using rats which aimed to examine the effectiveness of deepened extinction with cocaine cues. Three cues, tone, click, and light were used to train cocaine self-administration during the acquisition stage. The extinction that followed was split into two phases, during phase one all three cues were presented by themselves in separate non-reinforced trials. Next, during the second extinction phase the rats were divided into two groups, for one group the tone and the light were presented in compound and the click continued to be presented alone while for the other group the roles of the tone and click were reversed. The two extinction phases were followed by a spontaneous recovery test consisting of 16 presentations of each the click and the tone. Overall, the results showed that there was no difference between single cue and deepened extinction when the whole recovery test performance was analysed, but the two techniques were different when only the first four trials of the test stage were compared. The first four trials were compared to match the design used by Rescorla (2006) who used a test stage comprising of four trials. These results suggests that, although deepened extinction was observed to lead to less recovery this effect might not be stable over time. Furthermore, Krypotos and Engelhard (2019) reported no differences between cue alone and deepened extinction in fear and avoidance responses, concluding that deepened extinction did not produce a more stable extinction.

Super-extinction and deepened extinction were rarely compared against one another in the literature, most investigations focusing on the effectiveness of one if the procedures compared to a control cue alone extinction procedure. One of the only studies that directly compared the two compound extinction procedures was conducted by Griffiths et al. (2017) who used a series of two experiments to compare deepened extinction to super-extinction first and then deepened extinction to cue alone extinction. The experiments used two independent samples of students who were asked to take part in a classic learning task where they had to learn about the foods that are safe to eat for Mrs. X. The results of Experiment 1 revealed that the cues used in super-extinction showed more recovery compared to the cue that was extinguished through deepened extinction. As a result, deepened extinction was assumed to be the more effective extinction technique, however when compared to cue alone extinction in Experiment 2, deepened extinction was observed to lead to more recovery. Collectively, the two experiments suggest that compound extinction is not more effective than cue alone extinction, and that super-extinction leads to most recovery out of the three techniques which contradicts the two studies by Rescorla (2000, 2006) who showed that both super and deepened extinction were superior to cue alone extinction (see also Pineño, 2007). Together, these mixed findings cast doubt over the effectiveness of deepened extinction, however it might be the case that compound extinction is only effective in certain situations or it is dependent on some underlying individual differences.

Overall, these two variations of extinction rely on cue interactions that are aimed at developing a more stable extinction learning that shows reduced levels of recovery and renewal. By combining excitatory cues in extinction more inhibition is assumed to be encoded, therefore the two procedures are similar in their aim of using more inhibition to reduce responding more permanently. As a result, it can be assumed that the success of extinction could be dependent on the organism's ability to develop/show inhibition. Having discussed inhibition in the context of occasion setting, conditioned inhibition, and extinction, the importance of inhibition in associative learning is evident. It was also previously mentioned that the concept of inhibition is not unique to associative learning, therefore learning in all the previously discussed situations/paradigms might depend on the wider concept of inhibition.

The second series of experiments presented in the current thesis focuses on assessing the effectiveness of compound extinction compared to cue alone extinction where compound extinction has been first defined as super-extinction only and later as either super-extinction or deepened extinction. Given the intense focus on inhibition when considering extinction the second series of experiments also aims to assess the previously discussed link between associative and non-associative inhibition. For this purpose, associative inhibition was defined on the basis of extinction, while non-associative inhibition was defined in the same manner as before. To quantify the inhibition resulting from extinction two approaches were used, first the extinction rates in the form of the speed of extinction was computed and the relationship between this and non-associative inhibition was assessed. Next, the inhibition developed during extinction was measures using a context inhibition test, which examined the levels of inhibition acquired by the extinction context as a result of the extinction training. The relationship between this and non-associative inhibition was also assessed.

The next section focuses on two widely used and influential formal models of associative learning, as they hold different predictions for the above mentioned compound extinction techniques.

1.3 Formal Associative Learning Models

In order to better understand extinction and the related phenomena discussed as part of this chapter it is important to consider the manner in which extinction was defined as part of various frameworks through the associative learning literature. Two main theories and the models proposed based on those are briefly discussed below: the Rescorla-Wagner Model and Pearce's Configural Model.

1.3.1 Rescorla-Wagner Model

Rescorla and Wagner (1972) proposed an associative model, which relies on a simple mathematical equation to predict behaviour. The model can account for simple phenomena such as acquisition and extinction, but it can also provide an account for cue interactions. As a result, the model is able to predict various associative phenomena such as overshadowing, blocking, and most relevant for the current thesis conditioned inhibition. The model was designed to provide a trial-by-trial representation of the changes expected in the associative status of a stimulus. In the development of the model, Rescorla and Wagner build on the research of Kamin (1969) whose work assumed that expectancy and surprise play a central role in conditioning. According to this view, organisms have particular expectations when encountering a stimulus. If these expectations are violated the organism would be surprised by the discrepancy between its expectation and the outcome. Consequently, this would facilitate the formation of an association between the stimulus and the surprising/unexpected outcome that occurred. Referring back to the initial example of the rat which learnt to expect food every time a light stimulus was presented, according to Kamin's theory, the rat came to associate the cue with the food because its expectation was violated. The light stimulus was novel prior to training so the expectation of the rat was that nothing would happen, however when it was presented with food immediately after the light stimulus the surprise of the pairing caused it to start developing an association between the cue and the outcome. This "surprise" element is coded into the Rescorla-Wagner Model, and it determines the "amount" of conditioning that occurs. The mathematical formula of the model was defined as follows:

$$\Delta V_x^{n+1} = \alpha_x \beta (\lambda - V_{total}^n)$$

(2)

$$V_x^{n+1} = V_x^n + V_x^{n+1}$$

(3)

In this model, V_x^n represents the associative strength of a given cue, X before the US is presented at the start of trial n, ΔV_x^n represents change in associative strength which results from the pairing of CS X with the US on the trial n, therefore ΔV_x^{n+1} which is the sum of the previous two represents the associative strength of cue X after the US was presented on trial n. The parameter αX represents the learning rate for the CS X, this is a value which shows how easy or difficult the acquisition of associative strength is for the given stimulus. The value of α can range from 0 to 1 and is related to the nature of the stimulus, more specifically to its salience/intensity. Similarly, β represents the learning rate of a given outcome (US) and it reflects how well associative strengths can develop from pairing a CS with that US. The value of β also ranges from 0 to 1 and is dependent on the salience/intensity of the US. The λ parameter represents the highest value of associative strength that a US can reach. The final parameter from (2), V_{total}^n represents the sum of the associative strengths of all the CSs which were present on the trial n. To summarise, the first equation represents the associative strength (ΔV_x^{n+1}) resulted from pairing CS X which has an associability value of αX with a US which has an associability of β on a trial n where the US can reach the maximum value of λ and the sum of the associative strengths of all the cues present on the n trial are V_{total}^n. The second equation shows how the associative strength of a CS is updated after a trial. Consequently, this part of the model shows how the associative strength of a CS X is updated after a trial by summing the associative strength of the CS before the trial with the change that occurred following the trial.

Table **2** shows a simple example of the predictions made by the Rescorla-Wagner model for a series of reinforced trial for a CS A, followed by a series of non-reinforced trials for the same CS. The alpha values of CS A and the US were both set to .5 and since the trials were reinforced the maximum associative value the US could reach (λ) was 1 in the first half of the table and 0 in the second half. The example shows the associative strength of the CS at the start of the trial, the amount of change to its associative strength, and the final value of the associative strength at the end of the trial. The previously discussed elements of surprise/expectation from Kamin's theory can be seen in action in Table 2 as the learning that occurs on every trial is dependent on the strength of the association before the trial, the lower the value at the beginning of the reinforced trial the more learning occurs and vice versa. This is due to the model's assumption that conditioning is contingent upon the difference between the associative strength at the beginning of the trial and the maximum associative strength. If the associative strength at the beginning of the trial is low, the difference between this value and the maximum strength value will be large, therefore the larger the change to V is and the more learning occurs on the trial. On every trial V increases and reaches toward λ, the maximum value, however V never reaches this value as the increase of V is asymptotical in nature. The increase of V is also dependent upon the values of α_x and β, these remain constant throughout and they control the speed of conditioning, large values allowing for fast conditioning while low values resulting in slow conditioning.

Table 2

α	β	λ	Trial	V start of the trial	λ − V	ΔV	V end of trial
0.5	0.5	1	A+	.00	1.00	.25	.25
		1	A+	.25	.75	.19	.44
		1	A+	.44	.56	.14	.58
		1	A+	.58	.42	.11	.68
		1	A+	.68	.32	.08	.76
		1	A+	.76	.24	.06	.82
		1	A+	.82	.18	.04	.87
		1	A+	.87	.13	.03	.90
		0	A−	.90	−.90	−.22	.67
		0	A−	.67	−.67	−.17	.51
		0	A−	.51	−.51	−.13	.38
		0	A−	.38	−.38	−.09	.28
		0	A−	.28	−.28	−.07	.21
		0	A−	.21	−.21	−.05	.16
		0	A−	.16	−.16	−.04	.12
		0	A−	.12	−.12	−.03	.09

Rescorla-Wagner Model Predictions for Acquisition and Extinction

The Rescorla-Wagner model encompasses five important assumptions which are central to its success in comparison to other early models of associative learning. The first assumption is that the amount of conditioning that occurs on a given trial is dependent on the associative strengths of all the CSs present on that trial rather than the associative strength of the CS alone. The second assumption is that learnt inhibition and learnt excitation are opposites; this is incorporated in the model by having the two concepts represented with opposite signs. Learnt excitation has a positive sign while learnt inhibition has a negative sign, making the two mutually exclusive in the mathematical model. The third assumption is that the associability of a given stimulus is constant and thus not susceptible to change. The fourth assumption states that the learning which occurs on a given trial for a given stimulus is independent of its associative history (i.e. the change that occurs to the associative strength of a stimulus is solely dependent on its current associative strengths along with the outcome of the trial, and not on how the associative strength was developed, the previous conditioning path of the stimulus is ignored). The final assumption is that the relationship between learning and performance is monotonic (Miller et al., 1995; Rescorla & Wagner, 1972).

The Rescorla-Wagner model became widely popular in the field of associative learning as a result of its simple mathematical formulation along with its ability to account for a wide range of phenomena from simple acquisition and extinction to blocking and conditioned inhibition.

1.3.2 Pearce's Configural Model

Pearce (1987) proposed an alternative model to the one formulated by Rescorla and Wagner, a configural model. This model's central assumption is that organisms encode the entirety of their environment when developing associative learning mechanisms, into their long-term memory. In addition, if something were to change in the environment from trial to trial it was assumed that a new CS-US association would be formed in the organism's long-term memory rather than changing the existing association. As a result, Pearce assumed that when an organism learns about their environment, this learning would be dependent on the similarity between the CSs that have been already encoded and the CSs experienced on a given trial. The learning that occurs on a given trial strengthens all the previously encoded CS-US associations dependent on the similarity between the CSs already encoded and the CSs present in the trial, the larger the similarity the more the association would be consolidated. Learning was deemed to be dependent on the similarity between the CSs because all stimuli were assumed to be composed of individual elements that enter in associations with the outcome. These individual elements are referred to as input units, and when a conditioning trials occurs, it is assumed that a unitary representation of the pattern of stimulation of that trial is formed. The pattern of stimulation is made of individual input units which form a network, each input unit being connected to a considerable number of larger configural units. The model proposed by Pearce uses the following equation:

$$\Delta E_A = \beta(\lambda - \overline{E}_A) \tag{4}$$

$$\overline{E}_A = E_A + \sum_{j=1}^{n} {}_jS_A \times E_j \tag{5}$$

In Equation (4) the ΔE_A parameter represents the change in the associative strength of a CS A, on a given trial. The β parameter serves as the learning rate of the outcome, the US that was used on the trial, and it can take a value ranging from 0 to 1. Similarly to the Rescorla-Wagner model, the λ parameter represents the highest value of associative strength that the US used on the trial can reach, Pearce referring to it as the asymptote of conditioning as this is the asymptotical value the associative strengths of the CSs are reaching towards. The $\overline{E}_A$ parameter represents the aggregate associative strength of CS A, on the trial and is calculated in accordance with Equation (5). In Equation (5) E_A is the associative strength of CS A on the given trial. The second part of the equation, $\sum_{j=1}^{n} {}_jS_A \times E_j$ or e_A in short is the total activation that generalises to CS A from the n number of stimuli which are similar to CS A, and which have been previously reinforced with the same US used to reinforce CS A. Having considered all elements of the mathematical form of the model, the equation shows that the change in the associative strength ΔE_A of a CS (CS A in this example) is dependent on the learning rate of the US (β) used on the trial, and on the aggregate associative strength of CS A. This aggregate associative strength is in turn dependent on the associative strength of CS A at the beginning of the trial as well as all previously trained cues that are similar to CS A.

Table 3 shows an example of the model's predictions for a series of reinforced trials followed by a series of non-reinforced trials. In a simple example like this, when only one cue is being reinforced throughout the training phase, the model behaves almost identically to the Rescorla-Wagner model, it is only when a more complicated design is used when the predictions of the two models start to differ. The examples in the tables differ only because in the early formulation of the model Pearce used only one learning rate parameter β but in a later revision introduced the α parameter too (Pearce, 1987, 1994).

Although the predictions of the models for simple extinction training are very similar showing the associative strength of a cue decreasing, effects such as spontaneous recovery and renewal demonstrated in multiple studies support the idea that extinction does not mean the permanent erasure/destruction of an association, and the way in which each of the models explain these phenomena is when the differences between predictions of the models become more apparent (Bouton, 1994).

The predictions of the two models, plus the predictions of the configural Rescorla-Wagner model for the last extinction study are assessed and compared against each other. The configural Rescorla-Wagner model is derived from the original Rescorla-Wagner model by adding a cue which represents stimulus configuration. This model is discussed in more detail in the final chapter. As previously mentioned these models hold different predictions for more complex cue interactions, and the final aim of the current thesis was to find which of the three models produce the most accurate predictions.

Table 3

Pearce's Configural Model Predictions for a Series of Reinforced and Non-reinforced Trials

β	λ	Trial	E_A before trial	$\sum_{j=1}^{n} {}_jS_A \times E_j$	$\overline{E}_A$	$\lambda - \overline{E}_A$	ΔE_A	E_A end of trial
0.5	1	A+	0.00	0.00	0.00	1.00	0.50	0.50
	1	A+	0.50	0.50	1.00	0.00	0.25	0.75
	1	A+	0.75	0.75	1.50	−0.50	0.13	0.88
	1	A+	0.88	0.88	1.75	−0.75	0.06	0.94
	1	A+	0.94	0.94	1.88	−0.88	0.03	0.97
	1	A+	0.97	0.97	1.94	−0.94	0.02	0.98
	1	A+	0.98	0.98	1.97	−0.97	0.01	0.99
	1	A+	0.99	0.99	1.98	−0.98	~0.00	~1.00
	0	A−	~1.00	~1.00	1.99	−1.99	−0.50	0.50
	0	A−	0.50	0.50	1.00	−1.00	−0.25	0.25
	0	A−	0.25	0.25	0.50	−0.50	−0.12	0.12
	0	A−	0.12	0.12	0.25	−0.25	−0.06	0.06
	0	A−	0.06	0.06	0.12	−0.12	−0.03	0.03
	0	A−	0.03	0.03	0.06	−0.06	−0.02	0.02
	0	A−	0.02	0.02	0.03	−0.03	−0.01	0.01
	0	A−	0.01	0.01	0.02	−0.02	~0.00	~0.00

1.4 Aims

The current thesis had a total of three aims which were restated below.

First, given that inhibition is a wide-spread construct defined in many different ways throughout the literature, the current thesis aims to assess what the relationship between associative inhibition and the wider concept of non-associative inhibition. Based on the review of literature it was noted that more often than not, inhibition measures/subtypes fail to correlate regardless of the seeming similarities between them. Even if the relationship between the various components of non-associative inhibition has been repeatedly assessed using a large variety of measures, associative inhibition was rarely included in these experiments. As a result, this relationship was assessed across a series of four experiments where associative inhibition was either defined as conditioned inhibition, extinction rate, or context inhibition. For all the studies non-associative inhibition was defined as cognitive inhibition, delayed discounting, and response inhibition in order to determine whether associative and non-associative inhibition are independent or related inhibition subtypes.

The second aim of the thesis was to assess different techniques of extinction using compounds in order to reduce recovery and produce a more long-lasting extinction. For this purpose, two studies assessed the effectiveness of super-extinction compared to cue alone extinction only, followed by a comparison between super-extinction, deepened extinction, and cue alone extinction.

The final aim was to evaluate the accuracy of the predictions made by three formal models of associative learning: Rescorla-Wagner, configural Rescorla-Wagner, and Pearce configural model for the three extinction techniques used in the last study: cue alone extinction, super-extinction, and deepened extinction.

Chapter 2 Conditioned Inhibition and Non-associative Inhibition

Associative learning processes allow organisms to adapt to changes in the environment, and inhibitory associative learning is one way to conditionally modify previously learnt behaviours (see Sosa, 2022 and Williams, 1995 for reviews of inhibitory associative learning phenomena). Conditioned inhibition and negative occasion-setting are forms of associative inhibition that can be established when an organism learns that a specific stimulus signals the omission of an otherwise expected event as seen in a simple feature-negative (FN) procedure. Take the example of a rat which learns that it will receive food every time a light flash occurs (A+ trials). In traditional Pavlovian terminology the light is a conditioned stimulus and the food is an unconditioned stimulus, alternatively known as cue and outcome, respectively. If, on some trials, cue A is presented together (in compound) with a second cue, B, a tone, and the outcome does not occur (AB− trials) then cue B may become a conditioned inhibitor or an occasion-setter (Bouton, 1997; Holland, 1992; Rescorla, 1987). As a result, the rat will no longer respond as if it was expecting food on the AB trials. The main difference between B as a conditioned inhibitor and B as an occasion-setter is that the response inhibiting properties of a conditioned inhibitor are general so that responding to a CB compound (a summation test), after C+ trials, would also be suppressed. If B's inhibitory properties were specific to A, then B would be said to have acquired occasion-setting properties (Holland, 1992).

Whether or not training in a FN discrimination will result in the feature (cue B in this example) acquiring conditioned inhibition or occasion setting properties can be determined by procedural as well as individual difference variables. In the case of procedural variables, serial presentation of cues (B then A) is more likely to lead to cue B becoming a specific negative occasion setter for cue A than simultaneous presentation of A and B, which could be due to a mixture of temporal and non-temporal factors (see Holland, 1992 for full review). In contrast, simultaneous presentation tends to result in B becoming a general conditioned inhibitor (Holland, 1992; Swartzentruber, 1995). Recent studies with human participants have provided evidence that there are individual differences in “strategy” adopted given fixed procedures (Glautier & Brudan, 2019; Lee & Lovibond, 2021). To expand, in Experiment 1 of the study by Glautier and Brudan, participants were classified as inhibitors or non-inhibitors based on a summation test carried out in a context that had been used for extinction. In Experiment 2 FN performance of those who had been classed as inhibitors in Experiment 1 was disrupted more than the performance of the non-inhibitors by reinforcing the feature. This pattern would be expected if the inhibitors and non-inhibitors had learned conditioned inhibition and occasion setting, respectively, because reduced responding to the target by the presence of the feature relies on the feature’s association with the outcome in the case of conditioned inhibition. In contrast, for occasion setting, the feature does not control responding by its association with the outcome. Instead, the feature appears to control the operation of the target-outcome association (c.f. Bonardi et al., 2017; Bouton, 1994; Nelson, 2002 for further analysis).

Inhibitory phenomena are not unique to the domain of associative learning. For example, in the literature on impulsivity there is frequent reference to behavioural inhibition which in various forms incorporates a wide range of phenomena including those that fall under the headings of impulsive actions and impulsive choices (Bari & Robbins, 2013). Elaborating further, inhibitory processes in the context of impulsive action would facilitate stopping responses that have already been initiated, and in the context of impulsive choice would facilitate waiting for delayed rewards (e.g. Bari & Robbins, 2013; Broos et al., 2012).

The current series of experiments aims to connect these two areas of research by exploring the relationship between response inhibition produced in a FN predictive learning task as traditionally studied in relation to associative learning under the headings of occasion-setting and conditioned inhibition, and inhibition as traditionally studied in other domains.

Surprisingly, as noted by Sosa and Ramírez (2019; c.f. also Sosa, 2022), there are few studies that have assessed the relationships between the aforementioned forms of inhibition, and in the studies that have, the results have been mixed (He et al, 2013; He et al., 2011; Migo et al., 2006).

Migo et al. (2006) assessed the relationship between conditioned inhibition and scores on Carver and White's (1994) Behaviour Inhibition System/ Behaviour Activation System (BIS/BAS) scales. They unexpectedly found that conditioned inhibition was positively correlated with the BAS-reward responsiveness subscale but no relationship was found between conditioned inhibition and BIS (nor with the other BAS subscales). He et al., (2011) also assessed the relationship between conditioned inhibition and other forms of inhibition by comparing a group of individuals with a history of offending who were characterized by impulsive/violent behaviour to a control group from the general population using their performance on a conditioned inhibition task. The first group was further divided based on whether the criteria for personality disorder (PD) or dangerous and severe personality disorder (DSPD) was met. The control group showed a conditioned inhibition effect in the summation test while the group with a history of offending did not, suggesting that weak conditioned inhibition may be linked to impulsive behaviour, no differences were observed during acquisition. This effect was more notable in the DSPD group. In a follow-up study He et al. (2013) examined the relationship between conditioned inhibition and the BIS/BAS scales in a sample of university students. They found no relationship between inhibitory learning and BAS, but reported a significant negative correlation between the BIS and inhibitory learning. This result was, once again, unexpected based on the assumption that there is a common process underlying conditioned inhibition and response inhibition as measured with the BIS subscales. Thus, as shown in these examples the relationship between conditioned inhibition and the BIS/BAS is not as clear as it might be, but there is some evidence of weaker conditioned inhibition in offenders with a history of impulsive behaviour.

Therefore, the goal in the current investigation was to assess further the evidence for a common inhibitory process that contributes to performance across different domains of inhibition. In particular, the focus was on the relationship between associative inhibition acquired in a FN learning task and four “non-associative” measures of inhibition: a) stopping responses that have already been initiated using the Stop-Signal Reaction Time task (SSRT) and b) stopping responses that would lead to the choice of smaller-sooner rewards in order to obtain larger rewards in a delay-discounting task. These have been selected as examples of non-associative measures of inhibition because of their currency in the literature and, in the case of the SSRT, because the task itself closely resembles the procedure used in FN learning tasks. Additionally the relationship between associative forms of inhibition and two widely used questionnaire-based measures c) the Behavioural Inhibition System/Behavioural Activation System (BIS/BAS) questionnaire (Carver & White, 1994; Patterson & Newman, 1993) and d) the Barratt Impulsivity Questionnaire (BIS-11; Patton et al., 1995) was also assessed. The BIS/BAS questionnaire is derived from Gray's (1982) reward sensitivity theory, which involves the interaction of a behavioural inhibition system and a behavioural activation system. The BIS is assumed to react to novel stimuli and signals for non-reward and punishment by inhibiting ongoing behaviour and this is reflected in the BIS subscales of the BIS/BAS questionnaire. The BIS subscales have items to assess sensitivity to stimuli which are anxiety and fear provoking (Carver & White, 1994; Gray, 1987). The BAS is assumed to react to reward, non-punishment and punishment avoidance by activating reward-related behaviours. Correspondingly, the BAS subscales of the BIS/BAS have items which assess sensitivity to reward-related stimuli (Carver & White, 1994). The BIS-11 assesses impulsivity on a number of sub-scales (e.g. motor, self-control) which contain items directly relevant to inhibition as a complement of impulsivity (e.g. “I act on impulse”, “I am self-controlled”).

In order to separately assess conditioned inhibition and occasion setting the procedure for evaluating associative inhibition involved two stages. First, associative inhibition defined by performance in FN discriminations was assessed. However, as previously mentioned, solving FN discriminations could be due to the participant learning conditioned inhibition or occasion setting but these possibilities cannot be distinguished purely based on the FN discrimination performance. Therefore, in the second stage, conditioned inhibition was assessed in summation tests. These summation tests gave a direct measure of the extent to which each participant had developed conditioned inhibition during the FN phase but, in addition, enabled the classification of participants as inhibitors and non-inhibitors. Since inhibitory and non-inhibitory strategies are relatively stable within individuals (Glautier & Brudan, 2019) the FN discrimination performance was re-examined separately for inhibitors and non-inhibitors. To allow for conditioned inhibition to be assessed strict learning criteria have been applied to ensure that learners have been selected for the analysis, these criteria are described separately for each study below.

Three experiments are presented in the current chapter, the first and last aiming to assess the relationship between associative and non-associative inhibition to determine if the two subtypes of inhibition are independent or whether they rely on common underlying mechanisms. Additionally a pilot study[1] was used to test a new conditioned inhibition design following study 1.

[1] Study 1 was a lab-based study, and so was the pilot study, however the data collection of the pilot study was stopped by the COVID-19 pandemic, and study 2 was carried out online.

2.1 Study 1

The current study aims to assess the relationship between associative and non-associative inhibition, and investigate whether the underlying inhibitory mechanisms involved are distinct or whether they stem from a common underlying inhibitory construct. Of all the previously mentioned inhibition subtypes, conditioned inhibition and response inhibition are most similar, both in the way they are defined and in the manner in which they are assessed. For these tasks, a participant would have to respond to a cue, but withhold responding when the cue is presented in compound with another cue (the conditioned inhibitor or the stop signal). The key difference between a conditioned inhibition learning task and a stop signal reaction task is the fact that the stop signal is presented with a varying delay after the go signal as opposed to the learning task during which the CS and CI are presented simultaneously. In addition there is a time limit to respond to the cues in the stop signal reaction task, while the cues usually remain on display until a response is given during an associative learning task. Because of the time pressure and delay along with the lack of other cues, the stop signal reaction task measures strictly the ability of a participant to stop a response that was already initiated. The horse-race model (previously discussed) is applied to understand and interpret the results of the stop signal task, which assumes that the go and stop signals are competing to reach completion. The speed of the two signals is assumed to vary between participants due to some individual differences, some participants being slower or faster. If the same model is applied to a conditioned inhibition task and we assume that the CS and CI both initiate a signal (go and stop respectively), then the failure to show conditioned inhibition might be due to the participant's stop signal reaction rather than a failure to learn. Some participants might learn the meaning of the conditioned inhibitor but might be unable to stop responding to the CS because the CS signal wins the race with the stop signal and reaches completion. As a result, out of all the inhibition measures used, conditioned inhibition and response inhibition are assumed to be the most likely to be associated.

To facilitate a more in-depth exploration of this assumed relationship, the conditioned inhibition task used in the current study was customised to resemble a stop signal reaction task by adding a time limit and having two groups, a control and a delay group. For the control group only an overall time limit was added, but for the delay group a short delay was included between the presentation of the CS and the CI (full details provided in the method section). If response inhibition is associated with or plays a role in the display of conditioned inhibition it would be expected that all participants in the control group, regardless of their response inhibition, would be able to show conditioned inhibition. For the delay group however, only participants with high response inhibition (fast stop signal reactions) are expected to show signs of conditioned inhibition.

2.1.1 Method

2.1.1.1 Participants

A total of 118 student participants (of which 108 identified as female and 10 identified as male, with a mean age of 19.35 years, SD = 1.49) were recruited to take part in the current study from the Southampton University Highfield Campus. Participants were awarded course credit for their participation and the study took between 45 minutes and one hour to complete.

2.1.1.2 Questionnaires

Three questionnaire-based measures were used: 1) The BIS/BAS scales (Carver & White, 1994), 2) the BIS-11 (Patton et al., 1995), and 3) an adjusting amount delay discounting questionnaire. The delay discounting questionnaire consisted of 10 blocks of choices between hypothetical monetary rewards. Each choice was between a smaller immediate reward and a later larger reward. The blocks used five delays: one week, one month, six months, one year, and two years. Each delay was used twice, once in an ascending block and once in a descending block. Questions were all of the form "Would you prefer S now or L in D?" where S was a (variable) small sooner reward value, L was a (fixed) large later reward value and D was the delay until L. In ascending blocks S started at £5 and each time the participant chose L, S would increase in the next question until chosen. S was one of £5, £100, £250, £550, £800, £950, £990, and £1000 whereas L was always £1000. In descending blocks S started at £1000 and each time the participant chose S it would be decreased in the next question until L was chosen. This procedure allowed the estimation of indifference points (average of indifference points obtained in ascending and descending sequences) at each delay, following which, least-squares non-linear regression was used to fit Mazur's hyperbolic delay discounting (Equation 1) to the indifference points (Mazur, 1987). From this, a discounting parameter k was extracted for each participant, which was used as a measure of response inhibition – larger k values suggest weak response inhibition, corresponding to a pattern of impulsive choices biased towards smaller sooner rewards. These three questionnaires were the same for all studies presented that assessed non-associative inhibition.

2.1.1.3 Stop Signal Reaction Time Task

The fourth and final measure of non-associative inhibition was the stop signal reaction time (SSRT) which was assessed using the STOP-IT task designed by Verbruggen et al. (2008), which is a free software under the GNU General Public License (available at: https://expsy.ugent.be/tscope/stop.html). As part of the task participants were presented with either a square or a circle on the screen and were asked to press the "\" key when the square was presented and the "Z" key when the circle was presented. The instructions specified that the participants should aim to be as fast and accurate as possible. The participants were also instructed that on some trials a tone might follow the square or the circle and that they should not respond on these trials. The tone was the stop signal and it was presented with a varying delay after the square/circle cues participants were instructed to respond to the primary cues as soon as they appeared and not wait to see if there was going to be a stop signal or not. The tone was present on 25% of the trials, and participants had 32 initial practice trials. The start value of the delay was 250 ms, following every response to the stop signal the delay was automatically adjusted by 50 ms. If the participant failed to stop in time then the delay was decreased by 50 ms and if the participant succeeded in stopping the delay was increased by 50 ms. This procedure aimed to find the delay at which each participant showed a probability of stopping of 50%. Using the delay time of each participant the stop signal reaction time of the participants was individually calculated. The SSRT is a measure of the time required by the stop signal to reach completion, and for the current study large SSRT values are interpreted as reflective of weak response inhibition since the signal requires more time to reach completion.

2.1.1.4 Learning Task

Participants completed a custom built learning task, as part of which participants were asked to learn about the patterns presented inside a series of square shaped stimuli. During the task, one or two squares were presented on the screen, each with a predetermined pattern made out of randomly chosen shapes and colours (Appendix A). Some of the squares and pairs of squares were followed by a red flash, and participants were instructed to try and predict the red flashes as accurately as possible by pressing the “R” key whenever they thought the stimuli presented on the screen would be followed by a flash. The patterns used were randomly selected for each participants, while the outcome (the red flash) was the same for all participants. The task consisted of 90 trials, 5 practice trials, 72 acquisition trials, 2 summation trials, and 20 recovery trials. The stimuli were presented for a limited time, .875 of a second, and participants were instructed to make a prediction during while the stimuli were present on the screen. Furthermore, there were two groups to which participants were randomly assigned prior to the start of the task: delay and no delay. For the delay group, when a compound of stimuli was presented, instead of the stimuli being shown simultaneously one of the stimuli was presented with a delay of .250 of a second. The putative conditioned inhibitor was always the cues presented with a delay for the feature negative discrimination. For the no delay group, the compounds were presented simultaneously.

2.1.1.5 Design

Table 4 shows the design used in the learning task, it consists of three stages: acquisition, summation, and retardation. In each of the stages the trials were randomly ordered independently for each participant. The acquisition trials were organised in blocks consisting of three presentations of each of the cues, meaning that no more than four trials of the same type could occur in succession. The acquisition therefore comprised 4 blocks during which cue B was trained to become a conditioned inhibitor using a feature negative discrimination procedure. This was achieved by reinforcing cue A when it was presented alone but not when it was presented in compound with cue B. Cue C was the test cue for the summation test and received reinforcement training during acquisition. The DE compound was used to show that a pairing of cues is not enough to lead to non-reinforcement, therefore the pair received reinforced training. Cues F and G were non-reinforced throughout the acquisition stage with the aim of balancing the number of reinforced and non-reinforced cues. Following acquisition, two summation tests were presented in random order the transition to the test stage was not explicitly signalled. For the summation test, cue C which received reinforced training throughout the acquisition was presented in compound with cues B and N. Cue B was the putative conditioned inhibitor and cue N was a novel stimulus which was assumed to be associatively neutral. The suppression of responding to the test cue C caused by B was compared to response rates to cue C in the last block of training and the control compound CN. Finally, a retardation test was carried out using the conditioned inhibitor B, the neutral stimulus N, along cues F and G from the acquisition stage. During this test the first two cues were always reinforced while the latter two cues were always non-reinforced. All cues were presented five times, resulting in 20 total trials which were split into five blocks, each block containing only one presentation of each of the cues.

Table 4

Design of the Learning Task for Study 1

Acquisition	Summation	Retardation
A+ x12	CB− x1	B+ x5
AB− x12	CN− x1	N+ x5
C+ x12		F− x5
DE+ x12		G− x5
F− x12		
G− x12		

2.1.1.6 Data Selection and Analysis

Three of the 118 participants have been excluded for failing to correctly complete some parts of the experiment resulting in them having unusable or incomplete datasets. For the remaining 115 participants an exclusion criterion based on the acquisition performance was applied. The criterion focused on selecting participants who showed overall learning which allowed for the analysis of the participants' learning rates during the feature negative discrimination. The purpose of the exclusion criterion was to ensure that participants' responses are due to learning and can be interpreted as such when assessing the acquisition rate of the feature negative discrimination, the summation and retardation tests, otherwise non-responding during training and test could be attributed to some extraneous factors other than learning. For the exclusion criterion the last two blocks of the acquisition stage were used, from which the responses to cues C, DE, F, and G were selected. As a result, 12 trials were selected per participant and only participants who responded correctly to 10 or more of these trials were included in the final sample. The binomial distribution was used to determine the cut-off point, given that guessing is defined as p(success)= .5, the probability of correctly responding to 10 out of the 12 selected trials is less than .05. A total of 67 participants were excluded using this criterion leaving 48 participants (24 per group) in the final sample used for the data analysis.

The analysis was computed using R (R Core Team, 2021). The data analysis consisted of generalised linear mixed models, non-parametric ANOVAs, and multiple regressions.

First, the analysis focused on the acquisition stage using a linear mixed model (lmer4 package version 1.1.27.1) for binary data more specifically, the feature negative discrimination performance during training was used as a dependent variable. The model employed a maximum likelihood criterion to estimate the parameters along with a logit link function. To facilitate the development of the model, the feature negative performance for every participant was transformed into a binary vector with 12 elements (corresponding to the 12 acquisition trials). Values of 1 indicated that the participant responded correctly to both components of the feature negative, while values of 0 were indicative of one or two errors (i.e. a correct prediction of a flash for the A+ trial and a correct prediction of no flash for the AB− would be represented by 1, all other possible responses were coded as 0). The model development took place in two stages starting with a full factor model which was trimmed down by removing non-significant predictors to create the final model. Consequently, in the first stage, group and trial were used as fixed factors (group had two levels: delay vs no delay; trial 0-11) and participant was included as a random factor to allow for individual intercepts to be computed for every participant. The trial fixed factor was reverse coded (reverse trial = 12 – trial, i.e. trial 12 coded as 0, trial 1 coded as 11 etc.) which allowed for the intercepts to be interpreted as the terminal performance at the end of training. The intercepts therefore, represented the probability of the participant responding correctly to the feature negative discrimination at the end of the acquisition stage (trial 0 after reverse coding). The model included both a linear and a quadratic trend for trial, interactions between these trends and group, and it allowed individual intercepts and slopes to be computed for every participant. The slopes were interpreted as the acquisition rate of the feature negative discrimination during acquisition.

Next, for the second stage all non-significant predictors and interactions were removed from the model. From this final model the individual slopes and intercepts were extracted and used in a series of multiple regressions as dependent variables as part of which the standardised non-associative measures of inhibition (BIS11, BIS, BAS, DD, SSRT) were used as predictors.

For the summation test two non-parametric repeated measures ANOVAs (Friedman's ANOVA) were used, separately for the delay and no delay group, to assess the differences in responding between CB, CN, and C (the last trial of acquisition was used for cue C), followed by pairwise comparisons (in the form of Wilcoxon matched pairs). For each group there was a total of three comparisons to be made, therefore a Bonferroni correction was employed for the follow up comparisons. The pairwise comparisons aimed to examine the difference in responding to C when the cue was: presented alone at the end of the acquisition, in compound with a putative conditioned inhibitor B, and in compound with a novel cue N. The expectation was that the condition inhibitor B would reduce responding to cue C (C vs CB) more than the novel cue (CB vs CN), since it was trained using a feature negative discrimination.

Using the summation test performance, participants were then classified as inhibitors or non-inhibitors (Glautier & Brudan, 2019). The aim of the classification was to first assess whether the two procedures affected whether participants were classed as inhibitors and non-inhibitors, and second to assess any differences in non-associative inhibition between the two groups. Following the classification, the feature negative performance from the acquisition stage was revisited and the multiple regression models examining the relationship between acquisition performance and non-associative inhibition were updated to include the new classification along with interactions between the classification and the non-associative inhibition.

The retardation test was treated in the same manner as the feature negative discrimination from the acquisition stage. First, a linear mixed model for binary data was used to assess any differences in responding to the two cues of interest, B and N between the two groups. The model had trial and group as fixed factors, and participants as random factors. The model allowed for a linear and quadratic term, and participants were allowed to have individual linear and quadratic slopes as well. From this final model the slopes and intercepts were extracted and used a measure of acquisition, the slopes represented the speed of acquisition while the intercepts (where trial was again reverse coded) represented the terminal response rates for cues B and N.

2.1.2 Results

2.1.2.1 Non-associative Inhibition

Table 5 shows the means and standard deviations of all non-associative inhibition measures for the 48 participants who passed the learning criterion. All mean values were within the ranges of the heathy general population reported across the literature, and there were no significant differences between the two groups (Jorm et al., 1998; Klein et al., 2022; Lipszyc & Schachar, 2010; Stanford et al., 2009).

Table 5

Non-associative Inhibition Descriptive Statistics

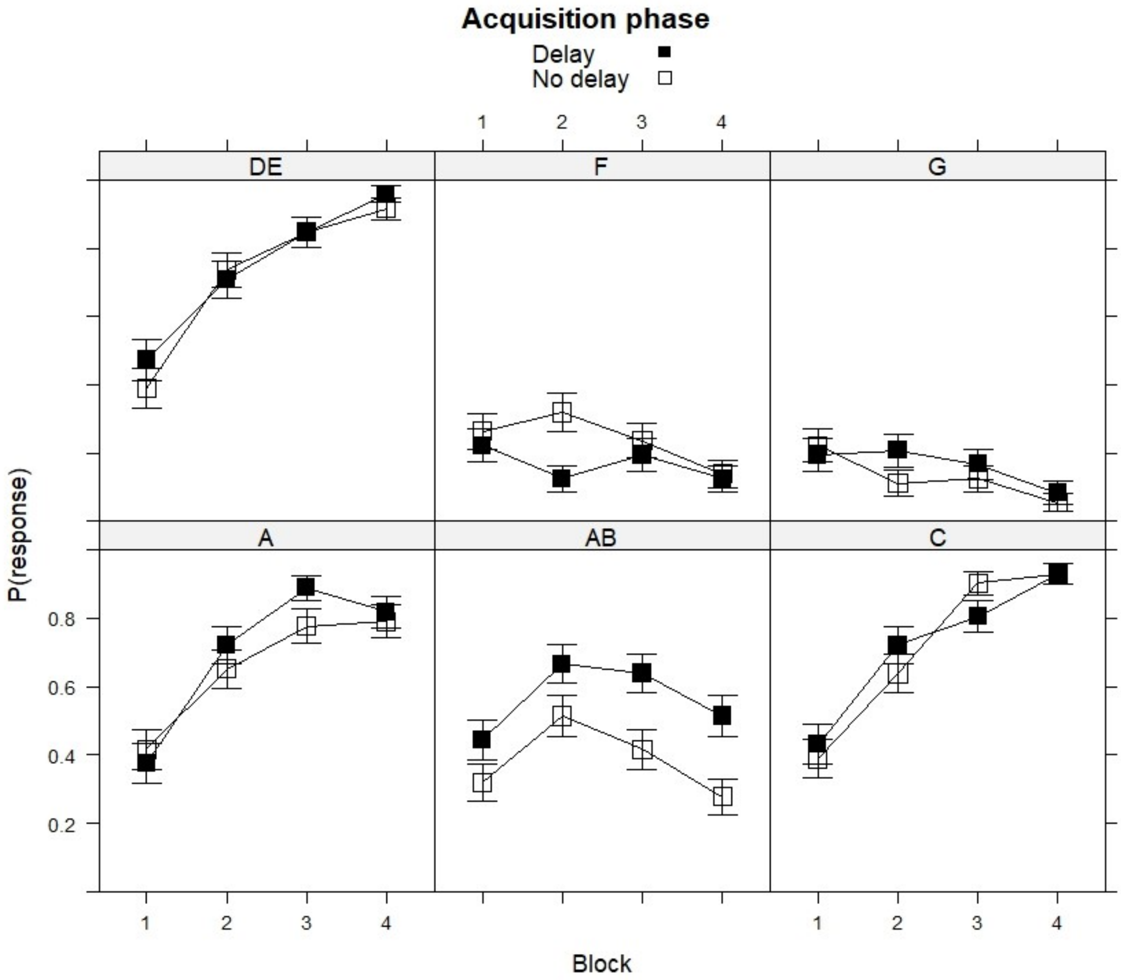
	Mean	Standard Deviation
BIS11	60.77	7.13
BIS	23.10	2.82
BAS	38.71	5.72
k	0.01	0.03
SSRT	255.76	37.66

2.1.2.2 Acquisition

Figure 2 shows the probability of responding to all the cues used in the acquisition stage divided by the group participants were assigned to. The 48 participants who passed the exclusion criterion had successfully learnt to respond more to reinforced than non-reinforced trials during acquisition, regardless of the group they were assigned to. The only noticeable difference between the two groups was the probability of response to the AB compound, the no delay group had a lower probability of responding than the delay group. This was expected due to the fact that B was presented with a delay which meant that some participants could have responded to cue A before B appeared, which made the feature negative discrimination more difficult for this group.

Figure 2

Probability of Responding during the Acquisition Stage by Cue and Group



2.1.2.2.1 Feature Negative Discrimination

To assess whether there was a difference in the speed with which the two groups learnt the feature negative discrimination an initial model with group (delay and no delay) and trial (0-11, both linear and quadratic) as fixed factors and participants as random factors was defined. As part of the model individual intercepts and slopes (linear and quadratic) were computed for every participant. The model revealed that the effect of group was significant along with the linear effect of trial, while the quadratic effect of trial and all other interactions were found to be non-significant (Table 6). According to this model the linear trends for trial was more appropriate than the quadratic term, as a result this was excluded from the model. The model also revealed a significant difference between the two groups, however this was just an overall difference in responding rather than a difference in the acquisition pattern, therefore all group and trial interactions were also removed.

The final model revealed that the effects of trial (linear) and group were still significant after the exclusion of the above mentioned factors (Table 6). As expected the effect of trial indicated that participants learnt to correctly respond to the feature negative discrimination more as the acquisition stage progressed. The effect of group confirmed that there was a difference between the overall response patterns of the two groups, the no delay groups showing more overall correct responses to the feature negative discrimination compared to the no delay group. The individual intercepts and slopes were extracted from this model and were used in the subsequent analysis as measures of feature negative discrimination learning.

Table 6

Feature Negative Discrimination Learning over Time by Group

Model	Fixed Effect	Estimate	SE	z	p
Group * Trial	Intercept	-1.23	0.25	-5.00	< .001*
	Trial (linear)	-9.98	4.64	-2.15	.03*
	Trial (quadratic)	-2.84	3.81	-0.75	.46
	Group	0.80	0.34	2.36	.02*
	Group * Trial (linear)	-2.64	6.13	-0.43	.67
	Group * Trial (quadratic)	5.00	5.00	1.00	.32
Group + Trial	Intercept	-0.46	0.34	-1.37	.17
	Trial (linear)	-0.13	0.04	-3.57	< .001*
	Group	0.72	0.31	2.29	.02*

2.1.2.2.1.1 Non-associative Inhibition

Using the intercepts and slopes extracted from the model as outcomes and non-associative measures of inhibitions as predictors, two multiple regression were computed. Group was also added as a predictor to the two regressions along with the interactions between group and the non-associative measures of inhibition. The regressions revealed a significant negative effect of BIS on the feature negative discrimination slopes. According to this effect participants who scored high on BIS had steeper slopes and therefore learnt the feature negative discrimination faster (Table 7, Figure 3). The effect of BIS on the intercepts, and the interactions between BIS and group (for both regression models) were not significant, however all were approaching significance. All other effects were not significant (Table 7).

Table 7

Effects of Non-associative Inhibition on Feature Negative Discrimination Learning

DV	R^2	dfs	F	p
FN Intercept	.16	11, 36	0.65	.78

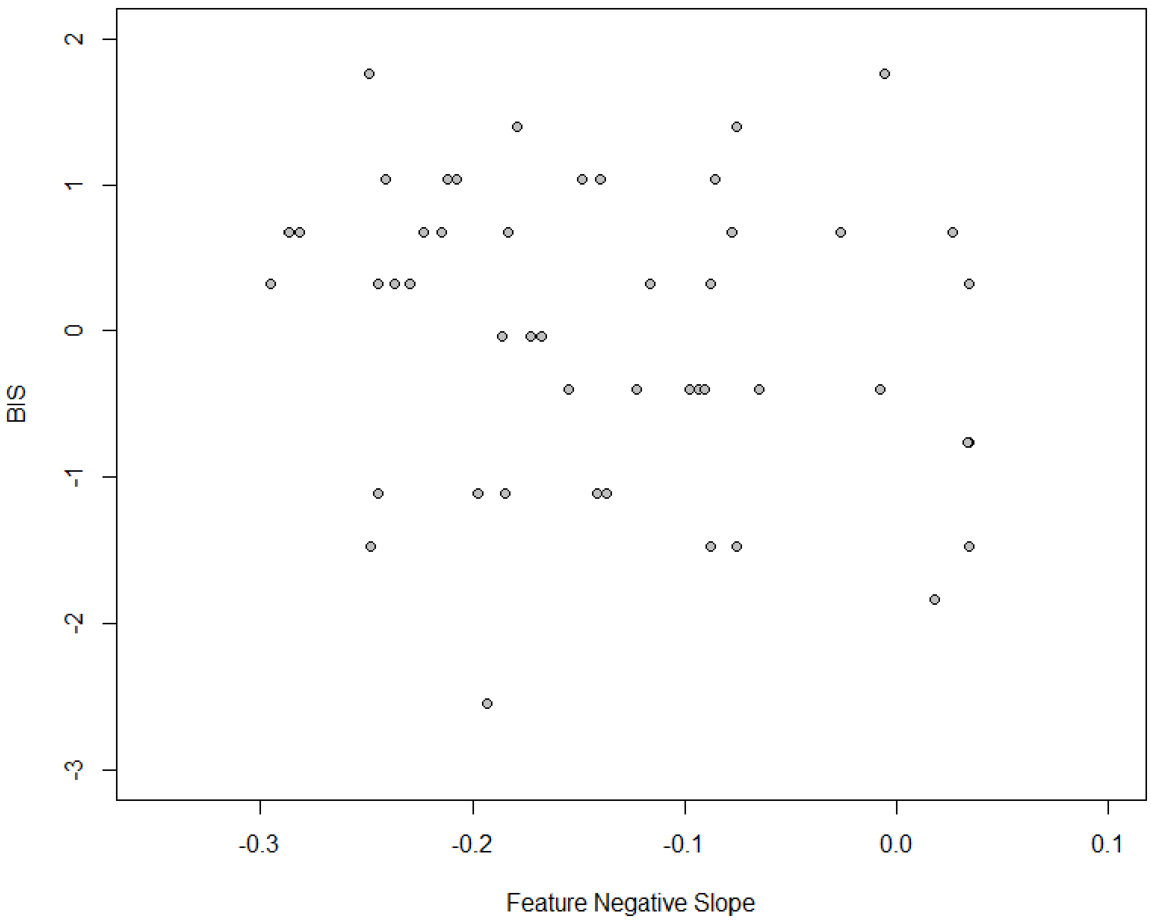
Non-associative Inhibition	Unstandardized β	t	p
Intercept	-0.02	-0.13	.90
BIS11	-0.10	-0.43	.67
BAS	0.06	0.26	.80
BIS	0.47	1.94	.06
DD	0.12	0.58	.57
SSRT	-0.25	-0.99	.33
Group	0.03	0.08	.93
Group*BIS11	0.07	0.23	.82
Group*BAS	-0.10	-0.28	.78
Group*BIS	-0.65	-1.84	.07
Group*DD	-0.07	-0.17	.87
Group*SSRT	0.13	0.38	.71

DV	R^2	dfs	F	p
FN Slope	.18	11, 36	0.70	.73

Non-associative Inhibition	Unstandardized β	t	p
Intercept	-0.13	-6.30	< .001
BIS11	0.01	0.50	.62
BAS	-0.01	-0.38	.70
BIS	-0.05	-2.13	.04*
DD	-0.01	-0.69	.49
SSRT	0.02	0.74	.46
Group	-0.001	-0.02	.99
Group*BIS11	-0.01	-0.32	.75
Group*BAS	0.01	0.44	.66
Group*BIS	0.07	2.01	.05
Group*DD	0.01	0.25	.81
Group*SSRT	-0.01	-0.24	.81

Figure 3

Effect of BIS on the Feature Negative Discrimination Learning (Slope)

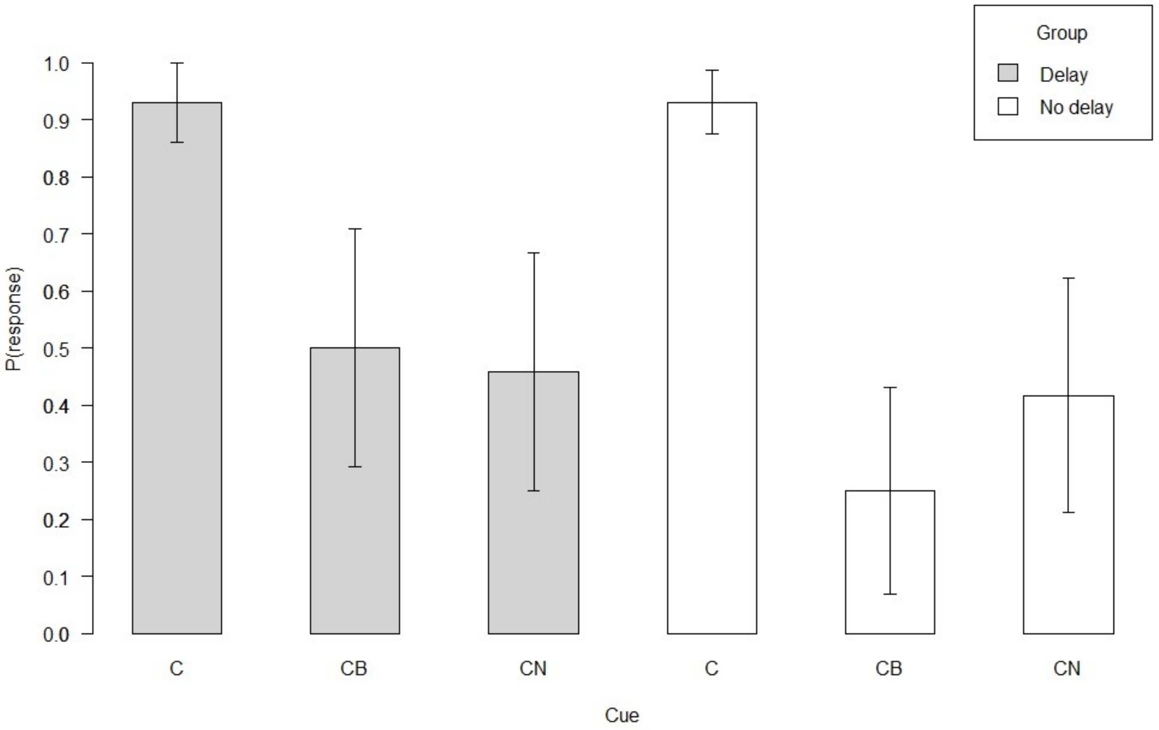


2.1.2.3 Summation Test

The summation test performance for the 48 participants who passed the exclusion criterion is shows in Figure 4. For the two groups, both the putative conditioned inhibitor (B) and the novel stimulus (N) reduced responding to C, however the responses to both compounds appear to be comparable.

Figure 4

Summation Test Performance by Group



Two Friedman's ANOVA were used to assess the summation test performance of the two groups separately. The tests confirmed that for both groups there was a significant difference in the way participants responses to C in the last block of acquisition and the two summation compounds CB and CN ($X^2(2) = 9.65$, $p = .008$; $X^2(2) = 20.82$, $p < .001$, for the delay and no delay groups respectively).

For the delay group, the follow-up Wilcoxon matched pairs tests with a Bonferroni correction showed that both cues B and N reduced responding to C significantly when compared to the response rates of C during the last block of acquisition ($Z = -3.18$, $p = .001$, $r = - .46$; $Z = -3.34$, $p < .001$, $r = - .48$, respectively). When comparing the two compounds however, there was no significant difference in response rates ($Z = -0.28$, $p = .78$, $r = - .04$).

For the no delay group, the same pattern was found, both cues B and N have significantly reduced responding to C, however there was no significant difference between responding to the two compounds (Z = -3.91, p < .001, r = - .56; Z = -3.48, p < .001, r = - .50; and Z = -1.27, p = .21, r = - .18, respectively).

2.1.2.3.1 Non-associative Inhibition

Based on their responses in the summation tests participants were classified as inhibitors and non-inhibitors. Participants were labelled as inhibitors if their summation test performance showed that the conditioned inhibitor B reduced responding to C more than the novel stimulus N (CB < CN). In the delay group, there were 17 inhibitors and 7 non-inhibitors, while in the no delay group there were 12 inhibitors and 12 non-inhibitors. A chi-square test confirmed that this difference was not statistically significant $X^2(1, N = 48) = 2.18$, p = .14. Using this binary classification, a logistic multiple regression was computed to determine whether the training group participants were allocated to or the measures of non-associative inhibition had an effect on how the participants were classified into inhibitors or non-inhibitors. The regression also assessed the interactions between group and all measures of non-associative inhibition. The only significant effect was the interaction between group and BIS, this effect indicated that in the delay condition participants who had scored higher on BIS were more likely to be labelled as inhibitors, while in the no delay condition participant who scored higher on BIS were more likely to be labelled as non-inhibitors (Figure 5, Table 8). None of the remaining effects were significant (Table 8).

Table 8

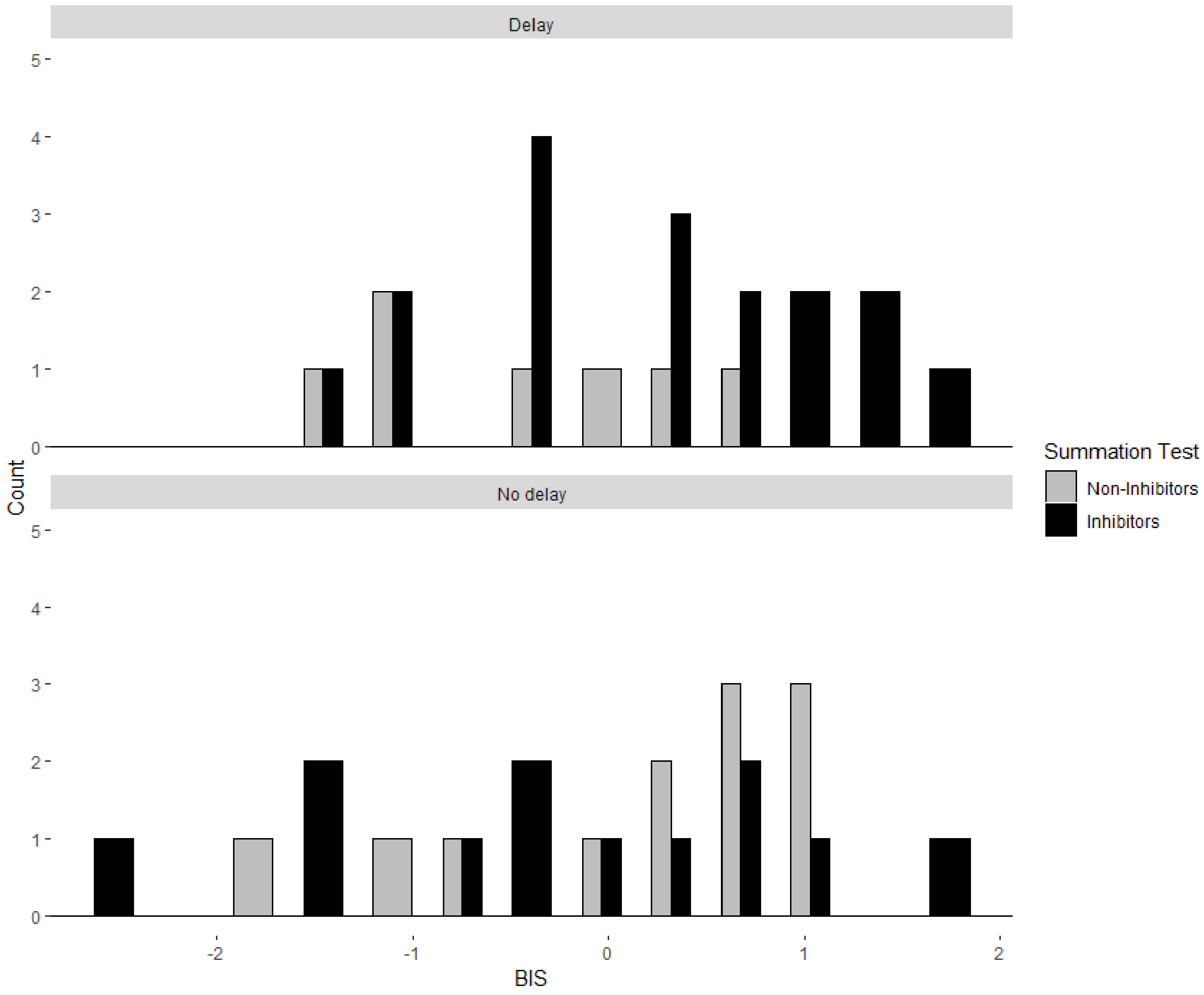
Effects of Non-associative Inhibition and Group on Summation Test Performance

Model	**Cox & Snell R^2**	**McFadden R^2**	**dfs**	**X^2**	**p**
	.24	.20	1, 36	12.84	.30

Non-associative Inhibition	**Estimate**	**Wald Statistic**	**p**
Intercept	-0.19	0.16	.69
BIS11	-1.27	2.76	.10
BAS	0.55	1.22	.27
BIS	-1.05	2.28	.13
DD	-0.33	0.53	.47
SSRT	-0.29	0.27	.61
Group	1.36	3.29	.07
Group*BIS11	1.44	2.50	.11
Group*BAS	-0.18	0.05	.83
Group*BIS	1.96	4.15	.04*
Group*DD	0.21	0.06	.81
Group*SSRT	-0.18	0.05	.82

Figure 5

Summation Test Performance by Group and BIS



2.1.2.4 Feature Negative Discrimination and Non-associative Learning

The two multiple regression models used at the beginning of the analysis to assess the effect of non-associative inhibition on the feature negative discrimination performance were recomputed after the addition of an inhibition factor (inhibitors vs non-inhibitors) along with the interactions between the new inhibition factor and all measures of non-associative inhibition. To simplify the models, the group factor has been removed, and instead each regression model was computed separately for the delay and the no delay groups. None of the effects were significant (Table 9; Table 10).

Table 9

Effects of Associative and Non-associative Inhibition on Feature Negative Performance for the Delay Group

DV	R^2	dfs	F	p
FN Intercept	.43	11, 12	0.83	.62

	Non-associative Inhibition	Unstandardized β	t	p
	Intercept	-0.84	-1.37	.20
	BIS11	-0.26	-0.71	.50
	BAS	-1.06	-1.69	.12
	BIS	-0.93	-1.05	.31
	DD	0.02	0.05	.96
	SSRT	0.28	0.27	.32
	Inhibition	0.99	1.53	.15
	Inhibition*BIS11	0.25	0.58	.57
	Inhibition*BAS	1.36	2.01	.07
	Inhibition*BIS	0.86	0.94	.37
	Inhibition*DD	0.50	0.90	.39
	Inhibition*SSRT	-0.75	-2.00	.07

DV	R^2	dfs	F	p
FN Slope	.44	11, 12	0.85	.60

	Non-associative Inhibition	Unstandardized β	t	p
	Intercept	-0.05	-0.92	.38
	BIS11	0.02	0.55	.59
	BAS	0.10	1.67	.12
	BIS	0.08	0.94	.37
	DD	-0.0001	-0.004	.99
	SSRT	-0.03	-1.17	.26
	Inhibition	-0.10	-1.53	.15
	Inhibition*BIS11	-0.02	-0.44	.67
	Inhibition*BAS	-0.13	-1.97	.07
	Inhibition*BIS	-0.07	-0.79	.44
	Inhibition*DD	-0.05	-0.94	.37
	Inhibition*SSRT	0.07	2.00	.07

Table 10

Effects of Associative and Non-associative Inhibition on Feature Negative Performance for the No Delay Group

DV	R^2	dfs	F	p
FN Intercept	.39	11, 12	0.68	.73

	Non-associative Inhibition	Unstandardized β	t	p
	Intercept	-0.32	-0.78	.45
	BIS11	-0.03	-0.08	.94
	BAS	0.95	1.14	.28
	BIS	0.59	1.18	.26
	DD	0.07	0.17	.87
	SSRT	-0.92	-1.08	.30
	Inhibition	0.32	0.49	.63
	Inhibition*BIS11	-0.48	-0.58	.57
	Inhibition*BAS	-1.16	-1.23	.24
	Inhibition*BIS	-0.29	-0.41	.69
	Inhibition*DD	-0.16	-0.23	.82
	Inhibition*SSRT	0.62	0.60	.56

DV	R^2	dfs	F	p
FN Slope	.38	11, 12	0.68	.73

	Non-associative Inhibition	Unstandardized β	t	p
	Intercept	-0.11	-2.71	.02*
	BIS11	0.003	0.11	.92
	BAS	-0.08	-1.07	.30
	BIS	-0.06	-1.28	.23
	DD	-0.01	-0.26	.80
	SSRT	0.07	0.88	.40
	Inhibition	-0.03	-0.47	.65
	Inhibition*BIS11	0.04	0.55	.59
	Inhibition*BAS	0.10	1.16	.27
	Inhibition*BIS	0.02	0.37	.72
	Inhibition*DD	0.01	0.17	.87
	Inhibition*SSRT	-0.04	-0.42	.68

2.1.2.5 Retardation Test

Figure 6 shows the participants' responses to cues B and N during the retardation test. To assess the participants' performance during the retardation test a linear mixed model with trial (reverse coded 0 to 4), cue (B vs N), and group (delay vs no delay) which also allowed for individual slopes and intercepts to be computed was defined. The model revealed that the effect of cue and group were not significant along with all interactions meaning that no significant retardation effect was detected (Table 11). Since responding to the two cues did not differ, no further tests were carried out for the retardation test.

Figure 6

Retardation Test Performance

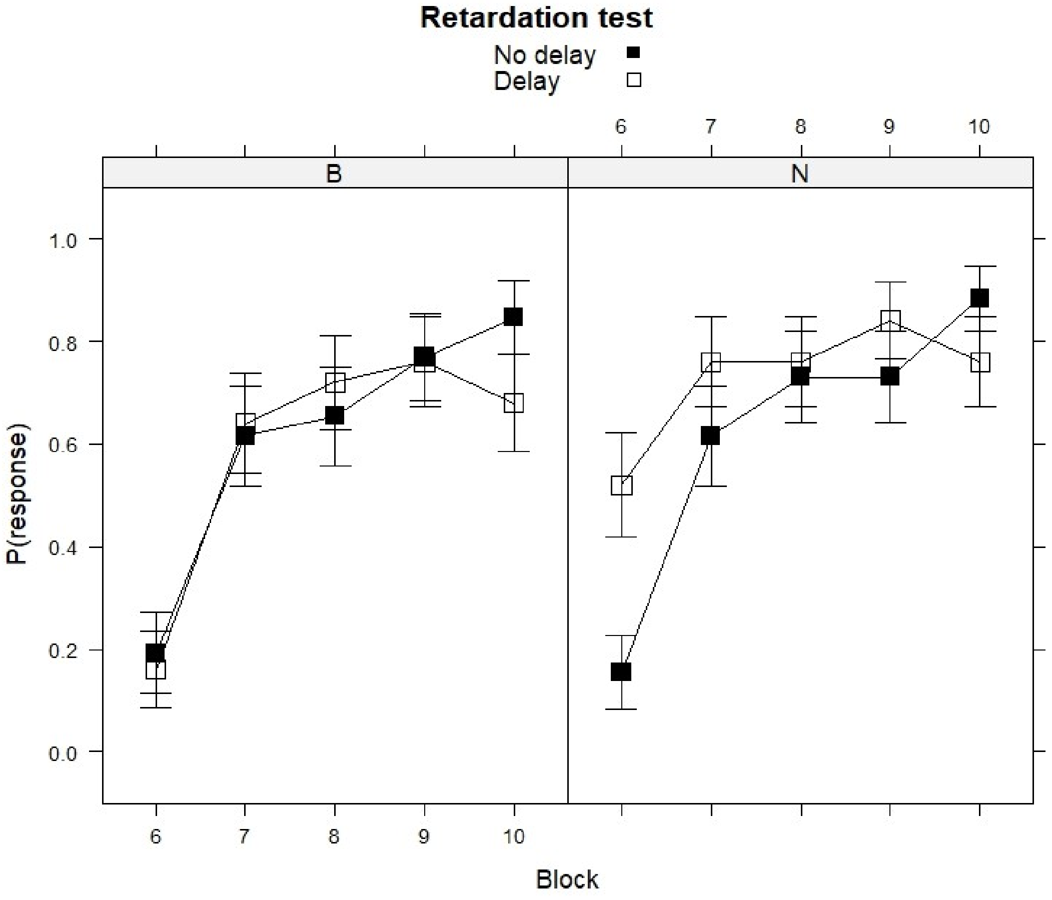


Table 11

Retardation Test Performance

Fixed Effect	Estimate	SE	z	p
Intercept	3.29	0.88	3.75	< .001*
Trial	-0.80	0.20	-3.95	< .001*
Cue	-0.34	0.87	-0.40	.69
Group	0.32	1.19	0.28	.78
Trial*Cue	0.28	0.24	1.18	.24
Trial*Group	-0.11	0.28	-0.40	.69
Cue*Group	0.88	1.28	0.69	.49
Trial*Cue*Group	-0.42	0.35	-1.21	.23

2.1.3 Discussion

In the current experiment a feature negative discrimination was used to train conditioned inhibition using two groups while also capturing non-associative inhibition using four distinct measures with the overall aim of assessing whether associative and non-associative inhibition share common underlying mechanisms. The two groups (delay vs no delay) were found to differ significantly when their feature negative discrimination responses were analysed, the no delay group showing more overall correct responses than the delay group (Table 6). This was to be expected as the delay added a degree of difficulty to the task and it was initially hypothesized that only participants who had fast SSRTs would be able to stop responding in time and show evidence of the feature negative discrimination learning during training. This hypothesis was not confirmed and none of the non-associative measures of inhibition were found to significantly predict training performance, with the exception of BIS (Table 7). BIS was found to have a significant effect on the speed of learning of the feature negative discrimination, participants who had high BIS scores learnt the feature negative discrimination faster. Although not significant, the effect of BIS on the intercepts along with the interactions between BIS and condition for both the slopes and intercepts were approaching significance. When assessing the summation test it was apparent that the current learning task has failed to obtain a robust conditioned inhibition effect. For both groups, although cue B significantly reduced responding to cue C, the reduction in response rates was not different from the reduction resulted from pairing cue C with a novel stimulus S. Participants were also classified as inhibitors and non-inhibitors based on their summation test performance, none of the non-associative measures of inhibition or the group were found to be significant predictors of this classification (Table 8). The only exception was the interaction between group and BIS which suggested that in the no delay condition participants who had higher BIS scores were more likely to be classified as non-inhibitors, while the reverse effect was observed for the delay group. This classification was not found to be a significant predictor of feature negative discrimination learning. During the retardation test no robust effect of inhibition was found and no further tests were carried out, along with the summation test results this suggests that the current learning task might have failed to produce strong conditioned inhibition.

Due to the overall lack of evidence that the current learning task has trained a reliable conditioned inhibition (associative inhibition) effect, the conclusions that can be drawn are limited. The large exclusion rate also raises concerns regarding the reliability of the associative inhibition effect, as 58% of the participants were excluded for failing to pass the learning criterion. Throughout the analysis the relationship between BIS and associating inhibition has been consistently present, however due to the previously mentioned reasons the current experiment can only be interpreted as partial evidence for a potential link between the two. Similarly, for the same reasons the lack of evidence in support of a relationship between the remaining measures of non-associative inhibition and conditioned inhibition cannot be considered definitive and uncertainty around the relationship remains. Furthermore, it seems unlikely that the lack of an overall conditioned inhibition effect was caused by occasion setting. Although not statistically significant, there were more inhibitors (participants showing conditioned inhibition) than non-inhibitors (participants showing performance indicative of occasion setting) in the delay group, in contrast in the no delay group there was an equal number of inhibitors and non-inhibitors. This could be due to the design of the learning task which was modified to resemble a stop signal reaction task and the target was presented before the feature in the feature negative discrimination for the delay group. It is generally agreed that a serial feature negative discrimination leads to occasion setting rather than conditioned inhibition, however the feature is presented before the target in order for this to be achieved. The key to the serial presentation is that the feature doesn't acquire strength by being presented first, when it is presented second it could become inhibitory (Holland, 1992; Holland & Lamarre, 1984; Rescorla, 1986). This seems to be the case in the current experiment, emphasized by the larger number of inhibitors in the delay group.

Currently in the literature, inhibition is defined in many various ways, the concept of inhibition being considered to be multidimensional with many, sometimes unrelated underlying factors all of which are usually referred to under the general umbrella term of impulsivity (Evenden, 1999; Paulsen & Johnson, 1980). When studied together, most of these subtypes of inhibition show little to no relationship, other than a similarity in the way they are defined: "inability to stop certain processes/mechanisms" (Reynolds et al., 2006). Using some of the most replicated underlying components of inhibition, Bari and Robbins (2013) put forward a structure for the concept of inhibition consisting of two main subdivisions of inhibition: behavioural, and cognitive. For the behavioural factor, there were three more underlying components: response inhibition, deferred gratification, and reversal learning. This is one of the few models which includes both associative and non-associative inhibition, usually associative inhibition is not considered when impulsivity/inhibition are discussed.

Furthermore, although the multidimensionality of impulsivity/inhibition is widely accepted, currently there is no comprehensive agreed upon model of the underlying structure. When building this structure however, associative inhibition should be considered alongside non-associative inhibition as a standalone factor given the current study suggests little to no relationship with the other non-associative measures of inhibition.

Associative and non-associative inhibition have been rarely studied together, the studies by Migo et al. (2006) and He et al. (2013) being one of the few examples of the crossover of the two concepts. The two studies had opposite results, Migo et al (2006) finding a significant relationship between associative learning and BAS-reward but no significant association with BIS, while He et al. reported a significant negative relationship between inhibitory learning and BIS but not significant association with BAS. Together these studies show the need for further research aimed at understanding the relationship between associative and non-associative inhibition. The current study did not replicate the findings of Migo et al (2006), but did find evidence to support the link between BIS and non-associative inhibition that was also found by He et al. (2013)

Although the results of the current study support some of the existing literature, there are concerns regarding the validity and robustness of the results due to the high exclusion rate and the lack of a robust overall inhibition effect. As a result, the design of the learning task was updated, tested in a pilot study, and the study was re-run with the updated methodology.

2.2 Learning Task Update

The learning task used to train conditioned inhibition in the previous study failed to produce a reliable conditioned inhibition effect, the conditioned inhibitor did not pass either the summation or retardation test (design shown in Table 4). Additionally, the task had a very large exclusion rate, 58% of the participant failed to pass the exclusion criterion. As a result, the design of the task was changed and a pilot study was run to assess the reliability/feasibility of the new design (Table 12). The changes made and the results of the pilot study are presented below.

Table 12

Updated Design for the Conditioned Inhibition Task

Acquisition	Summation
A+ x12	CI− x1
B+ x12	CN− x1
C+ x12	
IA− x12	
IB− x12	
JA+ x12	
JB+ x12	
K− x12	
L− x12	
M− x12	

First, the time pressure element of the task was removed, the cues were presented for three seconds instead of 0.875 of a second. Similarly, the delay group was also removed, all compounds were presented simultaneously for all participants in the new task. The aim of the time pressure and the delay was to accentuate a potential existing relationship between the stop signal reaction time of the participants and their ability to learn conditioned inhibition, however no indication of this association was found. The overall time pressure could have also been responsible for the large exclusion rate as it could have made the learning task too difficult for the participants.

Second, the feature negative discrimination was doubled with the aim of training a strong conditioned inhibition effect. Williams (1995) assessed the strength of the conditioned inhibition effect produced by a simple, double compound, and double elemental feature negative discrimination. In Experiment 1 the simple FN group was exposed to a traditional feature negative discrimination design P_1+/P_1N−, the double compound group was exposed to a double FN discrimination P_1+/ P_1N− and P_2+/P_2N−, and the double elemental group received a simple FN discrimination training with additional non-reinforced presentations of the feature P_1+/P_1N−/ N−. Williams (1995) found that the single group failed to show a conditioned inhibition effect, while the two double groups (compound and elemental) showed strong conditioned inhibition effects. Consequently, a double feature negative discrimination was chosen for the updated design of the learning task.

Finally, the retardation test was removed in order to simplify the design of the task. The two cues used in the summation and retardation test as part of the previous learning task were the same, meaning that the two tests could have affected one another given they were consecutive and the transition between stages was not signalled. Furthermore, as previously discussed a novel cue could reduce responding simply due to the fact that it is novel, therefore a neutral cue might be a better control.

The updated design of the task is shown in Table 12. The task consisted of 122 trials, 120 training trials and two test trials. During training, the trials were organised into six blocks, each block containing two trials of each type, therefore no more than three trials of the same type could occur in succession. All trials were randomly ordered within each block (including the summation test block). As part of the acquisition stage, a dual feature negative discrimination was used to train cue I to become a conditioned inhibitor. Accordingly, two cues A and B were reinforced when presented alone, but non-reinforced when presented in compound with cue I. The A and B cues were also presented as part of reinforced compounds with cue J to signal that a compound is not enough for the target cues to be non-reinforced. Cue C was the transfer test cue and cues K, L, and M were non-reinforced during training to balance the number of reinforced and non-reinforced cues.

The stimuli used were the same, therefore participants were asked to learn about the patterns within a series of square shaped cues. The patterns were made out of preselected patterns and colours that were randomly selected for each participant. Some of the patters and combinations of patters were followed by a red flash (used on reinforced trials) and participants were asked to press the "R" key when they thought a square or a pair of squared were going to be followed by the flash. Participants were instructed to maximise the number of correct predictions and minimise the number of errors.

2.2.1 Pilot Study

For the pilot study a total of 70 participants, students and visitors at the University of Southampton, were recruited via efolio (University based online study advertisement board for Psychology studies). No demographic information was recorded.

To test the new design the previously set methodology was used. First, the exclusion criterion was applied to the acquisition data followed by a non-parametric ANOVA with follow-up pairwise comparisons to assess the participants' performance during the summation test. For the exclusion criterion the last block of training was used for cues C+, AJ+, BJ+, K-, L-, and M- which meant that a total of 12 trials were used. Participants who responded correctly to 10 or more of the 12 trials were considered learners and their data was included in the analysis of the summation test. The cut-off was chosen based on the binomial distribution which shows that the probability of correctly responding to 10 out of 12 trials is less than .05 if the participant was guessing (p(success)= .5). Using this criterion a total of 32 of the 70 participants were excluded from the analysis.

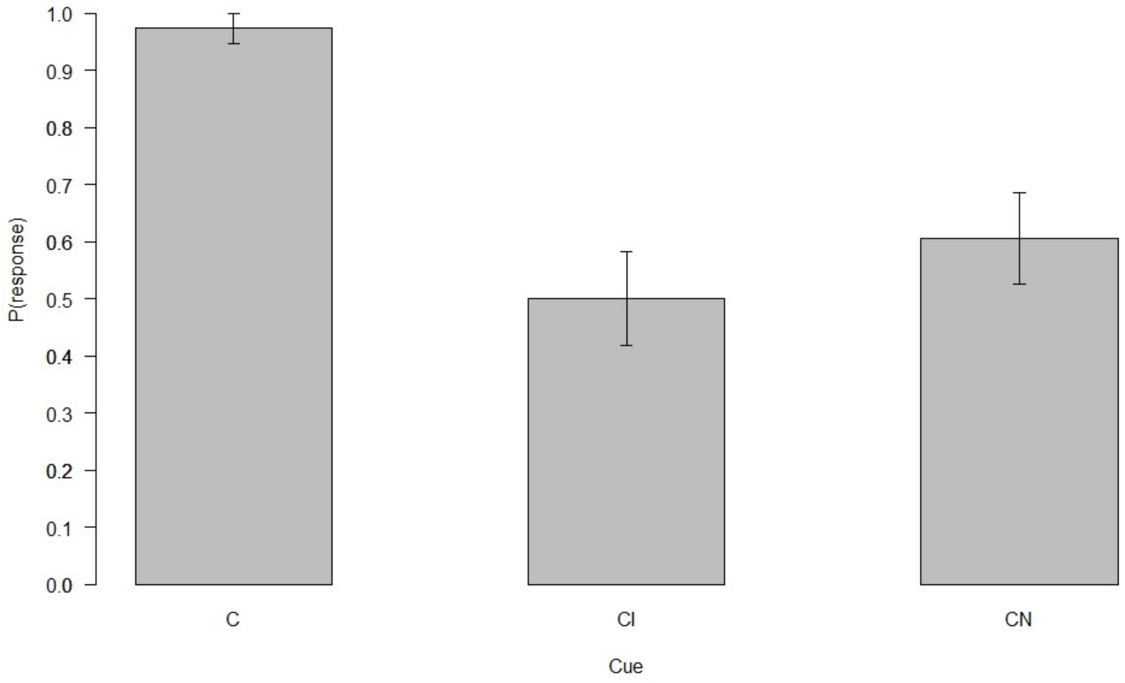
2.2.1.1 Results

The summation test performance for the 38 participants left in the sample is shown in Figure 7. According to Figure 7 the conditioned inhibitor I reduced responding to C slightly more than the novel stimulus N, however the difference was small.

A Friedman's ANOVA revealed that there were significant differences in responding to the three cues included in the summation test ($X^2(2) = 14.37$, $p < .001$). The follow-up Wilcoxon matched paired tests confirmed that both the putative conditioned inhibitor and the novel stimulus significantly reduced responding to C ($Z = -4.14$, $p < .001$, $r = - .11$, and $Z = -3.64$, $p = .003$, $r = - .10$, respectively), however there was no significant difference between responding to the two compounds ($Z = -1.07$, $p = .99$, $r = - .28$).

Figure 7

Summation Test Performance using the Updated Learning Task Design



Overall, the changes to the design have lowered the exclusion rate and improved the conditioned inhibition effect observed. The exclusion rate for the previous experiment was

58% while the exclusion rate for the current design was 46%. During summation, the participants who passed the learning criterion showed the expected pattern of response, the conditioned inhibitor reduced responding to the target cue more than a novel control cue. This difference although in the expected direction was not significant, which could be partly due to the relatively small sample or to the task itself, the patterns used for the cues might have been too difficult to remember/distinguish.

Prior to the re-run of the study a scoping review of the associative learning literature has been carried out to assess the features of previously used learning tasks/procedures.

2.2.2 Conditioned Inhibition Design Scoping Review

A total of 31 studies have been identified which provided the design of the conditioned inhibition training and testing used. The aim of the review was to identify ways of improving the design used in the previous experiment in order to obtain a reliable conditioned inhibition effect. Table 13 shows the summarised information extracted from these studies, which includes whether the study was an animal (A) or a human (H) study, the technique used for training conditioned inhibition, other key features of the conditioned inhibition training and testing, and whether or not the training was reported as successful.

One study included both humans and animals, 11 were animal studies and 19 were human studies. The majority of the studies used a feature negative discrimination to train conditioned inhibition or a variation of this procedure. Three studies have reported no conditioned inhibition effect while the remaining 18 have reported a significant conditioned inhibition effect. When it comes to testing whether the conditioned inhibition training was successful the studies had different approaches, seven studies used both a summation and a retardation test, one study used a retardation test only, and 21 used a summation test only. Out of the 21 that used a summation test alone, nine used a neutral control cue, six used a novel control cue, three used both, and three used no controls. Finally, out of the studies that used a summation test, 11 used a scale rating to assess participants' expectations.

As identified during the review a common practice for the studies that used a summation test is to use a neutral control cue, therefore a neutral control cue was introduced to the new learning task along with the novel control which was retained. As previously mentioned attention could play an important role in conditioned inhibition, more salient cues attracting more attention. As a result a novel control cue might influence the results of a summation test, the cue reducing responding dimply due to its novelty and salience, therefore a neutral cue is more appropriate for a control as its salience is comparable to the other pre-exposed cues. Furthermore, a large number of studies used a scale rating during the summation test which could make the detection of inhibition easier compared to a predictive test. As a result the new test featured two summation tests, one predictive and one evaluative. During the predictive test participants were asked to use the response key to indicate whether or not they expected the cues to be followed by the outcome. For the evaluative test, a scale from 0 to 100 was used. To prevent the predictive summation test being treated as non-reinforced trials, the new learning task signalled the transition from acquisition to test and then from the predictive to the evaluative test. Participants were informed that during the tests no outcomes would be used and that they should use their previous experience to indicate whether they believe the cues presented would have been reinforced or not. The new task retained the double feature negative discrimination, and did not include time pressure. Additionally, Lee and Lovibond (2021) showed that a stronger inhibition effect was obtained when a causal relationship between the cues and the outcome was implied, therefore a causal relationship was implied between the cues and the outcome as part of the new learning task.

The new learning task had a "game-like" design, participants were asked to learn about what a friendly unidentified life form (FULF) likes to eat. Pictures of foods in the form of fruit and vegetables were used as cues while happy, sad, or neutral reactions of FULF were used as outcomes. The task is described in detail in the following experiments.

Table 13

Scoping Review of Conditioned Inhibition Studies

Study	Participants	Design	Design Features	CI
Alarcón and Bonardi (2015)	H	A → O1 AB → O1 AX−	Feature negative discrimination A+/AX− with additional AB+ trials. Summation test with neutral cue. Scale ratings for summation.	Yes
Amundson, Wheeler, and Miller (2005)	A	A+ AX−	Feature negative discrimination A+/AX−. Varied the number of A+ trials, more A+ trials led to more inhibition. Summation test with no control.	Yes
Baetu and Baker (2010)	H	Phase 1 A+ AB− Phase 2 A+ B−	Feature negative discrimination A+/AB− in phase 1 plus A+/B− in phase 2. Summation tests with neutral cues. Scale ratings for summation.	Yes
Burger, Denniston, and Miller (2001)	A	A+ AX−	Feature negative discrimination A+/AX−. Retardation test only with novel cue.	Yes
González, Alcalá, Callejas−Aguilera, and Rosas (2019)	H	Ci− P+	Obtained conditioned inhibition by presenting Ci without reinforcement while reinforcing the rest of the cues. Summation and retardation tests in separate studies with novel cue.	Yes
Grillon & Ameli (2001)	H	A+ X→A−	Feature negative discrimination A+/ X→A−, serial presentation of the compound, the inhibitor (X) preceding A. Summation test with neutral cue.	No
Harris, Kwok, Andrew, and Harris (2014)	A	CS+ CSL−	Feature negative discrimination CS+/CSL−, + represented a higher reinforcement rate while – represented a lower reinforcement rate. Summation with no control cues.	Yes Cont…

He, Cassaday, Howard, Khalifa, and Bonardi (2011)	H	A+ AZ+ AP−	Feature negative discrimination A+/AP− with additional AZ+ trials. Summation test with neutral cues. Scale ratings for summation.	Yes
Horne and Pearce (2010)	A	Stage1 A+ Stage2 A+ AX−	Feature negative discrimination A+/AX− with prior A+ training. Summation with no control cues and retardation with novel cue.	Yes
Karazinov and Boakes (2004)	H	P+ PI− I− PA+	Feature negative discrimination P+/PI− with additional PA+ and I− trials in experiment 1. Summation test with both novel and neutral cues in all studies.	Yes
	H	P+ PI− I−	Feature negative discrimination P+/PI− with additional I− trials in experiments 2 and 3.	
Karazinov and Boakes (2007)	H	Block 1-3 P+ Block 4 PX−	Conditioned inhibition was not the main aim but obtained a conditioned inhibition effect using a feature negative discrimination P+/PI−. Scale rating for summation.	Yes
		P+ PX−	Summation test with neutral cue in first experiment and both neutral and novel in second.	Yes
Laing et al. (2021)	H	A+ AX− X− AD+ D+	Feature negative discrimination A+/AX− plus additional X− and AD+ trials. Summation test with neutral and retardation test with novel cue. Scale ratings for summation.	Yes
Laing, Burns, & Baetu (2019)	H	E+ EF−	Feature negative discrimination E+/EI−. Summation test with neutral cue. Scale ratings for summation.	No

Cont…

Lee and Livesey (2012)	H	A_1+ A_1X_1− A_2+ A_2X_2−	Dual feature negative discrimination. Aim of testing the effect of time pressure; time pressure did not lead to conditioned inhibition but second-order conditioning.	No
	H	Stage 1 A+ AX− Stage 2 A+ AX−	Summation test with neutral cues in both experiments. Scale ratings for summation.	No
Lee and Lovibond (2021)	H	A+ AB−	Feature negative discrimination A+/AB−. Summation tests with both novel and neutral control cues. Scale ratings for summation along with predictive ratings and open-ended questions.	Yes
Lotz and Lachnit (2009)	H	A+ AX− B+ BY−	Dual feature negative discrimination. Unidirectional conditioning showed better conditioned inhibition compare to bidirectional. Summation test with novel cue. Scale ratings for summation.	Yes
Lotz, Vervliet, and Lachnit (2009)	H	A+ AB−	Feature negative discrimination A+/AB−. Summation test with neutral cue. Scale rating for summation.	Yes
Melchers, Wolff, and Lachnit (2006)	H	Stage 1 A+ AX− Stage 2 A+ X−	Feature negative discrimination A+/AX− in stage 1 followed by A+/X− in stage 2. Unidirectional conditioning showed conditioned inhibition while bidirectional did not. Summation test with novel cue. Scale rating for summation.	Yes
Migo et al. (2006)	H	A+ ACI− B+ BCI−	Dual feature negative discrimination. Summation test with novel cue. Scale ratings for summation.	Yes
Miguez, McConnell, Polack, and Miller (2018)	A	L+ LCI−	Feature negative discrimination L+/LCI−. Showed that conditioned inhibition can transfer between contexts.	Yes Cont…

Miguez, Soares, and Miller (2015)	A	P+ PX− Q+ QY−	Dual feature negative discrimination. Showed that conditioned inhibition is context specific only when it is the second learnt relationship of the target cue.	Yes
Neumann, Lipp, and Siddle (1997)	H	A+ AB− B− AC+ C+	Feature negative discrimination A+/AB− plus B− trials and AC+, where C was the transfer cue. Summation test with novel cue.	Yes
Polack, Laborda, and Miller (2012)	A	D:Z+ B:Z−	Showed that contexts can become conditioned inhibitory using a feature negative discrimination where the context is the feature.	Yes
		D:Z+ D:Y− B:Z−	Summation and retardation tests with novel controls for both experiments	Yes
Redhead and Chan (2017)	A and H	AX+ AW+ AY−	Aimed to show that spatial learning follows associative learning models. Summation test with novel and retardation test with neutral cue.	Yes
Richardson, Michener, Gann, North, and Schachtman (2020)	A	A+ AX−	Aimed to show that both conditioned inhibition and CS alone extinction trials produce a CS that passes both a summation and retardation test.	Yes
Sansa, Rodrigo, Juan Jose, and Chamizo (2009)	A	A+ AZ−	Aimed to show that spatial learning follows associative learning models. Summation test with novel cue.	Yes
Stout, Escobar, and Miller (2004)	A	A+ AX−	Feature negative discrimination A+/AX−. Showed that the more training trials are used the better the conditioned inhibition effect. Summation test with no control.	Yes
Urcelay and Miller (2008)	A	A+ AX− X−	Feature negative discrimination A+/AX−. Showed that when a feature negative discrimination alone is used it is more effecting than when X− are also included. Summation and retardation tests with neutral control cues.	Yes Cont…

Urcelay, Perelmuter, and Miller (2008)	H	Stage 1 AX− Stage 2 A+	Feature negative discrimination backward conditioning: AX−/A+. Showed that backward conditioning produces conditioned inhibition. Summation and retardation tests with novel control cues. Scale ratings used for summation.	Yes
Williams (1995)	H	Single P_1+ P_2+ P_1N− Double/C P_1+ P_2+ P_2N− P_1N− Double/E P_1+ P_2+ N− P_1N−	Used three groups: a single feature negative discrimination, double feature negative discrimination and a feature negative discrimination with the non-reinforcement of the feature. The double discriminations produced the strongest inhibition. Summation test with novel cue.	Yes
Zaksaite and Jones (2019)	H	D+ DE^0	Feature negative discrimination or simple non-reinforcement, with the aim of assessing the effect of the non-reinforcement type (0 = no change/− decrease). Scale ratings for summation. Summation test with neutral cue.	Yes
		D+ DE^0		

2.3 Study 2

Using the updated learning task the study aiming to assess the relationship between associative and non-associative inhibition was run again. At each stage, participants were carefully selected for analysis using specific learning criteria. This was necessary to ensure the specificity of the tests for associative inhibition. For example, a failure to show response suppression in a summation test for conditioned inhibition could be due to failure to learn the preceding FN discrimination or due to the adoption of an occasion-setting strategy. By excluding from the analysis of the summation test those who failed to learn the FN discrimination it was ensured that weak suppression in the summation test was indeed indicative of weak general conditioned inhibition.

2.3.1 Method

2.3.1.1 Participants

The study was based on a sample of 133 participants (of which 70 identified as male, 60 identified as female and 3 preferred not to say, the mean age was 34.89 years, *SD* = 13.03) recruited via Prolific (https://www.prolific.co/). Participants were each paid £2.50 for taking part in an online experiment that involved completing a series of questionnaires and behavioural tasks which, altogether, took approximately 30 minutes to complete. All tasks were presented on web servers running at the University of Southampton.

2.3.1.2 Questionnaires

The study used the same questionnaires as Study 1 (see section 2.1.1.2 Questionnaires for full details).

2.3.1.3 Stop Signal Reaction Time Task

An online version of the STOP-IT task (Verbruggen & Logan, 2008) was used to measure SSRT as an index of response inhibition capacity. The task was developed following principles highlighted in a guide for measuring response inhibition (Verbruggen et al., 2019) and is available under a GNU license on Github (https://github.com/fredvbrug/STOP-IT). For this task participants were presented with left and right pointing arrows (with a black outline and white fill) and were asked to indicate the direction of the arrows using the left and right arrow keys. On some trials participants were presented with a stop signal (the arrow would turn red) to indicate they must not respond. The stop signal was presented with a variable delay after the arrow first appeared. The delay was adjusted depending on the participants' responses. Failure to stop responses led to a decrease in the delay while success in stopping responses led to an increase in the delay. The delay adjustments were made to converge on a value which resulted in a 50% successful stop rate; that value was taken as the estimate of the stop signal reaction time. The SSRT represents the time needed for the response generated by the stop signal to reach completion and large SSRTs were interpreted to be a reflection of weak response inhibition.

2.3.1.4 Learning Task

Participants took part in a custom built "game-like" learning task programmed using jsPsych. Participants were introduced to the learning task by being told that they are part of a research team that is trying to find what a friendly unidentified life form (FULF) likes to eat. The learning task consisted of 116 trials, 110 acquisition trials and 6 test trials. On each trial participants were presented with cues (either one or two images of foods) followed by FULF's reaction, an outcome, which was a tummy ache or no tummy ache. The cues were randomly selected (from a selection of 11 images) for each participant while the outcomes were the same for all participants, tummy ache being used on reinforced trials (+ trials) and no tummy ache was used for the non-reinforced trials (− trials) (Appendix B). Participants were instructed to respond while the food was present, before seeing the reaction, in order to predict FULF's reaction. The instructions also asked participants to try and maximize the number of correct predictions and minimise the number of incorrect predictions. The foods were present for two seconds during which the participants had to make a prediction, next the participants were shown the outcome for one and a half seconds, and finally a fixation cross was presented for a further two seconds before the next trial started. Additionally, after completing the learning trials, participants were asked to first predict then rate the likelihood of specific food item combinations to cause a tummy ache, in a predictive then evaluative summation test.

2.3.1.4.1 Design

The design used to train conditioned inhibition is shown under the acquisition phase in Table 14. Also in Table 14, after acquisition, there were two test blocks, each containing three conditioned inhibition summation tests. Trials in each phase were randomly ordered independently for each participant subject to the constraint that, in the acquisition phase, no more than two trials of each type could occur in succession. Thus, there were 11 blocks of 10 trials each containing one of each trial type. During this stage cue I was trained to become a conditioned inhibitor by using a dual demonstration. The dual demonstration, during which a conditioned inhibitor indicates non-reinforcement in compound with two separate excitatory cues, has been shown to facilitate acquisition of conditioned inhibition compared to a single demonstration (Williams, 1995). Accordingly, cues A and B were reinforced when they were presented alone, but not when they were presented in compound with cue I. Additionally, cues A and B were reinforced when presented in compound with cue J to highlight the fact that it was not enough for cues A and B to be presented in compound in order for them to be non-reinforced, but they need to be in compound with I, the conditioned inhibitor (Williams, 1995). Finally, cues K, L, and M were presented non-reinforced so that there were equal numbers of reinforced and non-reinforced trials on the single cue trials, as well as on the compound cue trials. After acquisition there were two conditioned inhibition summation test blocks. During the first test block (the predictive response summation test) participants responded using the keys just as they had done in the acquisition phase, the transition to the test was explicitly signalled. In the second test block (the evaluative summation test) participants were asked to rate the likelihood of a tummy ache occurring on a scale from 0 to 100. In each summation test, excitatory cue C was presented in compound with the putative inhibitor I. Suppression of responding to test compound CI was assessed relative to responding to C alone in the last block of the acquisition phase and relative to compounds of C with two 'associatively neutral' control stimuli i.e. CN and CK. Cue N was novel, but in previous unpublished studies in this laboratory, strong suppression of responding to compounds containing novel stimuli was observed that may obscure, through floor effects, the differences between CI and CN. CK was therefore used as a second compound to compare with CI.

Table 14

Design of the Learning Task for Study 2

Acquisition	Summation: Predictive response	Summation: Evaluative response
A+	CI−	CI?
B+	CN−	CN?
C+	CK−	CK?
AI−		
BI−		
AJ+		
BJ+		
K−		
L−		
M−		

Note. Reinforcement, tummy ache, is denoted as + while non-reinforcement, no tummy ache is shown as "–". In the evaluative summation test, "?" indicates that participants were asked to rate the likelihood of a tummy ache on a scale from 0 to 100 rather than using the training keys. Each trial type was presented 11 times in the acquisition phase and once in each of the two summation tests.

2.3.1.4.2 Task Instructions

Based on the recent research of Lee and Lovibond (2021), to further facilitate the training of conditioned inhibition, a causal component was included in the instructions. Lee and Lovibond (2021) showed that implying a causal relationship between cues (the foods in this case) and outcomes (the tummy states in this case) could lead to more robust conditioned inhibition effects. Therefore, our instructions included the following: "So far the research team suspects that there is at least one food which causes FULF to have a tummy ache. Also there may be another food that suppresses FULF's tummy ache.".

2.3.1.5 Data selection and analysis

Eighteen of the 133 participants were excluded entirely from the analysis for failing to complete some parts of the experiment resulting in them missing scores on one or more measures. For the remaining 115 participants two sequential exclusion criteria were applied which ensured that participants met critical learning thresholds for assessment of associative inhibition, as measured in a) FN discrimination performance and then in b) the conditioned inhibition summation tests. The study aimed to assess variation in strength of inhibition in the FN discrimination and in summation tests so it was needed to select suitable participants independently of their performance in these parts of the experiment. To ensure that performance in the FN discrimination was indicative of strength of associative response inhibition it was elected to exclude non-learners from the analysis of FN performance – poor FN discrimination would not indicate weak response inhibition learning in participants who were simply failing to learn the task overall due to inattention, failure to understand and/or follow task instructions, or due to cognitive overload. Learners were therefore defined on the basis of their responses on trials that were not part of the FN discrimination during the last two blocks of the acquisition phase i.e. the last two C+, AJ+, BJ+, K-, L-, and M- trials. This defined 12 trials and participants responding correctly on 10 or more trials were classed as learners. Participants responding correctly on less than 10 trials were classed as non-learners and excluded from further analysis. This cut-off was chosen using the binomial distribution; with p(success)= .5 defined as guessing, the probability of getting 10 or more successes on 12 trials is less than .05. Application of this criterion excluded 16 participants leaving 99 whose FN data was analysed below. The second exclusion criterion was then applied to select participants for analysis of conditioned inhibition in the summation tests. Again, since it was intended to study variation in performance in this task to assess strength of conditioned inhibition, participants who failed to learn the FN discrimination were excluded. Learning the FN discrimination is a necessary (but not sufficient) condition for acquiring conditioned inhibition and it was intended to distinguish between failure to learn the FN discrimination and weak conditioned inhibition. The performance in the last two blocks of the FN discrimination (the last two A+, B+, AI-, and BI- trials) was used to define an 8 trial performance criterion. Participants with 7 or more trials correct on this basis were included in the analysis of the conditioned inhibition summation tests (binomial distribution p(success = .5) 7 or more successes on 8 trials $p < .05$). This excluded a further 24 participants leaving 75 participants whose conditioned inhibition data was analysed below.

All analyses were carried out using R (R Core Team, 2021). The main data analyses used generalised linear mixed models, parametric, and non-parametric ANOVAs, and multiple regressions.

For the analysis of the FN discrimination a generalised linear mixed model (lmer4 package version 1.1.27.1) for binary data was computed using FN discrimination performance and block as the dependent variables. The model estimated the parameters using a maximum likelihood criterion and a logit link function. For each participant performance was encoded in a 22 element binary vector with 1s indicating correct responses on both components of the FN discriminations in a block (e.g. an outcome prediction on an A+ trial and no outcome prediction on an AI- trial would be coded 1 but any other pattern would be coded 0). There were 11 blocks for each of the two FN discriminations (A+/AI- and B+/BI-) hence the 22 element binary vector. The model was computed in two stages. In the first stage, discrimination and block were used as fixed factors (discrimination, two levels: FN A+/AI- versus FN B+/BI-, coded 0,1 and block: 0-10) and participant as a random factor, meaning that an individual intercept was computed for every participant. Block was reverse coded (e.g. block 11, coded 0, block 1 coded 10). Reverse coding of block allowed interpretation of the intercepts as terminal performance, intercepts reflecting the probability of the participant responding correctly in the FN discriminations at the end of the acquisition phase. For this initial model, block was allowed to have both a linear and a quadratic term. This model was used to confirm that the two FN discriminations were not learned at different rates.

For the second stage, since the FN discriminations were not learned at different rates, the model was updated by removing the discrimination factor and allowing random quadratic slopes for participants in the random structure. Only quadratic slopes were included in the model as they reflected the performance of the participants more accurately than the linear ones, furthermore by excluding the linear slopes the intercepts could be interpreted as performance at the end of training. The slopes reflect the rate of acquisition of the FN discrimination. The slopes obtained from this second generalised linear mixed model for each participant were then used as measures of FN discrimination performance and included in a series of multiple regressions as dependent variables with the (standardised) non-associative measures of inhibition (BIS11, BIS, BAS, DD, SSRT) as independent variables.

For the analysis of the summation tests, repeated measures ANOVAs were employed, followed up by pairwise comparisons to examine the differences between the test cues CI, CN, CK, and C (for C the last trial of acquisition was used in the predictive summation test only). Bonferroni corrections were applied to these pairwise comparisons. Non-parametric tests were used for the binary data from the predictive response summation test (Friedman's ANOVA followed by Wilcoxon matched pairs) and parametric tests were used for the continuous data from the evaluative summation test (parametric ANOVA and Student's t-tests). These tests aimed to assess the reduction in responding to C on compounding with cues: I − putative conditioned inhibitor, K – neutral familiar control, and N – novel cue. It was expected that the conditioned inhibitor would reduce responding more than the control cues K and N.

Following the overall analysis of the conditioned inhibition tests, participants were classed as either inhibitors or non-inhibitors on the basis of their performance in the summation tests (Glautier & Brudan, 2019). The purpose was to look for any difference between measures of non-associative inhibition in those individuals showing clear and unambiguous conditioned inhibition and those who did not. The FN discrimination analysis was revisited by looking at the regressions of the FN coefficients on non-associative measures of inhibition and adding the new inhibition classification and interaction between the classification and the non-associative measures of inhibitions to the regression model.

For the analysis of the predictive summation test, participants were classified as inhibitors or occasion setters using their responses to CN and CI as follows. Participants were classified as inhibitors if cue I reduced responding to C more than cue N (CI − CN), otherwise they were assumed to be occasion setters since they had successfully solved the FN discriminations but the inhibitory properties of I did not transfer to the CI compound. The data from the second summation test was analysed following the same steps with the only difference being that a continuous score of conditioned inhibition was computed for every participant using their reported probabilities of tummy ache/no tummy ache as opposed to the previously used categorisation into inhibitors and occasion setters. The conditioned inhibition score was computed as the difference between CI and CN (CI – CN). High conditioned inhibition scores would suggest that the participant had learnt conditioned inhibition while low scores would suggest that the participant had learnt occasion setting. Similar to the first summation test, the conditioned inhibition score was used as a dependent variable in a multiple regression with the non-associative measures of inhibition (BIS11, BIS, BAS, DD, and SSRT) as independent variables. Then the FN multiple regressions were revisited, and the measure of inhibition along with interaction between this inhibition score and the other non-associative measures were included.

2.3.2 Results

2.3.2.1 Non-associative Measures of Inhibition

The means and standard deviations of the 99 participants who passed all the inclusion criteria on the non-associative measures of inhibition are provided in Table 15. All mean values were within the ranges of the heathy general population reported across the literature (Jorm et al., 1998; Klein et al., 2022; Lipszyc & Schachar, 2010; Stanford et al., 2009).

Table 15

Descriptive Statistics for Non-Associative Inhibition

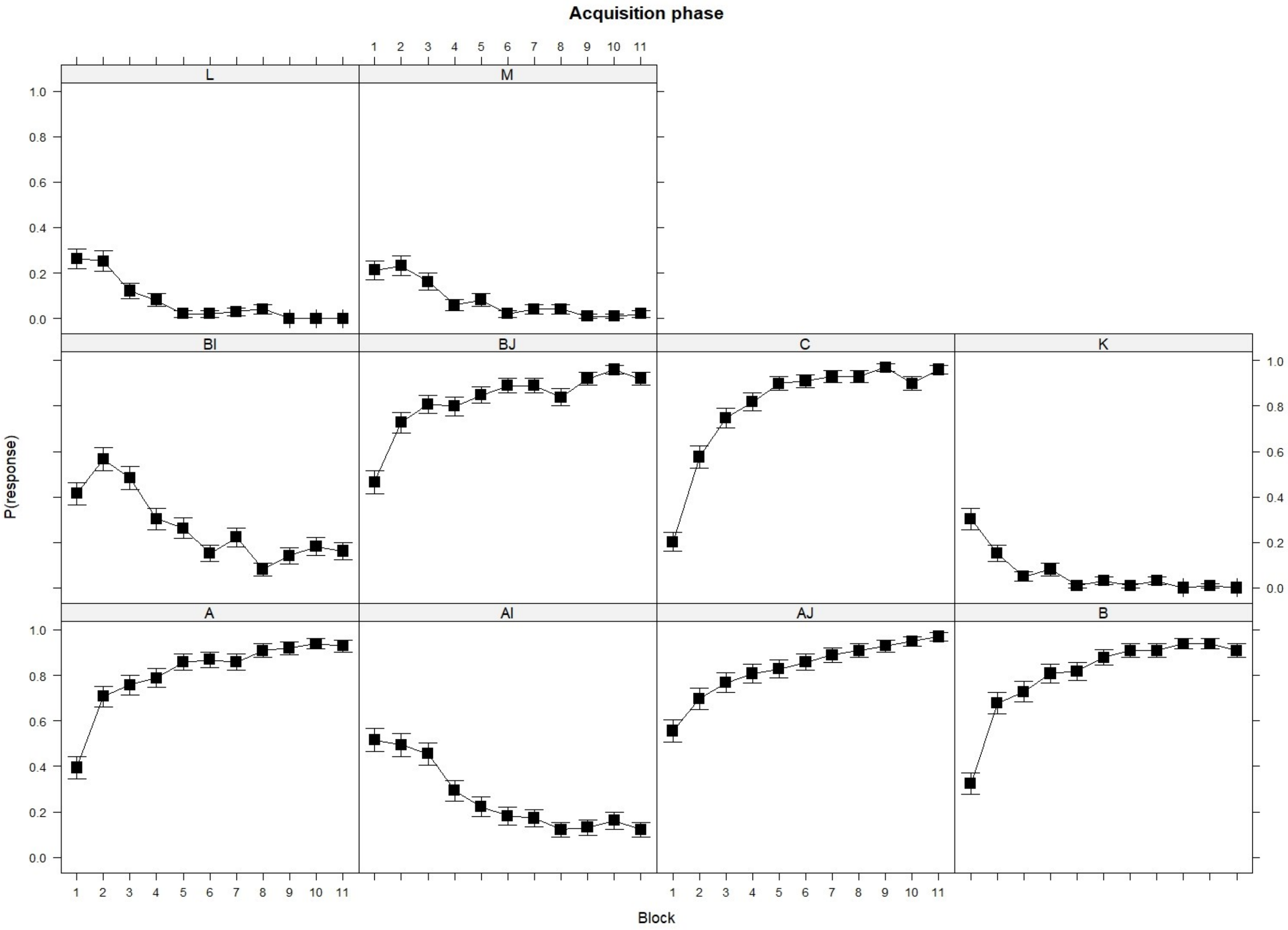
	Mean	Standard Deviation
BIS11	59.65	9.66
BIS	21.76	3.83
BAS	38.72	5.84
k	0.06	0.32
SSRT	240.41	52.10

2.3.2.2 Acquisition

The acquisition stage performance of the 99 participants who passed the learning criterion is shown in **Figure 8** , which indicates that these participants learned to respond more to the reinforced, than to the non-reinforced cues over the course of the acquisition blocks.

Figure 8

Acquisition Phase of the Learning Task.



2.3.2.2.1 Feature Negative Discrimination

An initial generalised linear mixed model with discrimination and block (linear and quadratic) as fixed factors and participants as random factors was used to assess whether the two feature negative discriminations differed. The model revealed that the fixed effects of both the linear and the quadratic terms for block were significant, however the main effect of discrimination was not significant (Table 16). The interactions between discrimination and block (linear), and between discrimination and block (quadratic) were also not significant (Table 16). Accordingly, the initial model showed that overall participants did not perform differently on the A+/AI- and B+/BI- feature negative discriminations during the acquisition phase.

As a result, for the final generalised linear mixed model the discrimination factor and the linear slope were removed, and individual intercepts and quadratic slopes were fitted for each participant (the linear slope was excluded to allow for the interpretation of the intercepts produced by the model and due to the fact that learning rates were expected to be quadratic in nature rather than linear). The model revealed that the quadratic effects of block was still significant (Table 16). The individual slopes and intercepts from the model were then used as measures of performance to assess the role of the non-associative measures of inhibition on the FN discrimination learning.

Table 16

Feature Negative Discrimination Learning over Time by Cue

Model	Fixed Effect	Estimate	SE	z	p
Cue * Block	Intercept	0.66	0.13	5.06	< .001
	Cue	−0.03	0.11	−0.26	.80
	Block (linear)	−54.45	2.91	−18.70	< .001
	Block (quadratic)	−21.22	2.71	−7.84	< .001
	Cue * Block (linear)	−1.29	5.32	−0.24	.81
	Cue * Block (quadratic)	−3.77	5.32	−0.71	.48
Block	Intercept	2.24	0.21	10.63	< .001
	Block (quadratic)	−0.04	0.002	−14.77	< .001

2.3.2.2.2 Non-associative Inhibition.

Two multiple regressions were computed using these slopes and intercepts extracted for each participant as the DVs and the non-associative measures of inhibition as IVs. No significant effect of the non-associative inhibition on the FN discrimination learning were found, meaning that the participants' learning performance was not associated with their performance on the non-associative inhibition tasks/questionnaires (Table 17).

Table 17

The Effect of Non-associative Inhibition on Feature Negative Discrimination Learning

DV	R^2	dfs	F	p
FN Intercept	.05	5, 93	0.90	.49

Non-associative Inhibition	Unstandardized β	t	p
Intercept	0.83	43.22	< .001
BIS11	-0.03	-1.40	.16
BAS	0.006	0.30	.76
BIS	0.02	0.92	.36
DD	-0.01	-0.70	.49
SSRT	0.02	1.18	.24

DV	R^2	dfs	F	p
FN Slope Quadratic	.01	5, 93	0.16	.98

Non-associative Inhibition	Unstandardized β	t	p
Intercept	-0.04	-31.85	< .001
BIS11	0.001	0.65	.52
BAS	-0.0003	-0.26	.80
BIS	-0.001	-0.42	.68
DD	-0.0001	-0.11	.92
SSRT	-0.001	-0.45	.65

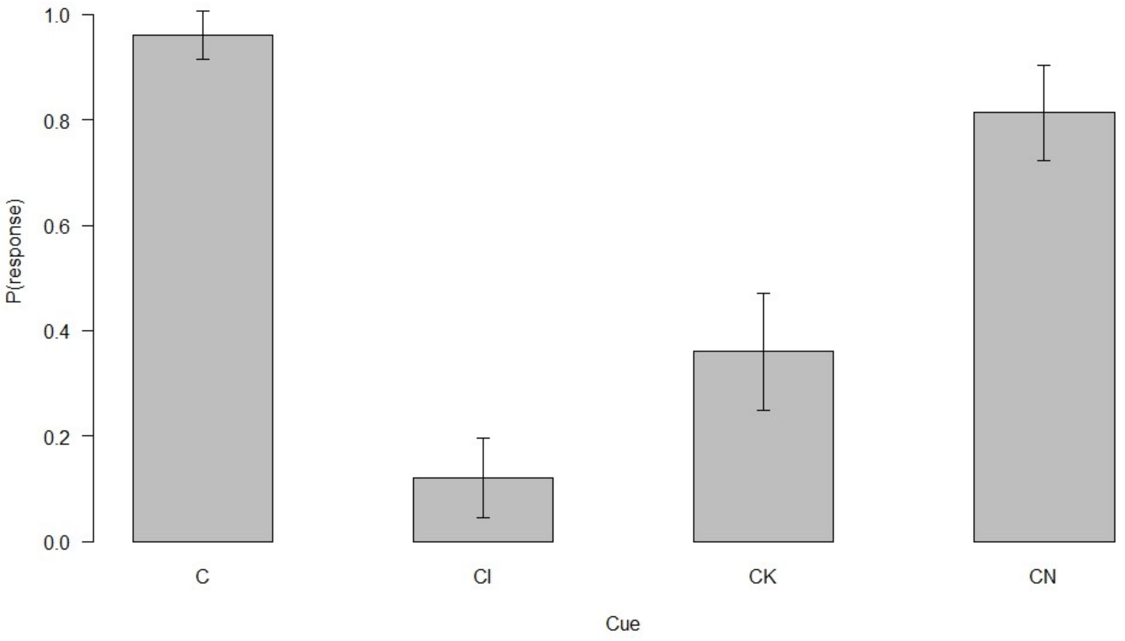
2.3.2.3 Summation Test

2.3.2.3.1 Predictive Summation Test

The predictive summation test performance of the 75 participants who passed the second exclusion criterion and solved the FN discrimination by the end of the acquisition is presented in Figure 9. Figure 9 shows that responding to CI was markedly suppressed compared to responding to C at the end of the acquisition and compared to control compounds CK and CN.

Figure 9

Predictive Summation Test Performance



A Friedman's ANOVA showed that there were significant differences in responding to the four cues ($X^2(3) = 124.08$, $p < .001$). Follow-up Wilcoxon matched paired tests with a Bonferroni correction showed that CI elicited reduced responding compared to C, CK, and CN ($Z = -9.94$, $p < .001$, $r = - .92$, $Z = -3.40$, $p = .003$, $r = - .39$, and $Z = -6.83$, $p < .001$, $r = -$ .79, respectively). Similarly, responses to cues CK and CN were significantly reduced compared to cue C ($Z = -6.56$, $p < .001$, $r = - .76$ and $Z = -2.67$, $p = .03$, $r = - .31$, respectively). Finally, responses to the two control test cues were also significantly different, participants responded less to CK than to CN ($Z = -5.67$, $p < .001$, $r = - .65$).

2.3.2.3.1.1 Non-associative Inhibition

Participants were classified as inhibitors and non-inhibitors based on their responses to the predictive summation test. Participants were classified as inhibitors if cue I reduced responding to C (C-CI) more than cue N (C-CN), there were 55 inhibitors and 20 non-inhibitors. A logistic multiple regression was used to assess whether the non-associative inhibition scores had an effect on the how participants were classified into inhibitors and occasion setters. None of the effects were significant (Table 18). The regression was repeated for the classification based on CK, instead of CN, this also produced no significant effects.

Table 18

Effects of Non-associative Inhibition on the Predictive Summation Test Performance

Model	Cox &Snell R^2	McFadden R^2	dfs	X^2	p
	.04	.03	1, 69	2.75	.74

Non-associative Inhibition	Estimate	Wald Statistic	p
Intercept	1.05	14.56	<.001
BIS11	.15	0.28	.60
BAS	.27	0.94	.33
BIS	.21	0.67	.41
DD	-.03	0.01	.91
SSRT	-.15	0.23	.63

2.3.2.3.1.2 Feature negative discrimination and non-associative inhibition revisited.

Two multiple regressions with slopes and intercepts as DVs were computed again with the addition of the inhibition group (inhibitor versus non-inhibitor) and the interactions between inhibition group and the non-associative measures of inhibition as IVs. The models revealed a significant effect of inhibition grouping on both the intercepts and slopes of the participants (Figure 10, Table 19). Inhibitors had a better performance on the FN training at the end of training and learnt the FN discrimination faster than the non-inhibitors (Table 19). The regressions were repeated for the classification based on CK but this produced no significant effects.

Table 19

Effects of Non-associative Inhibition on Predictive Summation Test Performance

DV	R^2	dfs	F	p
FN Intercept	.13	11, 63	0.88	.57

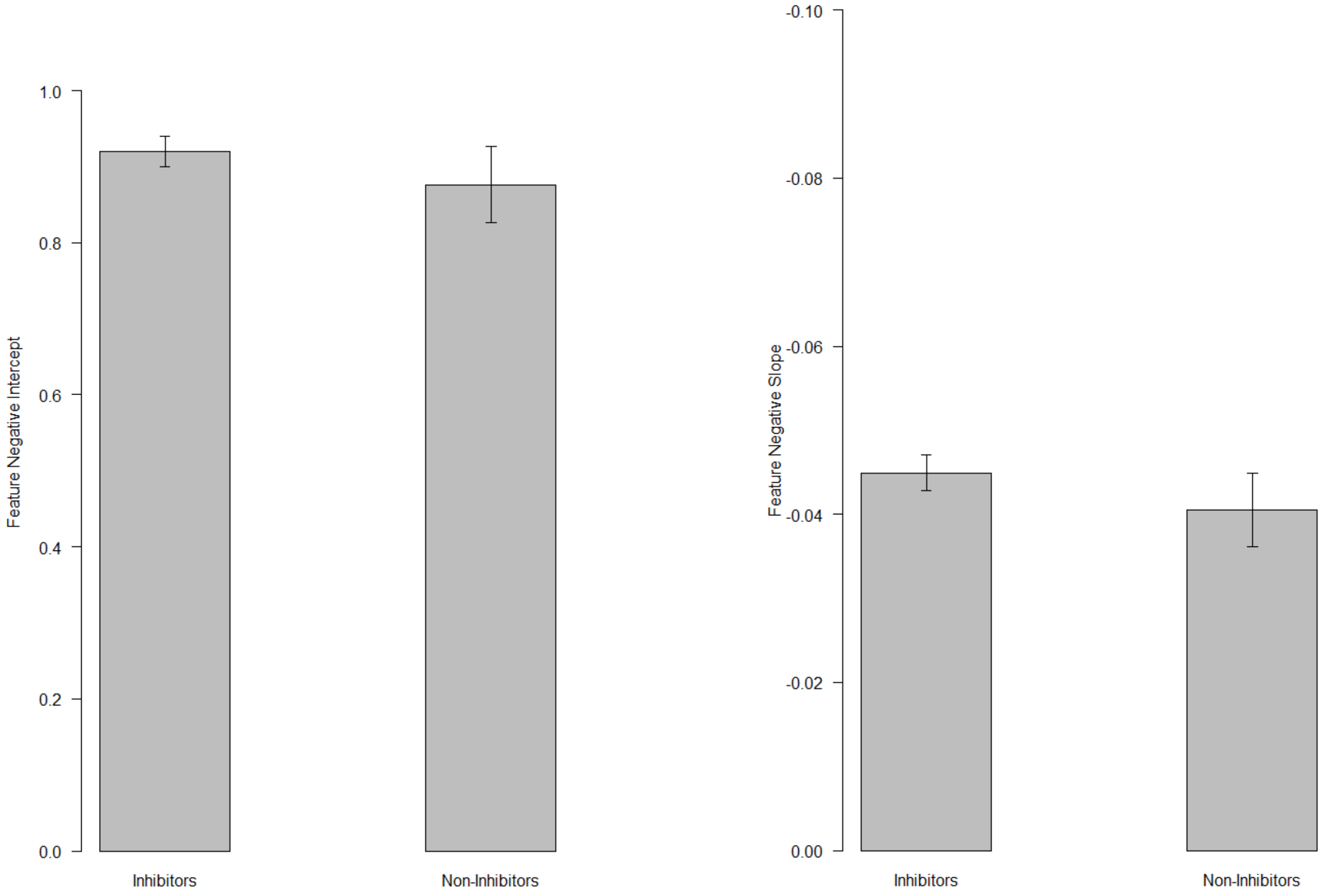
Non-associative Inhibition	Unstandardized β	t	p
Intercept	0.87	40.60	< .001
BIS11	-0.03	-0.90	0.37
BAS	-0.005	-0.23	0.82
BIS	-0.01	-0.76	0.45
DD	-0.01	-0.66	0.51
SSRT	0.001	0.02	0.98
Inhibition	0.06	2.37	0.02*
BIS11*Inhibition	0.04	1.14	0.26
BAS*Inhibition	-0.01	-0.57	0.57
BIS*Inhibition	0.003	0.11	0.91
DD*Inhibition	0.01	0.44	0.66
SSRT*Inhibition	0.003	0.10	0.92

DV	R^2	dfs	F	p
FN Slope Quadratic	.13	11, 63	0.89	.55

Non-associative Inhibition	Unstandardized β	t	p
Intercept	-0.04	-19.25	< .001
BIS11	0.001	0.49	0.62
BAS	-0.0004	-0.22	0.83
BIS	0.001	0.69	0.49
DD	0.56	0.57	0.57
SSRT	0.001	0.53	0.60
Inhibition	-0.01	-2.16	0.03*
BIS11*Inhibition	-0.002	-0.79	0.43
BAS*Inhibition	0.002	0.82	0.42
BIS*Inhibition	0.0001	0.03	0.97
DD*Inhibition	-0.002	-1.07	0.29
SSRT*Inhibition	-0.001	-0.48	0.63

Figure 10

Effect of the Inhibition Classification on Feature Negative Learning (Intercept on the left panel and Slope on the right panel)

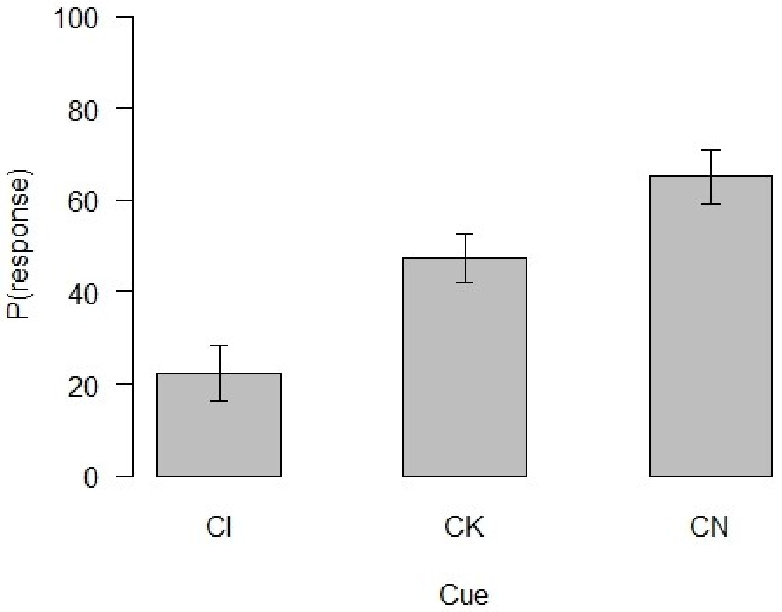


2.3.2.3.2 Evaluative Summation Test

The evaluative summation test performance of the 75 participants who passed the second exclusion criterion and solved the FN discrimination by the end of the acquisition is shown in Figure 11.

Figure 11

Evaluative Summation Test Performance



A repeated measures ANOVA was used to assess the differences between participants' evaluations of the test cues and C in the last trial of acquisition. The ANOVA revealed a significant effect of cue $F(2,148) = 63.19$, $p < .001$, $\omega^2 = .33$. All possible comparisons were then computed using paired samples t-tests with a Bonferroni correction. The t-tests revealed that participants rated the likelihood of CI ($M = 22.37$, $SD = 26.64$) to be reinforced significantly lower than both CK ($M = 47.53$, $SD = 22.89$) $t(74) = -7.35$, $p < .001$, $d = -1.70$ and CN ($M = 65.15$, $SD = 25.62$) $t(74) = -9.61$, $p < .001$, $d = -1.00$. These differences confirm the results of the predictive summation test and show that there was an overall effect of conditioned inhibition. The two control compounds, CK and CN, were also rated statistically differently $t(74) = -5.01$, $p < .001$, $d = -0.70$, CK was rated lower than CN. In summary, all comparisons were significant with compound CI rated the lowest in terms of likelihood of reinforcement, followed by CK and CN in that order (Figure 11).

2.3.2.3.2.1 Feature Negative Discrimination

An inhibition score was computed for all participants using their evaluative summation test performance (CN − CI). On this scale high scores represent higher levels of inhibition while low scores represent low inhibition. Similarly to the predictive summation test, a linear multiple regression was used to assess whether the non-associative inhibition had an effect on the participants' inhibition scores. There was a significant effect of BIS on the inhibition scores, participants who have scored high on BIS showed more inhibition on the evaluative summation test (Figure 12, Table 20). None of the other effects were significant. The regression was repeated for the classification based on CK and none of the effects were significant.

Table 20

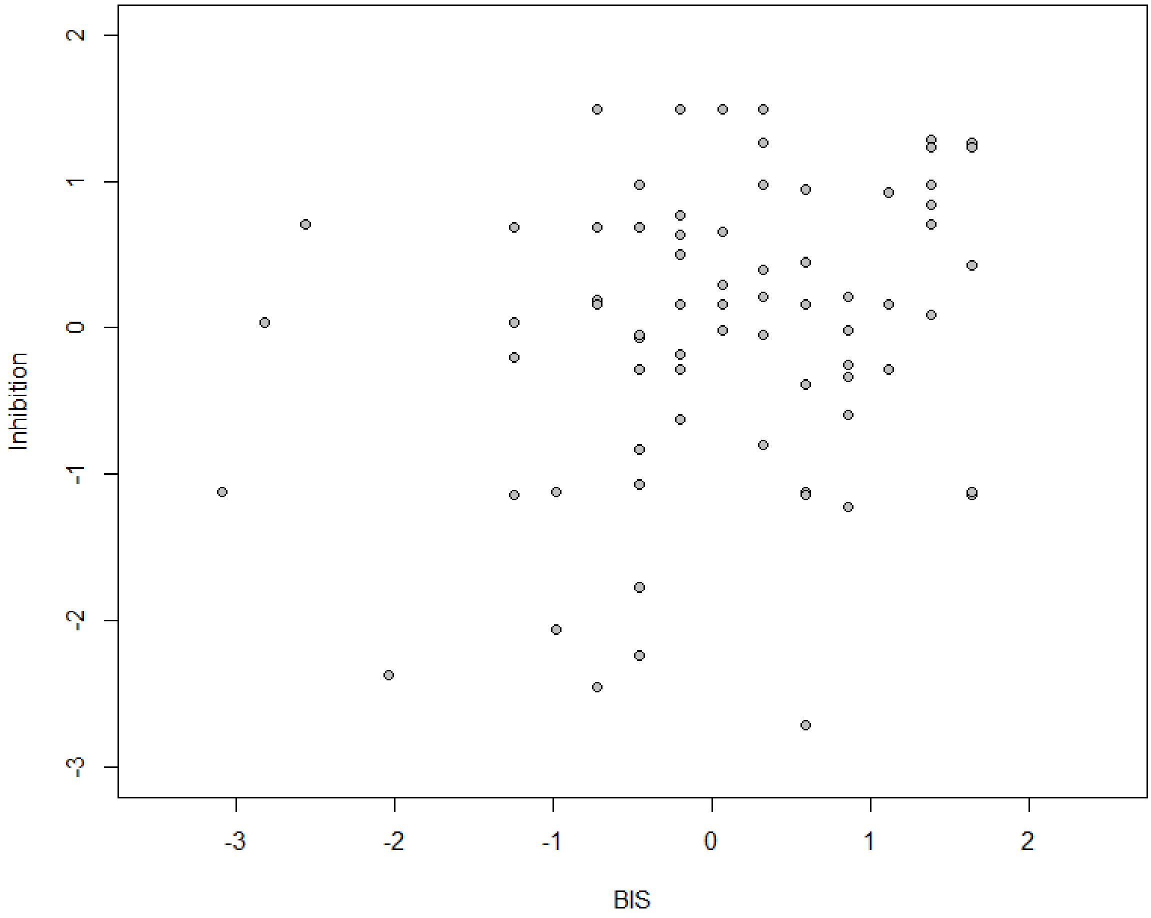
Effects of Non-associative Inhibition on the Evaluative Summation Test Performance

R^2	dfs	F	P
.10	5, 69	1.49	.21

Non-associative Inhibition	Unstandardized β	t	p
Intercept	-0.03	-0.23	.82
BIS11	-0.12	-1.04	.30
BAS	-0.12	-1.04	.30
BIS	0.27	2.28	.03*
DD	-0.02	-0.13	.90
SSRT	-0.09	-0.65	.52

Figure 12

The Effect of BIS on the Evaluative Summation Test Performance (Inhibition Score)



2.3.2.3.2.2 Feature negative discrimination and non-associative inhibition revisited

Following the methodology set out in the predictive summation test, two multiple regression with slopes and intercepts as DVs were computed again with inhibition group as a factor along with interactions between inhibition group and the non-associative inhibition scores as IVs. None of the effects were significant (Table 21). The regression was repeated for the classification based on CK and it revealed a significant effect of inhibition score on both the learning intercepts and slopes (Figure 13, Table 22). Higher inhibition scores were indicative of larger intercepts and slopes, meaning that participants who showed more inhibition were more likely to have learnt the FN discrimination by the end of the acquisition, and this learning has occurred faster. The effect of inhibition score and SSRT interaction on the intercept was also significant, according to this effect participants who had low SSRT scores (meaning they were fast in stopping their responses) and showed less inhibition performed worse in the FN discrimination at the end of training compared to fast participants who showed more inhibition. On the other hand participants who had high SSRT scores (meaning they were slow in stopping in their responses) showed the same level of performance in the FN discrimination regardless of how much inhibition they showed (Figure 14). None of the other effects were significant (Table 22).

Table 21

Effects of Non-associative Inhibition (Evaluative Summation Test) CN on the Predictive Summation Test Performance

DV	R^2	dfs	F	p
FN Intercept	.12	11, 63	0.79	.65

Non-associative Inhibition	Unstandardized β	t	p
Intercept	0.91	85.52	< .001
BIS11	0.01	0.82	0.42
BAS	-0.01	-1.20	0.23
BIS	-0.01	-1.12	0.27
DD	-0.01	-0.94	0.35
SSRT	0.01	1.07	0.29
Inhibition	0.02	1.46	0.15
BIS11*Inhibition	0.003	0.31	0.76
BAS*Inhibition	0.01	0.51	0.61
BIS*Inhibition	-0.01	-0.45	0.66
DD*Inhibition	0.01	0.71	0.48
SSRT*Inhibition	-0.03	-1.81	0.08

DV	R^2	dfs	F	p
FN Slope Quadratic	.08	11, 63	0.51	.89

Non-associative Inhibition	Unstandardized β	t	p
Intercept	-0.04	-40.29	< .001
BIS11	-0.001	-1.10	0.28
BAS	0.001	0.69	0.50
BIS	0.001	1.10	0.28
DD	-0.0003	-0.24	0.81
SSRT	-0.001	-0.40	0.69
Inhibition	-0.001	-1.32	0.19
BIS11*Inhibition	0.00004	0.04	0.97
BAS*Inhibition	0.0001	0.04	0.97
BIS*Inhibition	0.0002	0.19	0.85
DD*Inhibition	-0.001	-1.11	0.27
SSRT*Inhibition	0.002	0.99	0.32

Table 22

Effects of Non-associative Inhibition (Evaluative Summation Test) CK on the Predictive Summation Test Performance

DV	R^2	dfs	F	p
FN Intercept	.20	11, 63	1.39	.20

Non-associative Inhibition	Unstandardized β	t	p
Intercept	0.91	89.75	< .001
BIS11	0.01	0.55	0.59
BAS	-0.01	-0.99	0.33
BIS	-0.01	-0.90	0.37
DD	-0.01	-0.81	0.42
SSRT	0.004	0.31	0.76
Inhibition	0.02	2.10	0.04*
BIS11*Inhibition	0.01	0.93	0.35
BAS*Inhibition	0.001	0.13	0.90
BIS*Inhibition	-0.001	-0.05	0.96
DD*Inhibition	-0.001	-0.77	0.94
SSRT*Inhibition	-0.03	-2.46	0.02*

DV	R^2	dfs	F	p
FN Slope Quadratic	.19	11, 63	1.30	.24

Non-associative Inhibition	Unstandardized β	t	p
Intercept	-0.04	-43.62	< .001
BIS11	-0.001	-0.77	0.44
BAS	0.0005	0.47	0.64
BIS	0.001	0.91	0.36
DD	-0.0004	-0.31	0.76
SSRT	0.0005	0.42	0.68
Inhibition	-0.02	-2.07	0.04*
BIS11*Inhibition	0.0002	0.16	0.87
BAS*Inhibition	-0.001	-1.14	0.26
BIS*Inhibition	-0.001	-0.07	0.55
DD*Inhibition	0.0001	0.07	0.95
SSRT*Inhibition	0.002	1.90	0.06

Figure 13

The Effect of Inhibition Classification on Feature Negative Discrimination Learning (Intercept on the top panel and slope on the bottom)

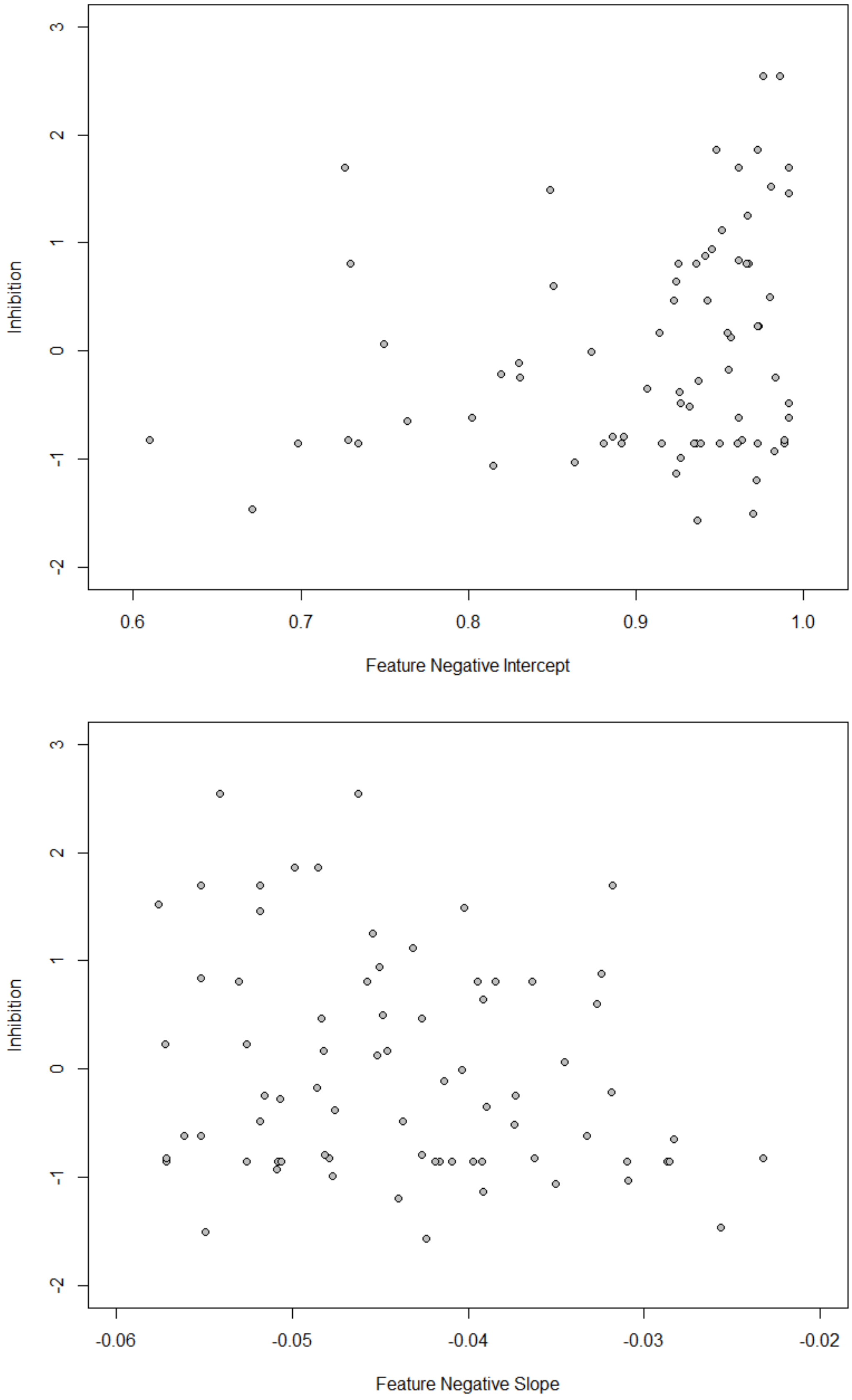
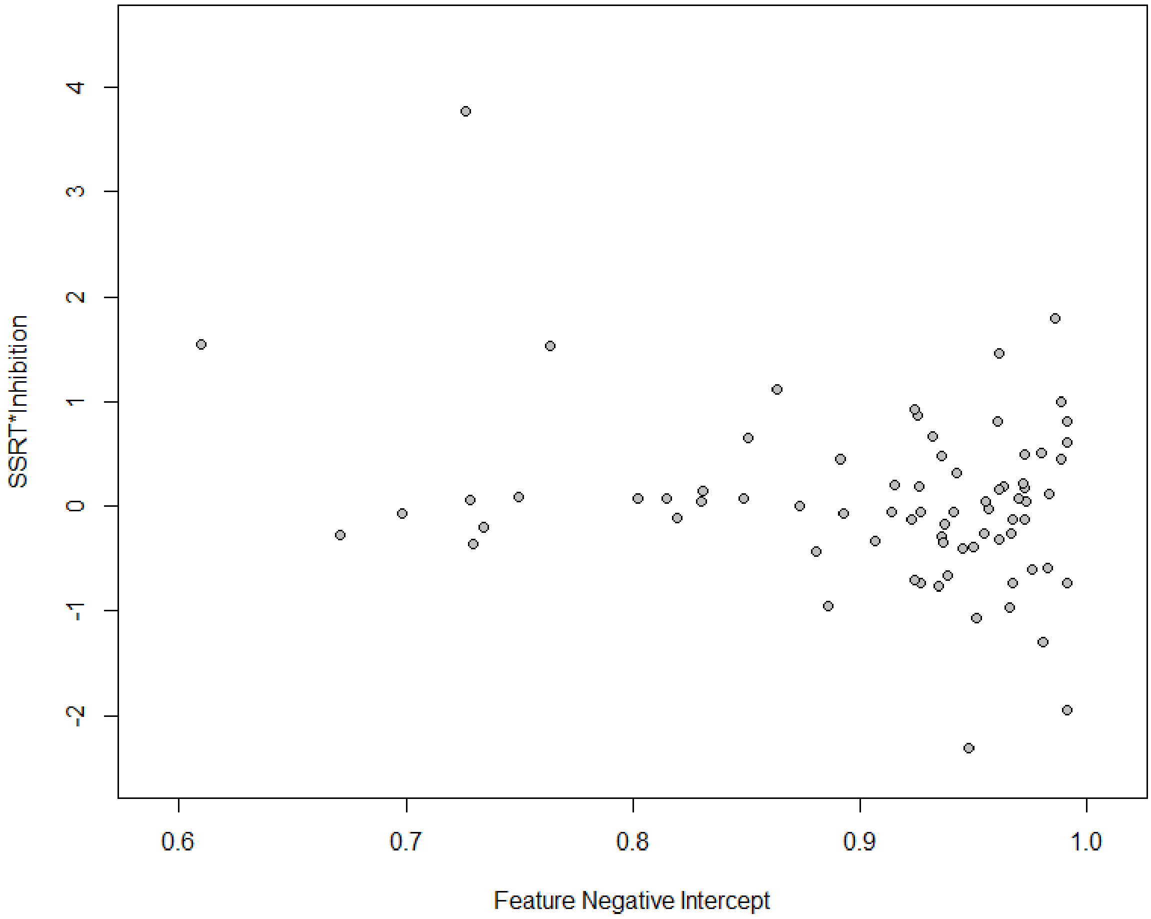


Figure 14

Effect of SSRT and Inhibition Interaction on Feature Negative Discrimination Learning (Intercept)



2.3.3 Discussion

The current experiment used a feature negative discrimination task to produce associative inhibition along with four measures of non-associative inhibition to assess whether a common underlying inhibitory mechanism exists to link these two domains of inhibition. The results showed that participants' performance in the feature negative discrimination task was not significantly related to any of our non-associative measures of inhibition (BIS/BAS, BIS11, Delay discounting, and SSRT, Table 17) regardless of whether or not the participants were classed as inhibitors or as non-inhibitors (Table 18) and regardless of whether or not the summation test used to classify participants was based on predictive or evaluative responses (Table 18 versus Table 20). The only exception was the relationship between BIS and associative inhibition shown to be statistically significant in the evaluative summation test, however this was not replicated with the predictive summation test. High BIS scores are indicative of a strong inhibitory system, therefore this positive correlation confirmed the *a priori* expectation of participants with high scores on BIS to show strong inhibition. Another link between non-associative inhibition and FN performance was seen between SSRT and the intercept of the FN discrimination but this was not replicated across the two cues used to assess conditioned inhibition or across the two summation tests (CI v CN and CI v CK, Table 20 and Table 21). It was however found that inhibitors did tend to perform better in the FN discrimination than the non-inhibitors (Table 19, Table 21, and Table 22) but these effects were not strong nor were they consistent across the two cues used in the summation tests. Nevertheless, the fact that effects showed in both the predictive (Table 19) and the evaluative summation tests (Table 22) gives a degree of confidence in this finding.

Since learning criteria were applied to select only those participants who learned task prerequisites sufficiently well so that their performance could not be easily explained at the level of chance there is confidence that performance in the FN discrimination and in the summation tests was actually indicative of the strength of inhibitory learning rather than reflective of a learning deficit. In conclusion, although associative and non-associative inhibition both involve some form of inhibitory process, they are likely to be independent subtyped of inhibition.

Various models of inhibition/impulsivity can be found in the literature, however associative inhibition is not often considered in attempting to map these concepts. For example Caswell et al. (2015) assessed the relationship between 10 behavioural and one self-reported measure of impulsivity. An exploratory factor analysis was used on the data to assess the relationship between all the measures and to try and understand the factors that comprise impulsivity/inhibition. As a result, a four factor model of impulsivity was proposed. The factors were: motor-impulsivity (action cancellation), reflection-impulsivity, action restraint, and temporal-impulsivity. Caswell et al. (2015) aimed to show the multidimensionality of impulsivity and the four factors proposed along with the fact that four of the measures used did not load on any of the factors supported this view. Furthermore, associative inhibition was not included in this study meaning that at least a few more factors were missing.

Bari and Robbins (2013) proposed a structure for the concept of inhibition which partitions inhibition into cognitive and behavioural inhibition based on the growing number of studies showing a lack of correlation between self-reported and behavioural measures of inhibition (Broos et al., 2012; Enticott et al., 2006; Reynolds et al., 2006). Behavioural inhibition was further subdivided into response inhibition, deferred gratification, and reversal learning. Associative inhibition was included in this proposed underlying structure of inhibition in the form of reversal learning which assesses a participant's cognitive flexibility and ability to adapt to changes. Although the fact that inhibition/impulsivity are multifaceted concepts is generally accepted, a clear classification of the underlying structure has not been agreed upon. The current thesis argues that associative inhibition should be considered as a facet of inhibition in future development of inhibition/impulsivity models.

In isolation associative and non-associative inhibition have been historically shown to have important clinical implications, being associated with disorders such as ADHD, substance abuse, and schizophrenia (Bauer, 2001; Enticott et al., 2008; Fillmore & Rush, 2006; Fillmore & Rush, 2002; Hoptman et al., 2002; Porter et al., 2011; Schachar et al., 1993). However the two types of inhibition have been rarely used together, one of the few examples is the study by He et al. (2011) that assessed the relationship between conditioned inhibition and personality disorder using a group of participants with a history of violent offences and a control group from the general population. The group first group was further divided into participants with personality disorder and participants with dangerous and severe personality disorder. Although not explicitly measured, impulsivity was assumed to be high in the group of participants with personality disorders and a history of violent offences. The results showed that this group also performed worse in the conditioned inhibition summation test, showing an impaired ability to develop conditioned inhibition, although no differences were observed during acquisition. Furthermore, this was more accentuated in the group that met the criteria for dangerous and severe personality disorder. The results of He et al. (2011) along with the evidence suggesting that associative and non-associative inhibition are independent processes highlight the importance of including both types of inhibition when studying disorders such as personality disorders.

Chapter 3 Extinction and Non-associative Inhibition

Once learning has taken place it can be changed through various routes if environmental changes lead to that learning becoming outdated or even maladaptive. The previous chapter focused on conditioned inhibition as a route for changing a learnt association, the current chapter focuses on a second route: extinction. A simple extinction procedure consists of presenting a previously trained conditioned stimulus (CS) without the unconditioned stimulus (US) it was previously associated with. As a result of these non-reinforced presentations of the CS, the conditioned response (CR) generated by the CS diminishes. However, phenomena such as recovery and renewal show that extinction involves more than simply unlearning a previously learnt association (e.g. Bouton, 1993, 1994, 2000).

Recovery refers to the re-emergence of an extinguished CR when the CS is re-presented following a delay after extinction. On the other hand, renewal refers to re-emergence of an extinguished CR when the CS is presented in a context which differs from the context in which extinction took place. Both phenomena can potentially be explained through mechanisms involving contextual stimuli, with recovery being treated as a special case of renewal in which the passage of time implicitly modifies the context. In the case of renewal contextual changes are explicit e.g. when the environment changes after extinction. Contextual stimuli are those that remain constant across the course of multiple learning trials and can be contrasted with the punctate CSs and USs that mark the learning trials. One approach to explaining renewal is through conditioned inhibition and "protection-from-extinction". According to this, based upon the Rescorla-Wagner associative model (Rescorla & Wagner, 1972), when the context changes during extinction the new context behaves as a CS and acquires inhibitory properties. Therefore, post-extinction, when the CS is presented outside this context a renewal effect may occur because the inhibitory influence of the extinction context is no longer present. Because the context acquired inhibitory strengths during extinction it is assumed that it could protect the target cue from extinction. Some experiments have shown that extinction carried out in the presence of a discrete inhibitory stimulus can protect from extinction (e.g. Rescorla, 2003) but the evidence for contextual stimuli functioning in that way is mixed (e.g. Bouton & Swartzentruber, 1986; Glautier et al., 2013; Polack et al., 2012).

Another associative model, developed by Pearce (1994), can also explain renewal but the mechanism differs from that proposed in the Rescorla-Wagner model. The Rescorla-Wagner model is an elemental model in that it treats the CSs involved in conditioning as discrete elements, each of which may enter into associations with USs. In contrast, Pearce's model is a configural model in which the discrete stimulus elements encountered on each learning trial form "configurations" and the configurations themselves are the candidates for forming associations with USs. Renewal in the Pearce configural model is determined by the similarity relations between the stimulus configuration used in the post-extinction test and the other stimulus configurations previously encountered during acquisition and extinction. Renewal occurs if the net of generalised excitatory and inhibitory influences produced by the post-extinction test configuration is greater than zero.

It is of practical and theoretical interest to get a better understanding of the mechanisms underlying extinction. On the practical side there are therapeutic interventions based on extinction which could be improved. In the case of addiction it has long been accepted that relapse is a major problem with typically less than 50% "survivors" three months after initiating abstinence and this applies across a range of substances and even in individuals receiving clinical interventions (e.g. Anton et al., 2006; Fortmann & Killen, 1995; Northrup et al., 2015). Cue-exposure for addiction is based on an underlying model of addiction in which drug-related stimuli – drug-cues become CSs because they are repeatedly paired with drug USs. The CRs produced by drug-cues are thought to play a part in relapse and cue-exposure treatment aims to reduce relapse risk by extinguishing CRs to drug-cues by repeated presentation of the cues without a drug US. Unfortunately, although cue-exposure is effective for treatment of some conditions (e.g. phobias c.f. Choy et al., 2007), its effectiveness in the treatment of addiction is not well established, but a small number of studies suggest it is an intervention worthy of further investigation (Kiyak et al., 2023).

Renewal effects may be one factor that limits the effectiveness of cue-exposure treatments (e.g. Bouton, 2000; Conklin & Tiffany, 2002) and some experiments have provided evidence that carrying out extinction in multiple-contexts may reduce renewal effects (Bustamante et al., 2016; Glautier et al., 2013). An alternative approach, which is the focus of the current series of experiments, is to carry out extinction in the presence of multiple excitatory cues (Craske et al., 2014). The objective of carrying out extinction in the presence of multiple excitatory cues is to increase the amount of associative change that occurs during extinction. According to associative models, such as the Rescorla-Wagner and the Pearce configural models, associative change is driven by prediction error. An error signal is generated during extinction because a cue that has previously signalled an outcome is presented in the absence of that outcome. It follows from these associative models that if the prediction error can be increased during extinction then the amount of learning during extinction will be correspondingly increased. One way to increase prediction error, instead of presenting single cues on each extinction trial, is to present compounds of multiple excitatory cues on each trial during extinction. To explain this further, the Rescorla-Wagner and the Pearce configural models both make use of an error term the form of which is given in Equation (6).

$$\lambda - \sum V \tag{6}$$

In Equation (6) the value of λ is used to indicate the status of the US on each learning trial. When λ is set =1 there is a US, as in acquisition, and when λ is set =0 there is no US, as in extinction. The subtrahend, $\sum V$, represents the summed associative strength of all cues present on that trial. So, in a simple case for the Rescorla-Wagner model, assuming cue A has been trained to asymptote during an acquisition phase then $V_A \rightarrow 1$ and then the extinction of cue A would begin. On the first extinction trial $\sum V = V_A$ since cue A is the only cue present and the error on this first extinction trial would therefore approach −1 and this value determines the amount of associative change for cue A. If, during acquisition, cues A and B had both been trained to asymptote there would be the option of presenting an AB compound for extinction. In this case, on the first extinction trial, the error term would approach −2 ($\sum V = V_A + V_B$) and therefore theoretically more extinction would be expected to occur for target cue A than if only A had been presented for extinction. But this is not a universal theoretical prediction. According to the Pearce configural model, presenting an AB compound for extinction in this simple procedure would not increase prediction error. This is because $\sum V$ in Pearce's configural model is determined as a weighted sum of the associative strengths of all configurations known to the system, with the weights being formed by the similarities between the configuration actually present (AB in this case) and all configurations in the system (A, B, and AB in this case). Assuming the similarity between each of the elements and the AB compound is ½ (Pearce, 1994) and since $V_{AB} = 0$ it would mean that $\lambda - \sum V \rightarrow -1$ which is the same as if A was presented alone for extinction. Furthermore, since the associative change would occur to configuration AB the impact would only be on responding to the target cue A via generalisation, the associative strength of configuration A itself would remain unaffected.

In fact there have been numerous demonstrations which have shown that increased prediction error during extinction can result in more extinction (Rescorla, 2000; Rescorla, 2006). In Rescorla (2000) rats were trained with two cues, A and X, as signals for food (A+ and X+ trials) and with a third cue B which was non-reinforced (B- trials). The animals were then divided into four groups, with one group receiving extinction trials with an AX compound stimulus (AX− trials), and the other groups receiving extinction trials with X alone

(X− trials), a BX compound (BX− trials), or no extinction trials at all. In a test presentation of X group AX- showed least responding of all indicating that the AX- extinction trials had resulted most complete extinction. In the current thesis this is referred to as "super-extinction" (as used in Hermans et al., 2006; Jacoby & Abramowitz, 2016) after its mirror analogue with "super-conditioning" (Williams & McDevitt, 2002), and is distinguished from a related procedure "deepened-extinction". In super-conditioning acquisition of associative strength for a target cue is enhanced by reinforcement of that target in compound with an inhibitory cue whereas in super-extinction extinction of associative strength for a target cue is enhanced by non-reinforcement of that target in compound with an excitatory cue.

The current series of studies had two objectives. First, it was aimed to further examine the extent to which extinction is impacted, in human participants, by compound extinction and whether or not there is an association between the inhibition developed during extinction and non-associative inhibition. Second, it was sought to examine which of three related associative models, each based on error correction, would provide the best account of participant behaviour during our extinction procedures (the second aim is addressed in the next chapter). The series consists of two main experiments and one pilot study in-between. In the first study cue alone extinction was compared to super-extinction, however certain design feature could have prevented effects such as summation from being observed. As a result a pilot study[2] was used to test a new design for the learning task. The final study compared cue alone extinction, super-extinction, and deepened extinction. To allow for differences in extinction acquisition, context inhibition, and recovery to be assessed strict learning criteria

[2] The first study and the pilot study were lab based but the data collection of the pilot study was stopped by the COVID-19 pandemic, therefore the final study was run online. The pilot study for the current series of experiments was the same as the pilot study for the previous series. The chronological order of the studies was: conditioned inhibition study 1, extinction study 1, pilot study (testing a new design for conditioned inhibition and extinction), extinction study 2, conditioned inhibition study 2.

have been apply to ensure that learners have been selected for the analysis, these criteria are described separately for each study below.

3.1 Study 1

The first study of the series compared cue alone extinction with super-extinction using two separate groups. The differences in extinction rate, context inhibition, and recovery between the two groups were assessed. Using the extinction rates and context inhibition scores, the link between associative and non-associative inhibition was re-assessed.

3.1.1 Method

3.1.1.1 Participants

The sample consisted of 59 student participants (43 of the participants identified as female and 16 identified as male, the sample had a mean age of 28.88 years, SD = 4.68) recruited from the Southampton University Highfield Campus. Course credit was awarded for the participation, and the average completion time for the study was 50 minutes.

3.1.1.2 Questionnaires

The same questionnaire based measures used in the previous conditioned inhibition studies were used in the current study to assess non-associative inhibition. A full description of these measures was given under section 2.1.1.2 Questionnaires.

3.1.1.3 Stop Signal Reaction Time Task

The current study used the same stop signal reaction task as the first conditioned inhibition study, the full description can be found in section 2.1.1.3 Stop Signal Reaction Time Task.

3.1.1.4 Learning Task

A custom build learning task has been used to train the extinction of either a cue alone or a compound of two cues. As part of the task participants were asked to observe a series of objects that were "falling" from the top of the screen towards the bottom of the screen. The objects had different shapes and colours which were automatically randomly selected for each participant prior to the start of the experiment. At the bottom of the screen there was a triangular shaped "sensor" which could flash green, red, or not react at all when the objects passed it. The sensor was placed in a "room" and the context was represented by the colour and structure of the walls of the room (Appendix C). Participates were instructed to predict how the sensor would respond to every object while maximizing the number of correct predictions. Participants responded using the "R" key to predict a red flash and the "G" key to predict a green flash. Responses had to be entered once the object reached the response area which was a white rectangular shape on the screen. The task consisted of 67 trials, 48 acquisition trials, eight extinction trials plus eight trials continuing from acquisition intermixed with the extinction trials, two context inhibition test trials and one recovery test trial. Prior to the start of the experiment participants were randomly allocated to one of the two groups: control or super-extinction. The control group received cue alone extinction training, while the super-extinction group was exposed to compound extinction using two cues that were previously reinforced.

3.1.1.5 Design

The design of the learning task is shown in Table 23, and it consisted of four independent stages: acquisition, extinction, context inhibition test (test G), and recovery test (test A). The design followed an ABC procedure which meant that acquisition training took place in context A, extinction training followed in context B, after the extinction training the context inhibition test was carried out in context B followed by a final change in contexts for the recovery test which was carried out in context C. Each cue used in the task was paired with one of the following outcomes: X, Y, and Z. The outcomes X and Y represented reinforcement with either the green or the red flash, these were counterbalanced between participants meaning that for some participants X was representative of the green flash while for other it was representative of the red flash. Z meant no reinforcement, therefore the sensor would not react to the cue. The acquisition stage consisted of eight trials grouped into four blocks, within each block the order of the trials was randomised meaning that the maximum number of identical consecutive trials was three. Cues A and B received reinforced training during acquisition with outcome X while cue C received reinforced training with outcome Y. Cue G received the same training as A and B and was later used to test context inhibition. Cues D and E were non-reinforced to balance the number of reinforced and non-reinforced cues. The two groups: control and super-extinction received identical training throughout the acquisition stage. In the extinction stage the context was changed and the control group received single cue extinction training of cue A while the super-extinction group received compound extinction of cues A and B. There were a total of eight extinction trials groups into four blocks, the extinction being intermixed with the continuous reinforced presentations of cue C. Following extinction training, a context inhibition test was carried out in the extinction context using cue G. Finally the context was changed to a novel third context and the recovery of cue A was tested.

Table 23

Design of Super-Extinction Learning Task for Study 1

	Acquisition A:	Extinction B:	Test G B:	Test A C:
	A → X x8	A → Zx8	G → Z x2	A → Z x1
	B → X x8	C → Y x8		
Control	C → Y x8			
	D → Z x8			
	E → Z x8			
	G → X x8			
	A → X x8	AB → Zx8	G → Z x2	A → Z x1
	B → X x8	C → Y x8		
Super-extinction	C → Y x8			
	D → Z x8			
	E → Z x8			
	G → X x8			

3.1.1.6 Data selection and analysis

Prior to the analysis two exclusion criteria were applied to the data to ensure that the data allowed for the reliable assessment of extinction performance. The first exclusion criterion focused on ensuring that participants learnt to respond to the cue of interest, cue A. This was a critical criterion as the extinction of cue A was the main interest point for the study, therefore in order to study the extinction of the cue it is vital to ensure that the acquisition training was successful and participants have learnt to respond to the cue. As a result, the binomial distribution has been used to determine a cut-off point for participants' responses to cues A, C, D, and E in the last two blocks of the acquisition stage. As there were three possible outcomes, guessing was defined as p(success) = 1/3, there were a total of 12 trials selected for the criterion and the computed cut-off point was eight ($p < .05$). A total of 13 participants were excluded, leaving 46 in the dataset.

The second exclusion criterion referred to the non-associative measures of inhibition, participants who did not follow the instructions or did not complete all tasks were excluded from the second part of the analysis that focused on the relationship between associative inhibition and non-associative inhibition.

For the analysis R (R Core Team, 2021) was used, the data analysis used a variety of tests including: generalised linear models, Wilcoxon tests, and multiple regressions.

The first part of the analysis focused on the comparison between the two groups: control (cue alone extinction) and super-extinction. Using the extinction training a linear mixed model (lmer4 package version 1.1.27.1) for binary data was defined to assess whether there was a difference between two groups during the extinction phase. The model included group (control vs super-extinction), trial (numerical 1−8), and the interaction between the two as fixed factors, and participants as random factors, allowing individual slopes to be computed for every participant. The trial was reverse coded so that the last trial of extinction was trial 0 (reverse trial = 8 − trial) to aid with the interpretation as the intercepts would reflect terminal performance following extinction. From this model the extinction slopes were extracted and later used to assess whether the speed of extinction was linked with the non-associative inhibition. Next two Wilcoxon rank-sum tests data were used to assess whether there was a difference between the control group and the super extinction group in: 1) the level of context inhibition developed and 2) the revel of recovery observed.

The second part of the analysis focused on assessing the link between associative inhibition and non-associative inhibition. For the current study associative inhibition was defined in two distinct way. The first measure of associative inhibition was represented by the speed of extinction, more specifically the individual slopes of extinction extracted for each participant. These slopes represent the amount of inhibition displayed in the extinction training, steeper slopes being indicative of more inhibition and vice versa. The second measure of inhibition was the amount of context inhibition developed to the extinction context. Accordingly, less responding to the context inhibition test was interpreted as indicative of more context inhibition.

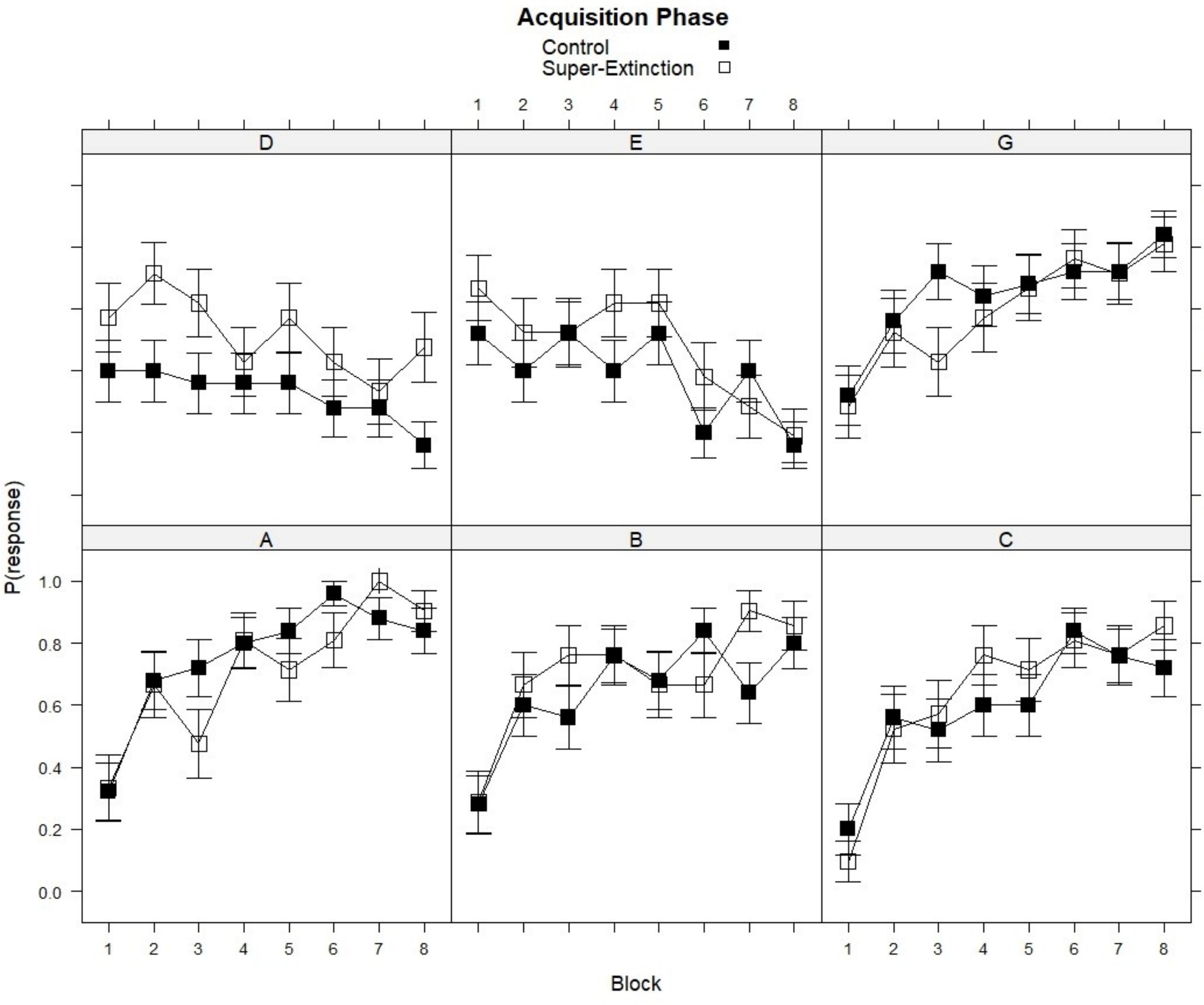
3.1.2 Results

3.1.2.1 Acquisition

The training performance of the two groups is shown in Figure 15, which indicates that participants successfully learnt to respond to the reinforced cues and not to respond to the non-reinforced ones.

Figure 15

Acquisition Phase of the Control and Super-Extinction Groups



3.1.2.2 Extinction

The extinction performance of the two groups is shown in Figure 16. According to Figure 16 there was little to no difference between the two groups in terms of how fast extinction was learnt. Additionally, no summation was observed for the super-extinction group.

The linear mixed model confirmed that there was no difference in extinction acquisition between the two groups (Table 24). This meant that both cue alone extinction and super-extinction were learnt at the same rates in two groups, contrary to the predictions of the Rescorla-Wagner model.

Figure 16

Extinction Phase Learning of the Control and Super-extinction Groups

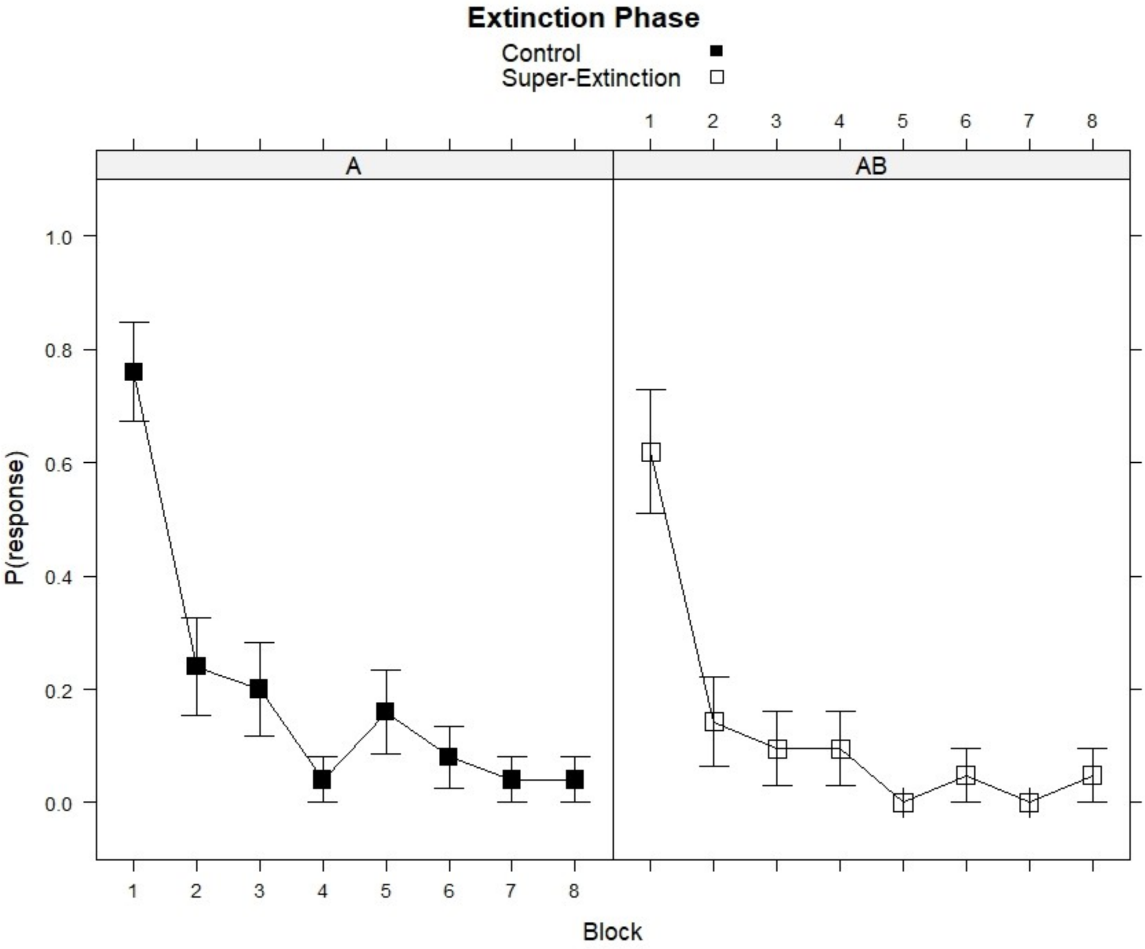


Table 24

Extinction Learning over Time by Group

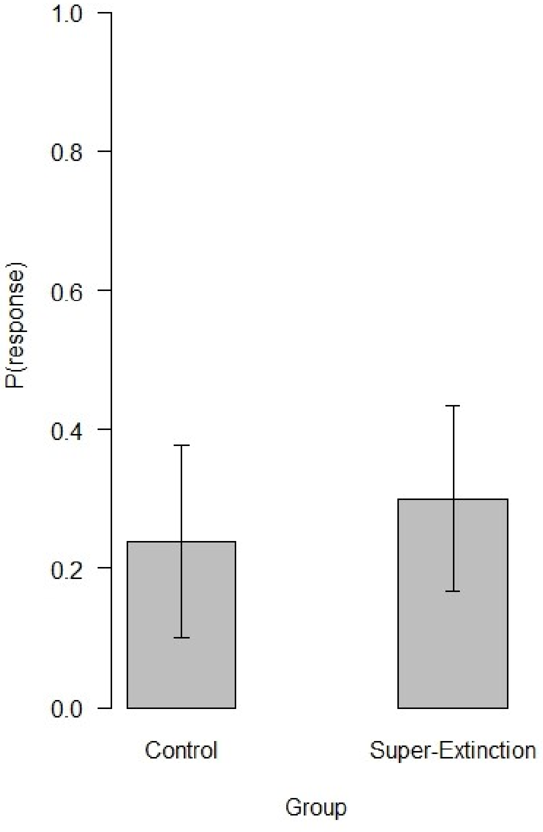
Model	Fixed Effect	Estimate	SE	z	p
Group * Block	Intercept	−6.33	1.42	−4.45	< .001*
	Group	−1.78	1.68	−1.06	.29
	Block	3.02	0.72	4.20	< .001*
	Cue * Block	0.66	0.82	0.81	.42

3.1.2.3 Context Inhibition

The context inhibition test performance of the two groups is shown in Figure 17, a Wilcoxon rank-sum test revealed that there was no significant difference in the amount of context inhibition developed by the two groups *Z* = -0.27, *p* = .79, *r* = - .04.

Figure 17

Context Inhibition Test by Group

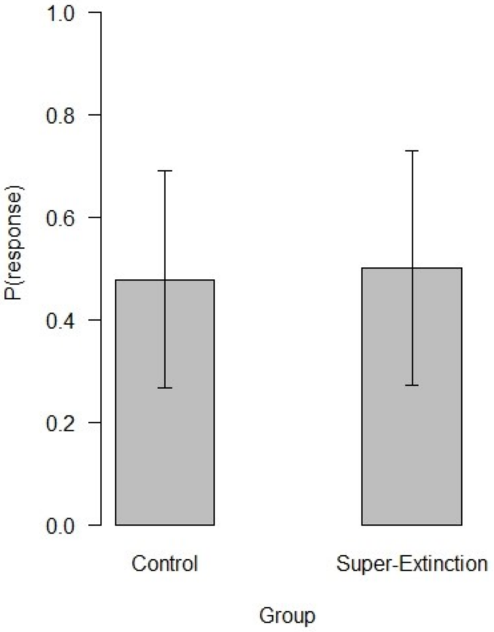


3.1.2.4 Recovery

The amount of recovery exhibited by the two groups following the context change is shown in Figure 18. A Wilcoxon rank-sum test confirmed that the two groups did not significantly differ in the amount of recovery shown after the extinction in a novel context Z = -0.03, p = .98, r = - .004.

Figure 18

Recovery Shown by the Two Groups



3.1.2.5 Non-associative Inhibition

3.1.2.5.1 Extinction Rates

The link between associative and non-associative inhibition was assessed using the individual extinction slopes computed for every participant as a dependent variable in a multiple regression with all measures of non-associative inhibition as predictors. The regression was computed separately for the control and the super-extinction group.

For the control group the test revealed a significant effect of BIS which suggested that participants with steeper slopes (higher values for the extinction slopes) had lower BIS scores meaning that participants with low BIS scores acquired extinction faster (Figure 19, Table 25). All other relationship were not significant (Table 25).

For the super-extinction group, none of the effects were significant (Table 25).

Table 25

Effects of Non-associative Inhibition on Extinction Slopes

Group	R^2	dfs	F	p
Control	.11	5, 17	1.70	.22

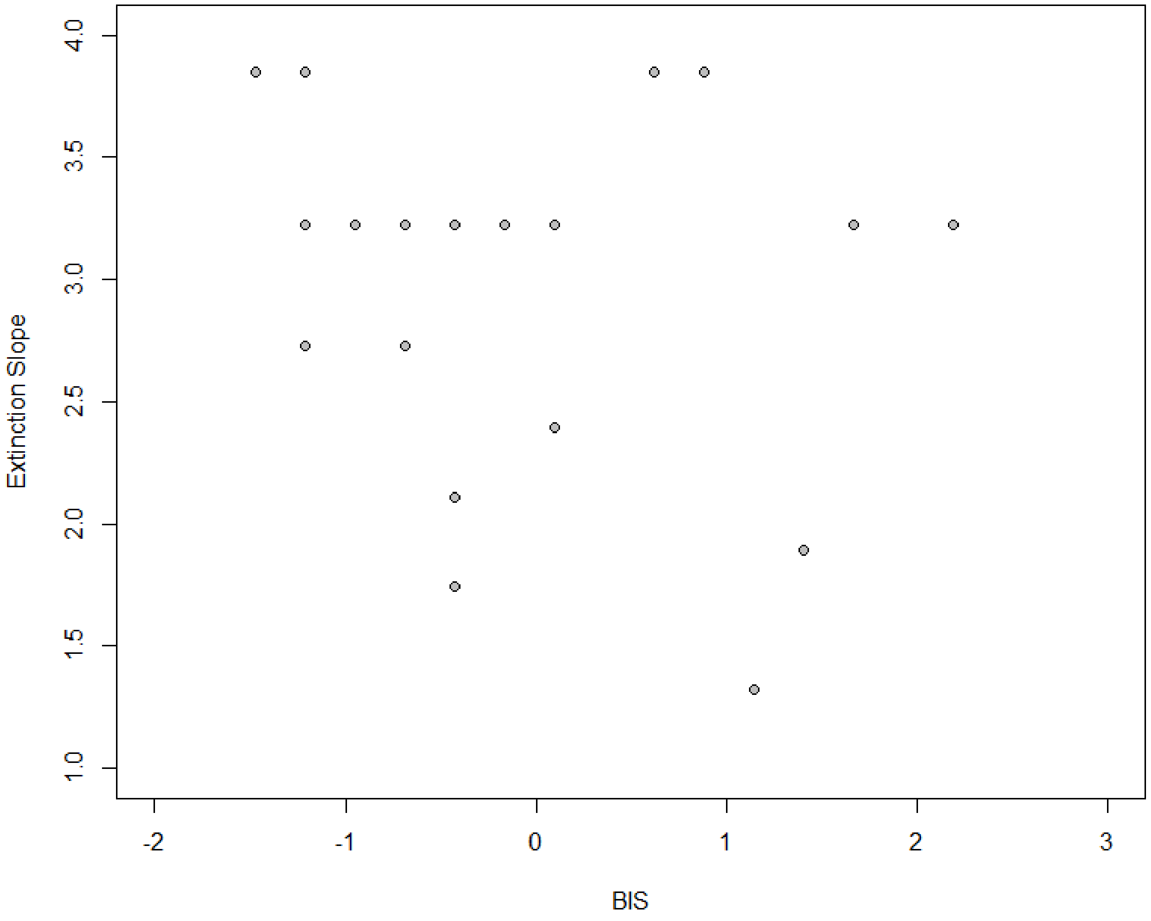
	Non-associative Inhibition	Unstandardized β	t	p
	Intercept	2.91	14.35	< .001*
	BIS11	−0.08	−0.42	.68
	BAS	0.06	0.28	.79
	BIS	−0.38	−2.19	.04*
	DD	0.29	1.17	.26
	SSRT	0.17	0.67	.51

Group	R^2	dfs	F	p
Super Extinction	.06	5, 14	0.79	.58

	Non-associative Inhibition	Unstandardized β	t	p
	Intercept	2.83	12.74	< .001*
	BIS11	0.16	0.50	.63
	BAS	−0.31	−1.45	.17
	BIS	−0.33	−1.01	.33
	DD	−0.22	−0.89	.39
	SSRT	−0.16	−0.83	.42

Figure 19

Effect of BIS on Extinction Slopes



3.1.2.6 Context Inhibition

A logistic regression was used to assess whether non-associative inhibition was linked to the amount of context inhibition developed by participants within each of the two groups independently.

For the control group the regression revealed that none of the non-associative inhibition measure were significant predictors of context inhibition (Table 26). No significant effects were found for the super-extinction either (Table 26).

Table 26

Effects of Non-associative Inhibition on Context Inhibition

Group	Cox &Snell R^2	McFadden R^2	dfs	X^2	p
Control	.27	.23	1,17	7.08	.22

Non-associative Inhibition	Estimate	Wald Statistic	p
Intercept	−0.50	0.82	.37
BIS11	−0.58	0.95	.33
BAS	−1.05	2.56	.11
BIS	0.72	2.15	.14
DD	−0.47	0.41	.52
SSRT	0.87	1.55	.21

Group	Cox &Snell R^2	McFadden R^2	dfs	X^2	p
Super Extinction	.18	.15	1,14	4.00	.55

Non-associative Inhibition	Estimate	Wald Statistic	p
Intercept	0.61	1.07	.30
BIS11	0.21	0.07	.79
BAS	0.20	0.15	.70
BIS	1.54	2.63	.11
DD	0.16	0.08	.78
SSRT	0.33	0.46	.50

3.1.3 Discussion

The current experiment used two groups of participants for whom a previously reinforced cue was extinguished using either cue alone extinction and super-extinction with the aim of assessing the differences between the two procedures. No significant differences were observed between the two groups on any of the comparisons carried out (extinction rate, context inhibition, recovery). The secondary aim of the experiment, which continued from the previous chapter was to assess the link between associative and non-associative inhibition. In the current study associative inhibition was defined as the speed of extinction (extinction slopes), and context inhibition. Using the extinction slopes as an outcome variable, a multiple regression model revealed that for the control group BIS was a significant predictor of extinction speed. According to this effect participants low on BIS showed faster extinction, however the effect was not replicated for the super-extinction group or by using the context inhibition as the outcome variable. No other significant effects were found.

The negative relationship between BIS and extinction acquisition was unexpected as high levels of BIS are assumed to be indicative of a strong inhibitory system, meaning that participants with high scores of BIS were expected to show faster extinction which would also be indicative of strong inhibition. A relationship of this nature was found in the first series of experiments presented in the current thesis where participants with higher BIS scores were observed to show more conditioned inhibition during an evaluative summation test. Together with other examples from the literature such as He et al. (2013) who found a negative relationship between BIS and conditioned inhibition and Migo et al. (2006) who found no link between BIS and conditioned inhibition, the current results do not clarify the relationship between associative inhibition and non-associative inhibition in the form of BIS. Although not systematic, given that BIS continuously reappeared as a significant predictor of associative inhibition however, could be interpreted as an early indication that associative inhibition and BIS share some underlying commonalities. In regards to the wider concept of non-associative inhibition (cognitive inhibition, response inhibition, and delayed discounting), the current results add to the conclusion drawn from the previous series of experiments that associative and non-associative inhibition are unrelated inhibition constructs. As a result, associative inhibition should be considered as an independent factor when considering the concept of inhibition and should be included in any further models as a standalone component alongside non-associative inhibition.

The current experiment did not find any significant differences in the extinction rates of the control and the super-extinction groups. According to the Rescorla-Wagner model the super-extinction procedure should have produced a faster extinction as the associative strengths of the two cues were expected to mathematically summate and drive a faster extinction acquisition. None of these a priori expectations were met, the groups performing similarly during the extinction stage, with the super-extinction group not showing summation in the first extinction trial or faster extinction. This could however be simply due to the fact that Pearce’s configural model which doesn’t predict any of the above mentioned effects is better at modelling behaviour (the possibility is considered in full in the next chapter). Summation, and faster extinction using super-extinction have been shown in previous studies (Rescorla 2000), therefore it could be the case that the design of the learning task did not allow for the effects to be observed. Although a compound was used for the super-extinction group during the extinction phase, this was the only time a compound was presented. As a result participants might have been confused by the occurrence of a second cue masking the summation effect. Additionally, even if summation did occur it would have been difficult to detect as the two cues used for the compound reached an asymptotic value by the end of acquisition. Consequently, when the associative strengths of the cues summate it would be impossible for the participant to show more expectation for reinforcement for the compound compared to the two parts of the compound. The limitations of the learning task are considered and addressed in the next section, and a new design is selected and tested.

3.2 Super-Extinction Pilot Study

Given the results of the previous study, it was decided that the design of the learning task would be updated to ensure that the conclusions of the study are robust and not due to shortcoming of the design. Two features of the previously used design were identified as potential sources of a masking effect which could explain the fact that no differences were observed between the two groups.

3.2.1 Design changes

First, summation in the first extinction trial for the super-extinction group was not observed, however this could be have been masked by a ceiling effect. The two cues used for super-extinction were approaching asymptote at the end of training, and when paired in the first trial although theoretically their associative strength would summate, participants could not express an expectation higher than 1. To avoid the ceiling effect and allow for summation to be detected during extinction, the target cues in the new design were reinforced at a 75% rate. Due to the partial reinforcement, the target cues could not reach asymptote towards the end of training allowing for a summation effect to be recorded, if present, when the two cues were paired.

Second, the previous design did not feature any compounds during training, the only compound used was the one presented during extinction for the super-extinction group. As a result, it would be reasonable to assume that not all participants, if any, expected the compound to be reinforced when first presented, as the idea of a compound itself was relatively novel to the participants. Accordingly, a compound was added to the training stage so that participants were made aware of the possibility of the cues being paired. The cues used for the compound received the same training as the target cues, meaning that they were reinforced at a rate of 75% however, the compound was always reinforced. This served as a summation demonstration aimed at increasing the chances of observing a summation when the two target cues are placed in compound for extinction.

The updated design is shown in Table 27. The four main stages were retained in the original order: acquisition, extinction, context inhibition test, and recovery test. The acquisition stage had 72 trials grouped into two blocks which were designed to deliver reinforcement with a rate of 75% to the target cues A and B, as well as the cues used for the summation demonstration K and L. Within each block there were four presentations of each type of cue, for the partially reinforced cues three of the trials were reinforced and one was non-reinforced. The order of the trials within blocks was randomised for each participant. No other changes were made to the design.

Table 27

Updated Design of Super-Extinction Learning Task

	Acquisition	Extinction	Test G	Test A
	A:	B:	B:	C:
Super-extinction	A → X x6	AB → Zx8	G → Z x2	A → Z x1
	A → Z x2	C → Y x8		
	B → X x6			
	B → Z x2			
	C → Y x8			
	D → Z x8			
Control	E → Z x8	A → Zx8		
	G → X x8	C → Y x8		
	K → Y x6			
	K → Z x2			
	L → Y x6			
	L → Z x2			
	KL → Y x8			

3.2.2 Exclusion Criterion

As a result of the target cue receiving partial reinforcement during training the previously used extinction criterion was no longer appropriate for selecting participants who showed learning. The criterion was updated to fit the new design: participants were selected on the basis of their responses to cues A, C, D, and E in the last block of acquisition. Using the responses to these cues two matrices were formed to show the number of X responses to 1) cue A and 2) cues C, D, and E together. Next a one sided Wilcoxon rank-sum test was computed for each participant to determine if participants showed a higher proportion of X responses to the cue that was partially reinforced with outcome X (A), compared to the cues that were either reinforced with outcome Y or non-reinforced (C and D/E respectively). Participant who showed a higher proportion of X responses to A compared to the rest of the cues together were included in the data analysis.

3.2.3 Participants

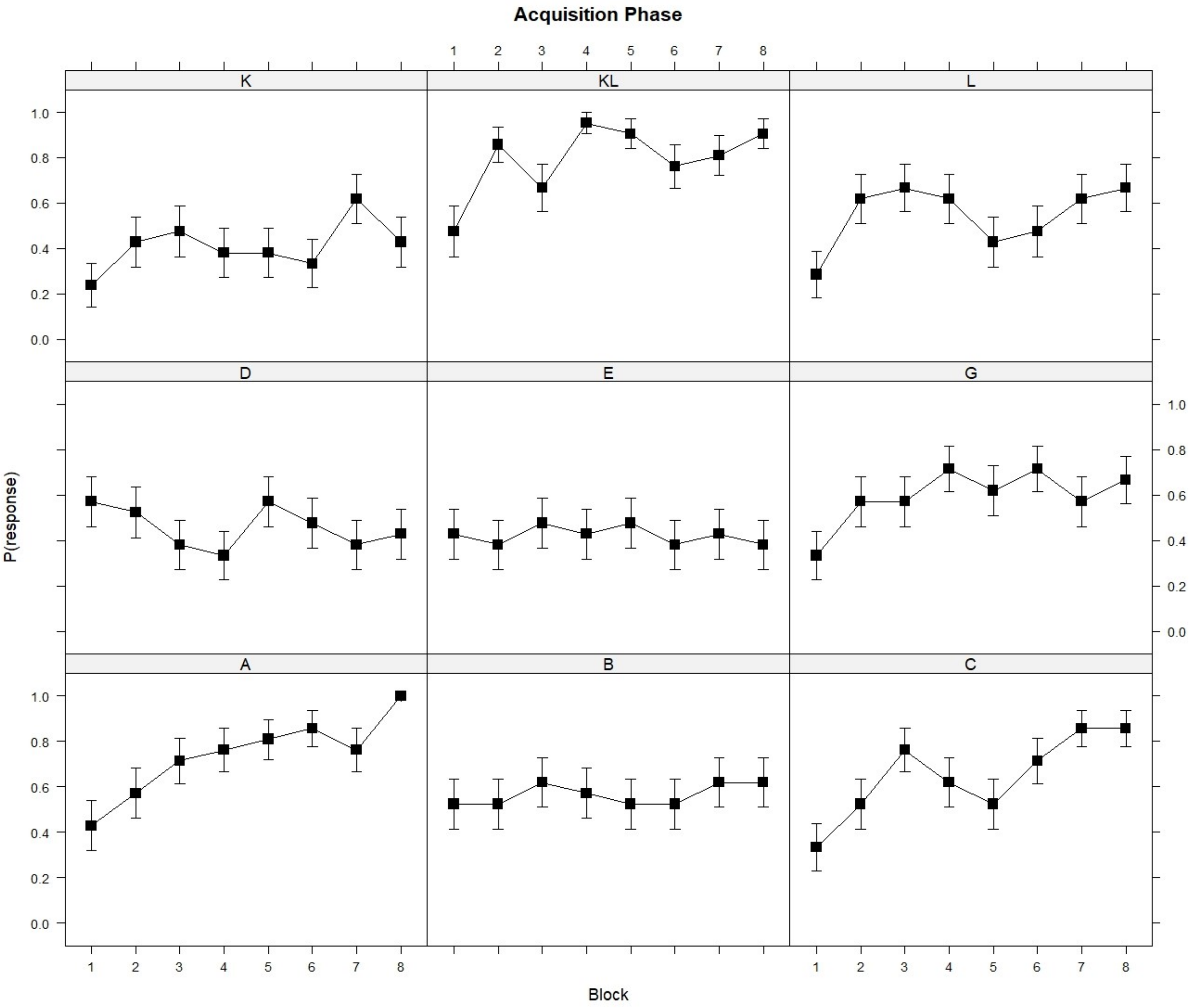
A total of 70 participants were recruited from the University of Southampton Highfield Campus and were awarded credit for their time. Demographic information was not collected.

3.2.4 Data analysis

When the new exclusion criterion was applied to the data, only 21 participants passed the criterion, 13 were in the control group and 8 were in the super-extinction group. The acquisition stage performance of the participants who passed the exclusion is shown in Figure 20. According to Figure 20 the learning that occurred during the training stage did not fully follow the expectations. Although participants did respond more to the cues that were always reinforced compared to the partially reinforced ones, there was still a relatively high level of responding to the non-reinforced cues. This indicates that participants did not have enough time to fully learn about the cues which could have been the result of having only two block of training and a total of eight distinct cues and a compound. Due to the small sample the results of the following analysis lacked robustness but were informative in developing the final design of the learning task.

Figure 20

Acquisition Performance

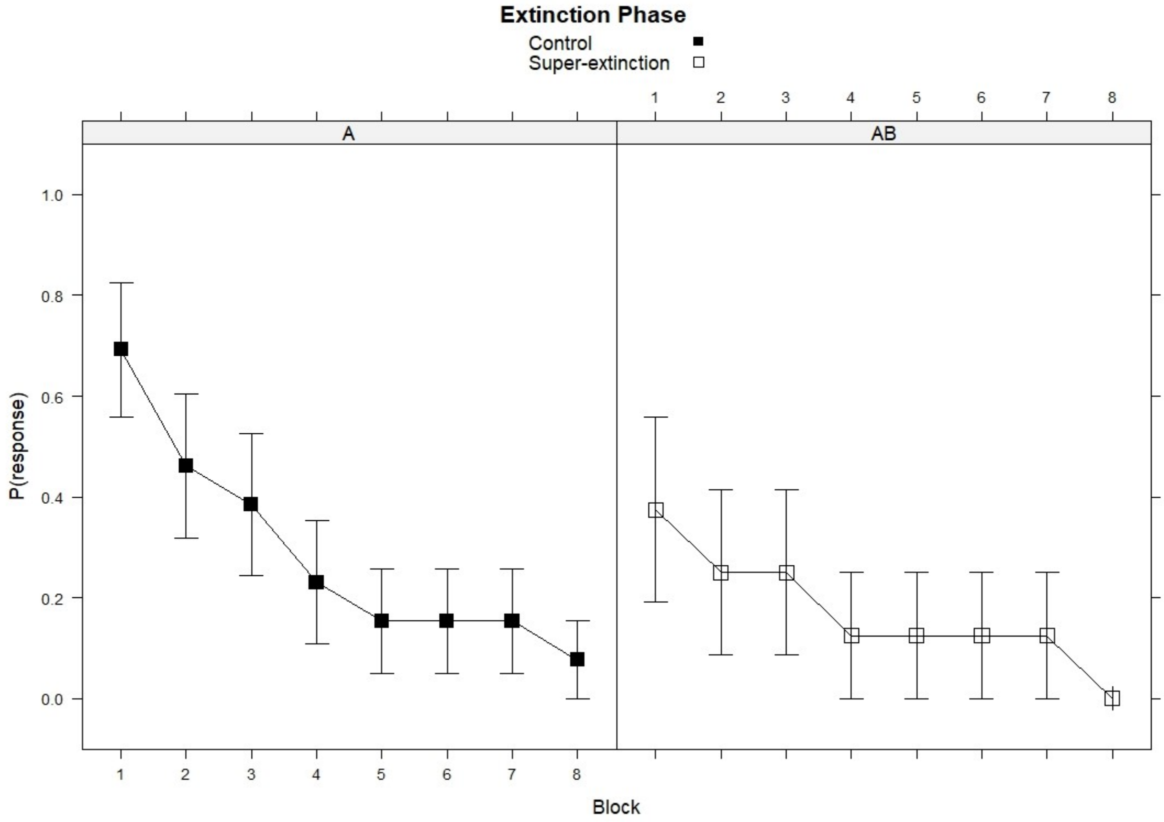


3.2.4.1 Extinction

The extinction of the two groups is shown in Figure 21, according to which the extinction of cue A seemed to be faster when it was carried out in compound however there was no summation effect in the first trial of extinction.

Figure 21

Extinction Phase



A linear mixed model with participants' responses throughout the extinction stage as a dependent variable and group (control vs extinction) and trial (reverse coded 0-7) revealed that the difference between the two groups was not statistically significant Table 28.

Table 28

Extinction Learning

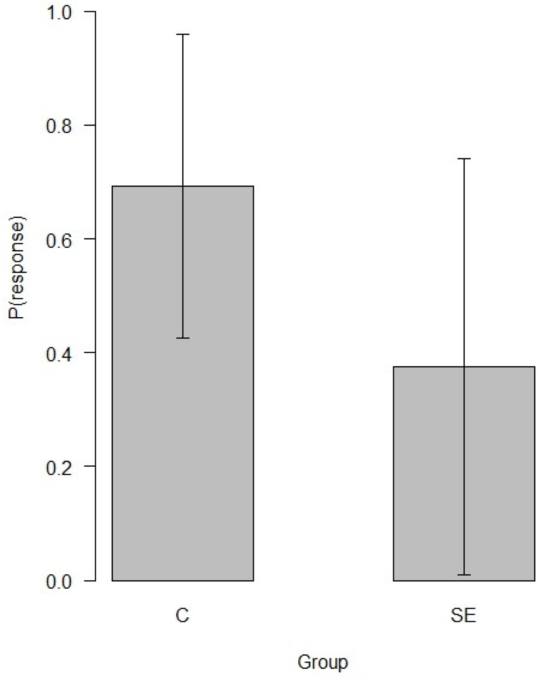
Model	Fixed Effect	Estimate	SE	z	p
Group * Block	Intercept	−12.10	4.33	−2.79	.005*
	Group	−4.44	5.90	−0.75	.45
	Block	2.95	1.06	2.78	.005*
	Cue * Block	0.61	1.39	0.44	.66

3.2.4.2 Context inhibition

The context inhibition test is shown in Figure 22, the difference between the two groups was found to be not significant using a Wilcoxon rank-sum test $Z = -1.39$, $p = .16$, $r = -0.30$.

Figure 22

Context Inhibition

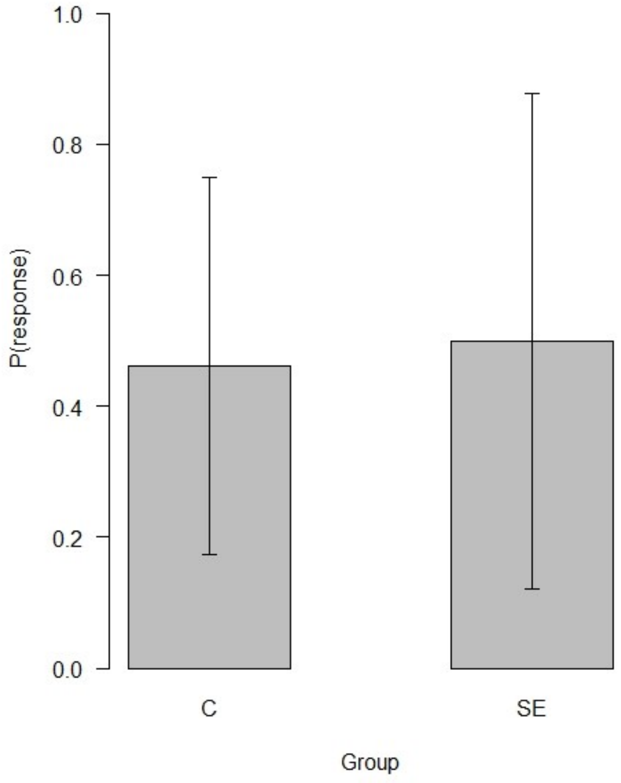


3.2.4.3 Recovery Test

During the recovery test both groups showed similar levels or recovery as shown in Figure 23. A Wilcoxon test confirmed that the two groups did not differ based on the level of recovery shown $Z = -0.17$, $p = .87$, $r = -0.04$.

Figure 23

Recovery Test



The pilot study results indicated that the updated design of the learning task resulted in some differences on extinction learning and context inhibition between the two groups, although the analysis returned non-significant results. The exclusion criterion was updated to match the partial non-reinforcement of the cues however it was clear that the two acquisition blocks were not sufficient for participants to show robust learning. As a result for the final iteration of the study the number of training blocks was doubled. Additionally a third group was added, a deepened extinction group which is described below.

3.3 Study 2

The final study from the current series added a third group to the design aiming to assess the differences in extinction between cue alone, super-extinction and deepened extinction. Super-extinction differs from deepened-extinction in that deepened-extinction is a "post-extinction" procedure (Leung et al., 2012) involving two extinction phases. In the first extinction phase of a deepened-extinction procedure the target cue is extinguished alone and only after this initial extinction of the target is a compound involving the target and a second, non-extinguished, excitatory cue introduced. In the second extinction phase this compound is presented non-reinforced. This difference could be theoretically as well as practically important since according to the Rescorla-Wagner model, in a simple super-extinction procedure, there should be more rapid extinction when compared to extinction of a single cue, but asymptotically both procedures would lead to the associative strength of the target cue falling to zero. In contrast, in a deepened-extinction procedure, it would be possible for the target cue to acquire inhibitory strength and so cue-exposure treatment with deepened-extinction may be more effective than single cue or super-extinction e.g. because the target would be less likely to have residual post-treatment associative strength. However, as will be discuss below, the superiority of the deepened-extinction procedure is not anticipated by the Pearce configural nor by a frequently cited development of the Rescorla-Wagner model, the configural Rescorla-Wagner model (Rescorla, 1973) which will be described in more detail below in the next chapter.

As previously mentioned, there are many studies showing that increased prediction error can increase extinction relative to simple single-cue extinction procedures. However, notably in human studies, there are several published investigations which indicate that strategies to increase prediction error do not always have the anticipated effect. For example, Griffiths et al. (2017), using a predictive learning task, found that whilst a super-extinction procedure resulted in faster compound extinction there was no evidence that extinction of the individual cues differed from a target which had undergone deepened-extinction, nor was there any evidence that deepened extinction resulted in greater extinction than obtained using a simple single-cue extinction procedure.

The current study follows the methodology previously established, with any differences detailed below. The main aim of the study was to assess the difference between the three types of extinction with a secondary aim of assessing the link between associative and non-associative inhibition.

3.3.1 Method

3.3.1.1 Participants

A sample of 207 student participants was recruited through a subject pool run in the Department of Psychology at the University of Southampton, by posted adverts, and by word of mouth. The average age was 19.5 years and 164 participants identified as female, 41 identified as males, one identified as non-binary, and one did not provide an answer.

3.3.1.2 Questionnaires

The questionnaires used in the current experiment were the same as the ones used in the first conditioned inhibition study described in more detail under section 2.1.1.2 Questionnaires

3.3.1.3 Stop Signal Reaction Time Task

The current study used the same stop signal reaction task as the last conditioned inhibition study, the full description can be found at 2.3.1.3 Stop Signal Reaction Time Task.

3.3.1.4 Learning Task

The learning task used was the same as the one used in conditioned inhibition study 2, the full description ca be found at 2.3.1.4 Learning Task. The only difference was FULF's reaction, the outcome, was one of three possibilities – happy, sad, or neutral, as per the experimental design. The happy and sad outcomes were the reinforced outcomes and participants were instructed to press the 'h' or 's' keys to predict these outcomes and to refrain from pressing any key if the neutral, non-reinforced, reaction was expected. Additionally, contexts were introduced which were represented by a faded wallpaper in the background (Appendix B).

3.3.1.4.1 Design

The design of the learning task is given in Table 29. There were a total of 179 trials split into five phases – 144 acquisition phase trials, 16 extinction phase 1 trials, 16 extinction phase 2 trials, and two test phases. The summation test phase came first and consisted of two trials and the experiment finished with the recovery test phase which was a single trial. Acquisition took place in context A:, the extinction and summation phases took place in context B:, and recovery was in context C:. Cues were presented in trials that were either reinforced by presentation with an outcome or non-reinforced by presentation without an outcome. There were two types of reinforced trials, those with happy and those sad outcomes which are coded X and Y in Table 29; the non-reinforced trials are coded Z. The assignment of happy and sad outcomes to X and Y was randomised so that for approximately half of the participants X corresponded to sad and Y corresponded to happy and vice-versa for the other half. The acquisition and extinction phases had multiple trials divided into blocks with trial order was randomised independently for each participant within block. The acquisition phase had four blocks. Within each acquisition block there were four presentations of each cue with outcomes delivered according to a continuously reinforced (e.g. four C → Y) or partially reinforced schedule (e.g. three A → X and one A → Z trials as per the design in Table 29. Throughout the experiment cues and outcomes were presented in one of three visually distinctive contexts as per the design. Screen background images were used to provide context cues. For each participant the backgrounds were selected at random, without replacement to serve each of the three contextual functions (A:, B:, and C:), from a collection of five possible backgrounds (Appendix B).

Each of the two extinction phases contained eight blocks with each block containing one trial of each of the types shown in Table 29. Cue A was the critical cue for testing the effects of deepened and super-extinction. For the control group cue A was extinguished alone during both extinction phases. In the deepened-extinction condition A was extinguished alone during extinction 1 and in compound with cue B during extinction 2. In the super-extinction condition cue A was extinguished in compound with B during both extinction 1 and extinction 2. Cue G was used in a summation test to assess the inhibitory strength of the extinction context after extinction was finished. Cue A was presented for a renewal test in a novel context, context C:, after the summation test.

It was assumed that if compound extinction was to increase extinction above that seen with single cue extinction that participants would have to summate outcome expectations generated by multiple cues in the manner suggested by associative models such as the Rescorla-Wagner model. In order to maximise the likelihood that such summation would occur cues A and B were partially reinforced with outcome X during the acquisition phase and cues K and L were partially reinforced with outcome Y. Cues K and L were also presented in a continuously reinforced KL compound as a ‘demonstration’ of cue additivity.

Additional cues C, D and E were used to equate the number of different outcome types on the single-cue trials during acquisition. Cue C was presented with outcome Y during the extinction phase, as in the acquisition phase, to provide some continuity between phases to avoid giving the impression that all reinforcement stopped suddenly after the change from acquisition to extinction context.

Table 29

Design of Learning Task for Study 2

	Acquisition	Extinction 1	Extinction 2	Summation	Recovery
	A:	B:	B:	B:	C:
Super Extinction	A → X x12	AB → Zx8	AB → Zx8	G → Z x2	A → Z x1
	A → Z x4	C → Y x8	C → Y x8		
	B → X x12				
	B → Z x4				
Deepened Extinction	C → Y x16	A → Zx8	AB → Zx8		
	D → Z x16	C → Y x8	C → Y x8		
	E → Z x16				
	G → X x16				
Control	K → Y x12	A → Zx8	A → Zx8		
	K → Z x4	C → Y x8	C → Y x8		
	L → Y x12				
	L → Z x4				
	KL → Y x16				

3.3.1.5 Data selection and analysis

All analyses were carried out in R (R Core Development Team, 2020). Thirty-three of the 207 participants were excluded due to poor performance during the acquisition phase leaving 174 participants for the analyses reported below. Since the primary aim was to study extinction of responding to cue A it was required that participants had acquired appropriate responding to cue A during the acquisition phase. For each participant two binary vectors were constructed that were then compared using one-sided Wilcoxon rank-sum tests. The first vector had length 4 and was used to represent responses to cue A during the last four trials of its presentation in the acquisition phase – X responses were coded 1 with any other responses coded 0. The second vector had length 12 and was used to represent responses to cues C, D, and E during their last four presentations of the acquisition phase, again X responses were coded 1 with any other responses coded 0. Cues C, D and E were never paired with outcome

X during the acquisition phase (C was continosly reinforced with outcome Y, D and E were continuously non-reinforced) and A was paired with outcome X on 75% of its presentations. Therefore participants were included if the cue A vector was significantly greater ($p < .05$) than the cue CDE vector and excluded otherwise.

3.3.1.5.1 Effects of extinction procedures

The three extinction procedures were compared in order to determine whether or not there was any evidence for a) more rapid extinction using a compound of two excitatory cues as compared to extinction of a single excitatory cue and b) more complete extinction in deepened and super-extinction procedures as compared to a standard single-cue extinction procedure. In the case of a) a general linear mixed model with a binomial link function and random effect intercepts was fitted. The fixed effect terms were a between subjects fixed effect contrast for group, within subjects fixed effect contrasts for trial, and interaction contrasts for group and trial. The dependent variable was a binary valued vector indicating whether or not participants predicted outcome X on the last cue A trial of the acquisition phase and on each of the eight extinction 1 trials involving cue A. Participants for the control and deepened-extinction groups were treated as one 'standard single-cue' extinction group, dummy coded 0, for the purpose of this analysis since they were treated identically up until the end of extinction 1 phase. They were contrasted with the super-extinction group, dummy coded 1, which had extinction of an AB compound during extinction 1 phase. Eight dummy coded variables were used to contrast each of the extinction 1 phase trials with the last cue A trial of the acquisition phase.

In the case of b) Kruskall-Wallis non-parametric one-way ANOVA was used to compare the group response in the Recovery test phase with follow-up tests using Wilcoxon rank sum tests. In addition, it was also of interest to assess whether responding in the Recovery test phase was linked with suppression of responding to cue G in the Summation test. According to the protection from extinction account of response recovery the extinction context becomes inhibitory and release from that inhibition causes recovery of responding. Additional Wilcoxon rank sum and Kruskall-Wallis tests were therefore carried to compare the amount of responding in the Summation test for our three experimental groups and for those who did and did not respond during the Recovery test phase.

Finally, the relationship between associative and non-associative inhibition was assessed, associative inhibition was defined as extinction acquisition speed in extinction stage 1, and context inhibition. Non-associative inhibition was defined as BIS11, BIS/BAS, delayed discounting, and response inhibition. To extract the extinction slopes the general linear model used for a) was modified to have a continuous predictor for trial, and participants were allowed to have individual slopes.

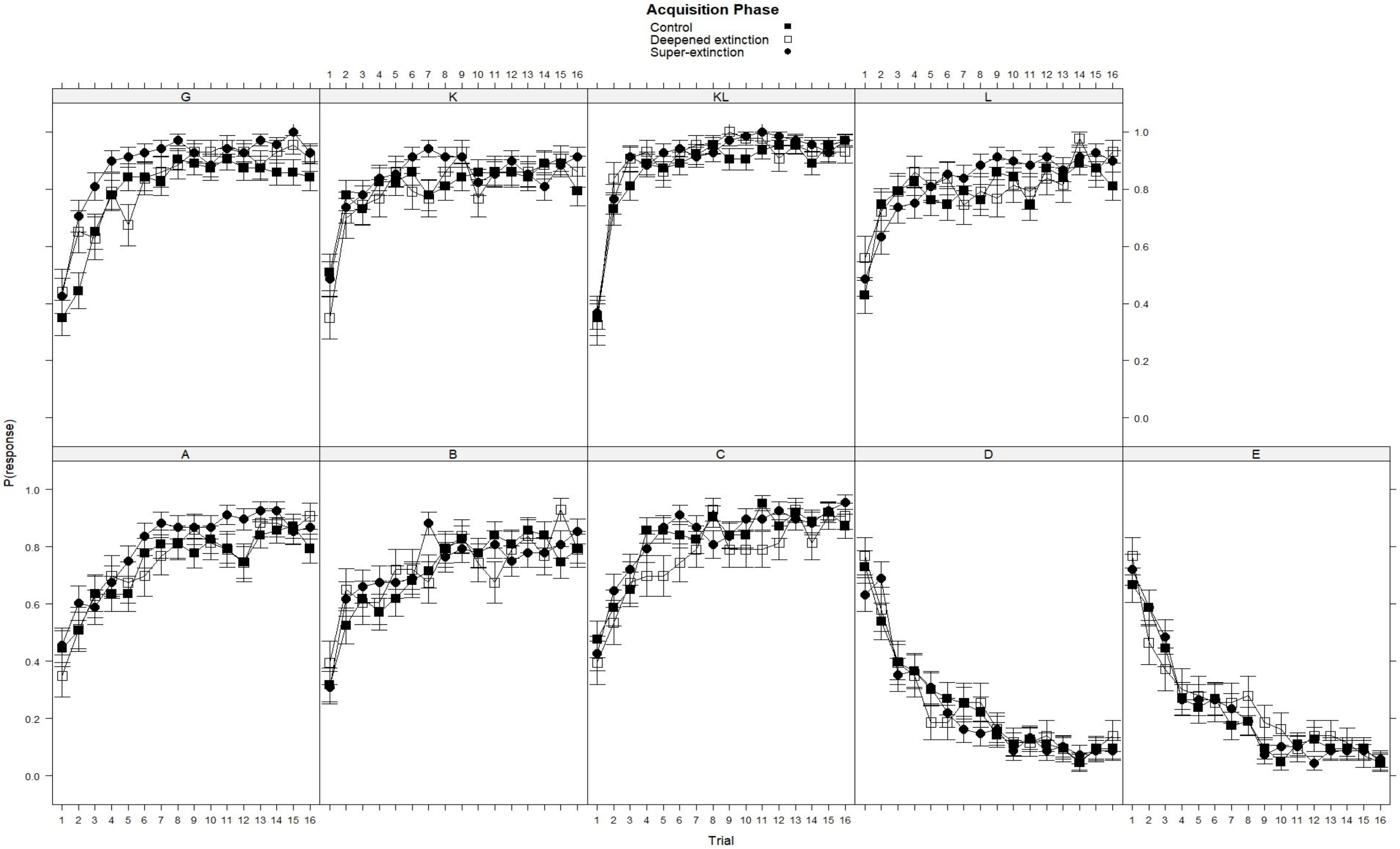
3.3.2 Results

3.3.2.1 Acquisition

Figure 24 shows that the 174 participants who passed the exclusion criterion learnt to respond to the cues as intended.

Figure 24

Acquisition Stage of the Super Extinction Study 3

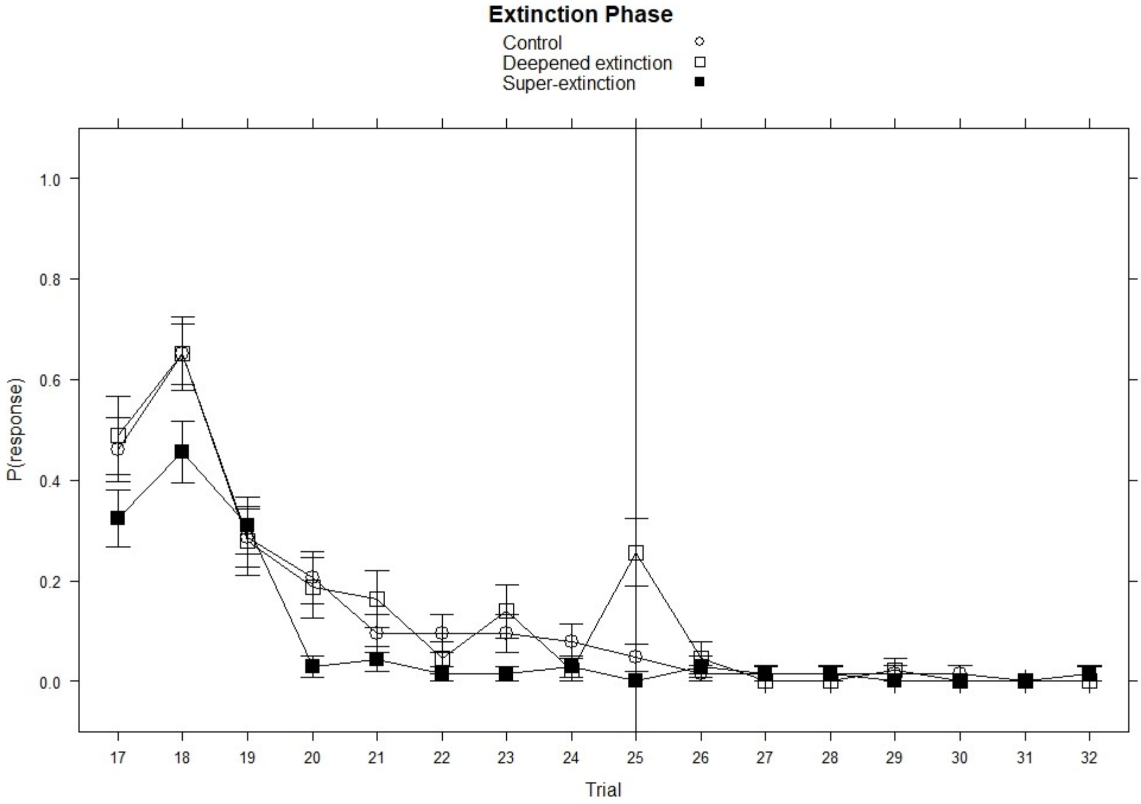


3.3.2.2 Extinction

The extinction performance of the three groups is shown in Figure 25, where the two extinction phases are separated by the middle vertical line. It can be seen that responding stopped entirely by the end of the extinction 2 phase, there were only two participants who responded on the last extinction 2 trial, one in the control group and one in the super-extinction group.

Figure 25

Probability of X-Responses during Extinction Phases 1 and 2 by Group



3.3.2.2.1 Effects of Extinction Procedures

Figure 26 shows that there is some evidence that the super-extinction group extinguished more rapidly across the extinction 1 phase trials compared to the aggregated control and deepened extinction groups. There is no indication of a summation effect on the first extinction trial when the super-extinction participants encounter AB compound cue for the first time. All groups show a marked reduction in responding on the first extinction trial. By the fourth extinction trial responding in the super-extinction group was markedly more suppressed than in the combined control and deepened-extinction group but thereafter responding equates by the end of extinction 1. Table 30 gives the fixed effect results from the general linear mixed effects model use to examine the extinction 1 phase data. Overall the Group × Block interaction was significant with a likelihood ratio test comparing models with and without the interaction contrasts yielding *χ2* = 18.50 (df = 8), *p* = .018. Confirming visual impressions the interaction contrast for the fourth extinction trial is significant (p < .01).

Figure 26

Average Probability of X-Responses by Group Over Extinction Phase 1

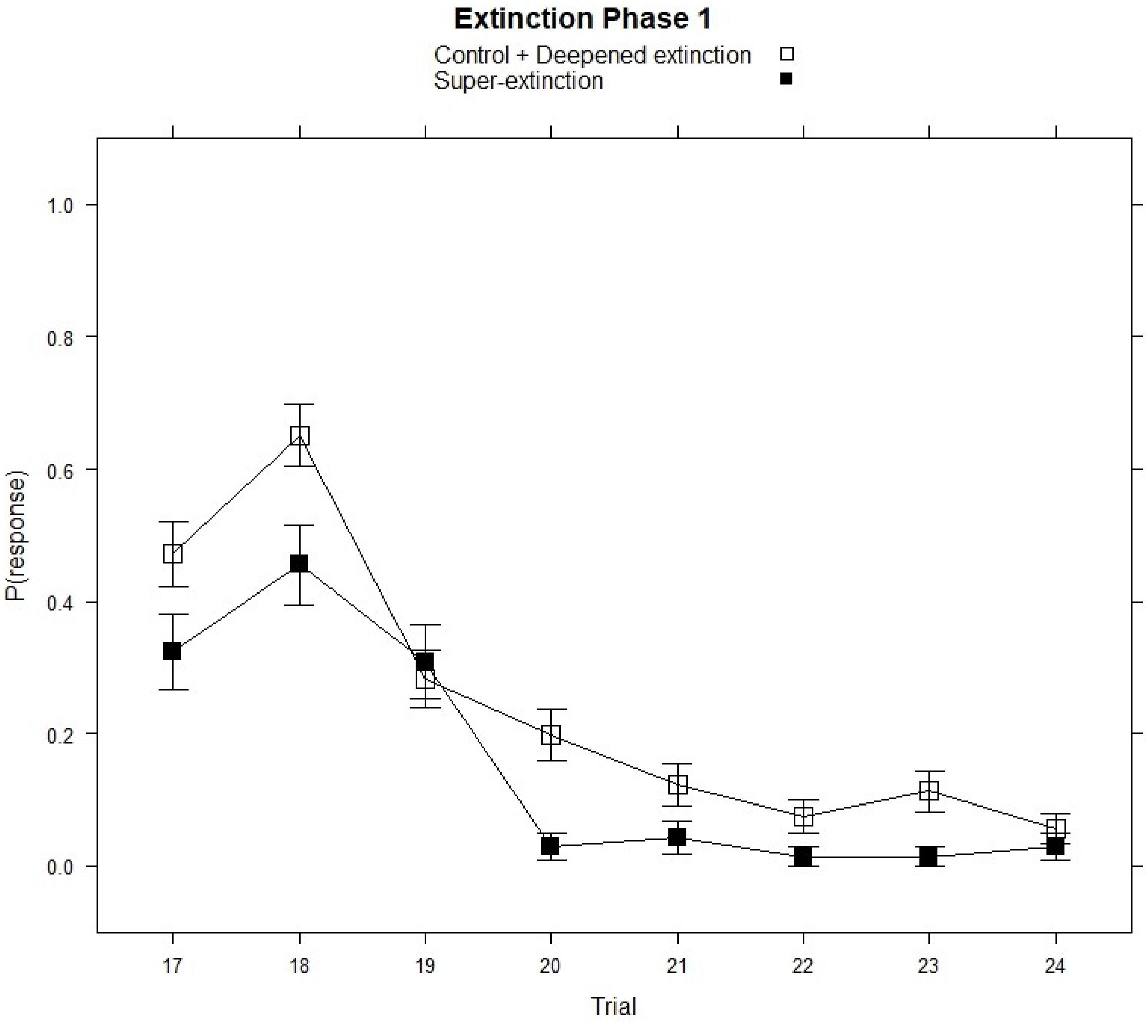


Table 30

Effects of Block and Groups for Extinction Stage 1

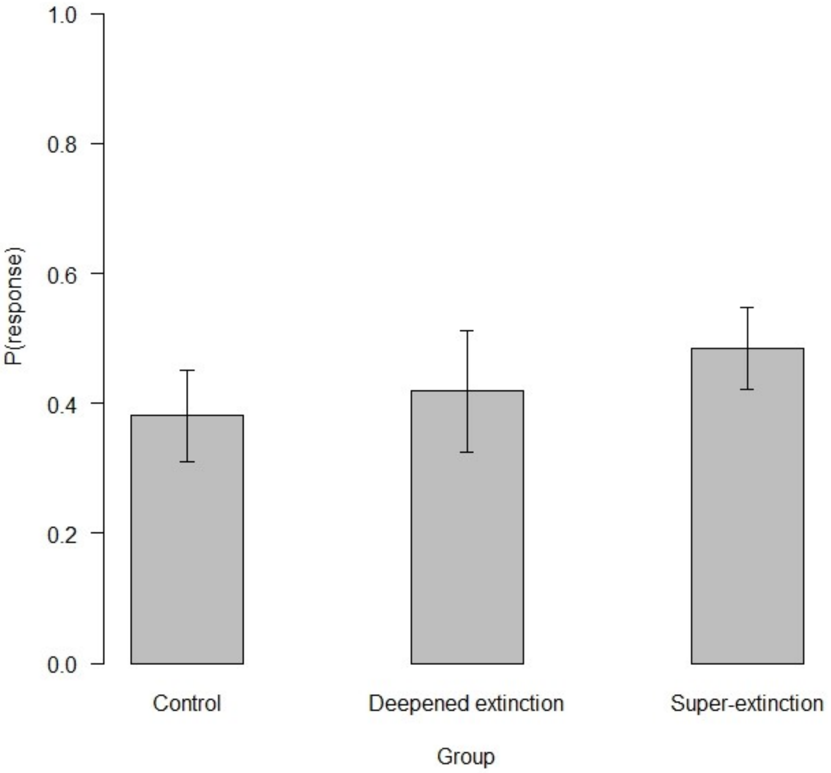
Model	Fixed Effect	Estimate	SE	z	p
Group*Block	Intercept	2.03	0.31	6.53	< .001***
	Group	0.18	0.50	0.37	.71
	Block1	−2.15	0.37	−5.88	< .001***
	Block2	−1.23	0.36	−3.37	< .001***
	Block3	−3.18	0.39	−8.17	< .001***
	Block4	−3.75	0.41	−9.08	< .001***
	Block5	−4.42	0.46	−9.70	< .001***
	Block6	−5.04	0.51	−9.79	< .001***
	Block7	−4.53	0.46	−9.75	< .001***
	Block8	−5.38	0.56	−9.64	< .001***
	Group:Block1	−0.96	0.60	−1.60	.11
	Group:Block2	−1.20	0.59	−2.03	.04*
	Group:Block3	−0.01	0.61	−0.02	.98
	Group:Block4	−2.38	0.92	−2.60	.01**
	Group:Block5	−1.27	0.84	−1.52	.13
	Group:Block6	−1.81	1.19	−1.52	.13
	Group:Block7	−2.32	1.17	−1.98	.05*
	Group:Block8	−0.75	0.98	−0.76	.45

3.3.2.3 Context Inhibition

The context inhibition test performance of the three groups is shown in Figure 27, according to which there were no differences between the groups. A Kruskal-Wallis test confirmed that the three groups did not significantly differ in the amount of context inhibition developed to context B: as a result of the extinction of cue A in this context, $X^2(2) = 4.78$, $p = .09$.

Figure 27

Context Inhibition Test by Group (Extinction Procedure)

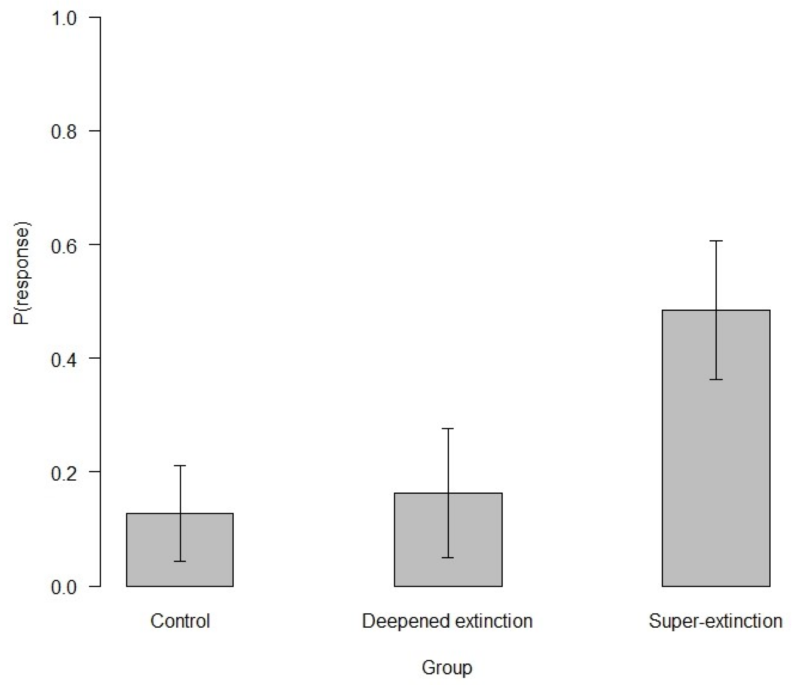


3.3.2.4 Recovery

Response recovery was observed when cue A was presented for test in context C (Figure 28). The recovery effect was much stronger in the super-extinction group than in the other groups with a Kruskal-Wallis test for the three groups producing $X^2(2) = 24.53$, $p < .001$. Follow-up Wilcoxon rank-sum tests showed that there was more recovery in the super-extinction group than in the other two groups ($W > 990.5$, $p < .002$) but the control group did not significantly differ from the deepened-extinction group ($W = 1306$, $p = .61$).

Figure 28

Recovery Effect for the Three Extinction Procedures



Finally, the average number of x-responses (minimum=0, maximum=2) to cue G in the Summation test was lower (.409) in those who did not respond in the Recovery test than it was for those who did respond in the Recovery test (.49), but the differences were not significant ($W = 2598.5$, $p = 0.08$). This suggests that increased context inhibition, which would have reduced responding in the Summation test, was not linked to greater responding in the Recovery test.

3.3.2.5 Non-associative Inhibition and Extinction

3.3.2.5.1 Extinction Rate

To extract the extinction slopes for every participant, the previous linear mixed model was recomputed with block as a continuous fixed factor. The slopes were then used as dependent variables in a series of three multiple linear regressions with the measures of non-associative inhibition as predictors. The regressions were computed independently for each of the three groups. None of the effects were significant (Table 31).

Table 31

Effects of Non-associative Inhibition on Extinction Slopes

Group	**R^2**	**dfs**	**F**	**p**
Control	.06	5, 37	1.55	.20
	Non-associative Inhibition	**Unstandardized β**	**t**	**p**
	Intercept	−0.81	−18.90	< .001
	BIS11	0.07	1.64	.11
	BAS	0.07	1.68	.10
	BIS	−0.07	−1.55	.13
	DD	−0.04	−1.02	.31
	SSRT	−0.01	−0.26	.79

Group	**R^2**	**dfs**	**F**	**p**
Super Extinction	.11	5, 38	0.17	.97
	Non-associative Inhibition	**Unstandardized β**	**t**	**p**
	Intercept	−0.78	−24.32	< .001
	BIS11	−0.02	−0.55	.59
	BAS	−0.004	−0.09	.93
	BIS	−0.01	−0.37	.72
	DD	0.03	0.76	.45
	SSRT	0.01	0.35	.73

Group	**R^2**	**dfs**	**F**	**p**
	.01	5, 29	1.10	.38
	Non-associative Inhibition	**Unstandardized β**	**t**	**p**
Deepened Extinction	Intercept	−0.80	−16.63	< .001
	BIS11	0.03	0.64	.53
	BAS	0.04	.94	.35
	BIS	−0.07	−1.39	.18
	DD	−0.04	−0.54	.59
	SSRT	0.06	1.28	.21

3.3.2.5.2 Context Inhibition

Context inhibition was the second measure of associative inhibition used as part of the current analysis. Similar to the extinction slopes the amount of context inhibition developed was used as a dependent variable in a series of three multiple regressions. For simplicity participants were classified into two groups: context inhibition (0 x-responses) and no inhibition (1 or 2 x-responses), therefore logistic regressions were used for the three groups independently. None of the effects were significant, however a few were approaching significance across the three groups (**Table 32**).

Table 32

Effects of Non-associative Inhibition on Context Inhibition

Group	Cox &Snell R^2	McFadden R^2	dfs	X^2	p
Control	.15	.13	1,37	7.02	.22
	Non-associative Inhibition		**Estimate**	**Wald Statistic**	**p**
	Intercept		1.10	7.05	.008
	BIS11		0.20	0.30	.58
	BAS		−0.13	0.10	.76
	BIS		−0.74	3.07	.08
	DD		0.61	3.06	.08
	SSRT		0.63	2.31	.13
Group	**Cox &Snell R^2**	**McFadden R^2**	**dfs**	**X^2**	**p**
Super Extinction	.27	.23	1,38	9.55	.09
	Non-associative Inhibition		**Estimate**	**Wald Statistic**	**p**
	Intercept		2.29	11.72	< .001
	BIS11		−1.08	3.24	.07
	BAS		−0.10	0.03	.87
	BIS		−0.61	1.39	.24
	DD		1.01	3.37	.07
	SSRT		−0.52	0.65	.42
Group	**Cox &Snell R^2**	**McFadden R^2**	**dfs**	**X^2**	**p**
Deepened Extinction	.14	.12	1,29	5.33	.38
	Non-associative Inhibition		**Estimate**	**Wald Statistic**	**p**
	Intercept		1.05	5.51	.02
	BIS11		−0.05	0.01	.91
	BAS		0.77	3.49	.06
	BIS		0.58	1.45	.23
	DD		−0.18	0.10	.75
	SSRT		0.42	1.09	.30

3.3.3 Discussion

The current experiment used three groups to assess the difference between cue alone extinction, super-extinction, and deepened extinction. During acquisition all groups received identical training in context A:, next extinction was carried out in context B:. The extinction phase was divided into two stages. The cue alone extinction was exposed to non-reinforced presentations of the target cue throughout both stages of extinction. The two stages were also identical for the super-extinction group for which the target cue was extinguished in compound with a cue that received reinforced training during acquisition. The deepened extinction group cue alone extinction in the first stage of extinction, followed by compound extinction with a cue that received reinforced training during acquisition in the second stage of extinction. First, the extinction rates of the three groups in extinction stage 1 were assessed. Cue alone and deepened extinction groups received identical training throughout acquisition and extinction stage 1 therefore they were grouped together. The analysis revealed that the super-extinction group seemed to have extinguished faster compared to the compound control and deepened extinction group. Although a faster extinction was observed, no summation effect was detected meaning that the faster extinction was not driven by a summation effect, or at least not a visible one. The three groups did not differ on the context inhibition test, however during the recovery test significantly more recovery was observed for the super-extinction group compared to the two other groups. No significant difference was observed between the control group and the deepened extinction group. The fact that the super-extinction group showed a faster extinction would suggest they used/developed more inhibition. This was however not supported by the context inhibition test as the three groups did not differ. Because of this lack of difference the recovery cannot be attributed to the context becoming more inhibitory for the super-extinction group and protecting the target from extinction. The current results contradict the predictions made by the Rescola-Wagner model which indicates that more complex processes are involved in cue interactions than the one assumed by this model.

For the second part of the analysis the link between associative and non-associative inhibition was assessed. Associative inhibition in the current study was defined as the rate of extinction in the form of extinction slopes, and context inhibition and non-associative inhibition was one of the following: BIS11, BIS/BAS, delayed discounting, and response recovery. No significant relationship was found between any of the variables.

The evidence in support or against the effectiveness of super-extinction and deepened extinction is divided in the current literature. Rescorla (2000) showed that super extinction produced less recovery than cue alone extinction, however in his studies a more rapid extinction for this group was not observed. Thomas and Ayres (2004) showed both a faster extinction for super-extinction and less recovery, however Griffiths et al. (2017) reported that although super-extinction resulted in faster extinction, this method along with the deepened extinction method did not lead to less recovery compared to single cue extinction. Furthermore, the cue used in the super-extinction group showed more recovery than the one used in deepened extinction, which is in line with the current results as a faster extinction was observed for the super-extinction group, followed by equivalent levels of recovery for the control and deepened extinction groups and more recovery for the super-extinction group.

Chapter 4 Formal Associative Models

The current chapter focuses on comparing three formal associative learning models: Rescorla-Wagner, Rescorla-Wagner with configural cues, and Pearce's configural model, with the aim of determining which of these models is best at predicting participants' behaviour. This was the secondary aim of the extinction series of studies, and the data from the final study (Study 2) was used for the model comparison. The three models were therefore compared on their ability to predict behaviour across three extinction procedures: single-cue extinction, super-extinction, and deepened extinction.

It is clear a) that the simple Rescorla-Wagner model can predict more rapid extinction in super-extinction than in single-cue extinction and that asymptotically deepened-extinction will be more effective than super-extinction and it is also clear b) that these predictions are not shared by the Pearce configural model nor by the configural Rescorla-Wagner model. This is driven mainly by the assumptions of the Rescorla-Wagner model that compounds are perceived to have the expectation for reinforcement equal to the sum of their parts while Pearce's configural model model assumes that this is equal to a weighted sum based on the similarity between previous trials and the current trial. As a result, the Rescorla-Wagner predicts significantly higher expectation for reinforcement in both previously-mentioned examples. However, despite this, there is no basis for a theoretically decisive test because these model predictions are dependent on both procedural and model parameters. For example, as super-extinction will asymptotically be equivalent to single cue extinction if there are too many extinction trials then differences between single cue and super-extinction conditions may not be detected. And if the second stage of a deepened-extinction procedure is introduced too early then any differences between super-extinction and deepened-extinction may also be difficult to detect. Furthermore, optimal procedural parameters will depend on model parameters. Additionally, since the predictions outlined above are based on associative strength, without assuming any more than a monotonic mapping to response strength, they are qualitative rather than quantitative. Therefore, in the work presented below, a softmax function (Ahn et al., 2008; Yechiam & Busemeyer, 2005) was applied to map between associative strength and response probability in order to estimate the likelihood of observed participant behaviour under maximum likelihood parameterisation of each of the three models. With these likelihood estimates an Akaike weight analysis (Burnham & Anderson, 2002) was used to provide further evaluation of the three models.

4.1 Model evaluation

Three primary models were studied – the Rescorla-Wagner model, the configural Rescorla-Wagner model, and the Pearce configural model in each of three steps. First, maximum likelihood parameter estimates were obtained for each model and participant. Second, using these parameter estimates, simulations of the experimental design were carried out and the expected (model) responses generated by simulation were compared to the observed (participant) responses. Third, models were compared using Akaike weight analysis to determine the best model overall and in order to assess the best model for each participant (Burnham & Anderson, 2002; Cavagnaro et al., 2016; Farrell & Lewandowsky, 2018; Wagenmakers & Farrell, 2004).

4.1.1 The Rescorla-Wagner model

The canonical form of the Rescorla-Wagner model is given in Equation (7) (Rescorla & Wagner, 1972). In (7) ΔV_{ijk} is the change in the associative strength (*V*) that occurs on trial *i* between cue *j* e.g. one of the foods eaten by the FULF on that trial (labelled *A . . . E, G, K, L* in Table 29) and outcome of that trial. ΔV is a function of two learning rate parameters, α a learning rate for cues and β a learning rate for outcomes, and the parenthesised error term. In the error term, λ_k represents the outcome of the trial and takes the value of 1 or 0 for the occurrence and non-occurrence of an outcome, respectively. $\sum V_{ijk}$ is the associative strength for outcome *k* summed over the *n* cues present on the trial.

$$\Delta V_{ijk} = \alpha\beta(\lambda_k - \sum_{j=1}^{n} V_{ijk}) \tag{7}$$

The Rescorla-Wagner model was implemented with two values of α, α_{ctx} and α_{cue}, to allow different learning rates for different categories of cue. The diffuse context cues provided by the screen background that were stable within different phases of the experiment were allowed to have different α learning rate than the discrete food cues which changed from trial to trial. Two values of β, β_{us} and $\beta_{\sim us}$ were used, to allow for the possibility that learning rate may differ on reinforced and non-reinforced trials.

4.1.2 The configural Rescorla-Wagner model

The configural Rescorla-Wagner model was implemented in the same way as Equation (7) except an additional class of cue was introduced to represented stimulus configurations. In the Rescorla-Wagner model cues are considered ‘standalone’ elements representing the intrinsic physical properties of a stimulus. However, this is generally believed to be an oversimplification with evidence indicating that configural cues can be produced when multiple stimuli occur together (e.g. Rescorla, 1973; Wagner & Rescorla, 1972 ;Woodbury, 1943). In the current implementation of the configural Rescorla-Wagner model a unique configural cue was coded to represent each pairwise cue combination. For example the cues on a trial involving the presentation of cue *A* in context *A:* would be coded *aAw*, where *a* is context *A:*, *A* is cue *A*, and *w* is the configural cue generated by the conjunction of *A* and *A:*. For an *AB* compound presented in context *B*: the encoding would be *bABxyz*. Here the configural cues are *x*, *y*, and *z* representing the pairwise cue combinations as follows: *bA* → *x*, *bB* → *y*, and *AB* → *z*. The configural Rescorla-Wagner model therefore has one more parameter than the Rescorla-Wagner model, an additional learning rate parameter

α_{cfg} allowing different learning rates now for three categories of cue (context cues, discrete cues, and configural cues).

4.1.3 The Pearce configural model

Pearce (1994) developed a widely cited configural model of associative learning which, despite the common moniker ‘configural’, operates on quite different principles than the configural Rescorla-Wagner model. The main difference between these models is in the way in which the cues are processed. In the Rescorla-Wagner model and the configural Rescorla-Wagner model each cue enters into individual associations with the outcomes. In contrast, in the Pearce configural model, cues are grouped into configurations and a configuration is formed by each unique pattern of cues encountered during learning and the configurations, rather than cues, are the units which enter into associations with the outcomes. For example, referring again to design Table 29, during the acquisition phase a configural unit *aA* would be used to represent the stimulus pattern when cue *A* was encountered in context *A*: and in the extinction phase a configural unit *bAB* would represent the cue compound *AB* presented in context *B*:.

In Equation (8) $\Delta V_{c_{ik}}$ is the change in the associative strength between the configuration present on that trial (c_i) and the trial outcome. Equation (8) is of the same form as the Rescorla-Wagner model but the error term is computed as the difference between λ_k and a weighted sum of the associative strengths of all the stimulus configurations known to the system. The weights are provided by the similarities between c_i and each of the *n* configurations in the system with the similarity between any two configurations *a* and *b* given as a function of the number of cues common to both configurations, n_{ab}, and the number of cues in each configuration, n_a and n_b, as shown in Equation (9). In Equation (9) *d* is a discrimination sensitivity parameter with larger values reducing the similarity and therefore increasing discrimination between configurations (Kinder & Lachnit, 2003).

$$\Delta V_{c_{ik}} = \alpha\beta(\lambda_k - \sum_{j=1}^{n} S(c_i, c_j)V_{c_{jk}}) \tag{8}$$

$$S(a,b) = \left(\frac{n_{ab}}{\sqrt{n_a}\sqrt{n_b}}\right)^d \tag{9}$$

4.2 Parameter estimation

The maximum likelihood parameters were estimated using R code written by Dr. Steve Glautier (code available with data at: https://osf.io/p59zu/) which was run in R version 4.0.3 using Nelder-Mead optimisation via package optimx version 2022-4.30 (Nash & Varadhan, 2011; R Core Development Team, 2020). The optimisations found, for each participant and model, a parameter vector for that model, ***θ***, which minimised ***L*** over the n = 178 trials of the experiment as shown in (10): models used one step lookahead, making probabilistic predictions for responses on trial n on the basis of what had been learned up to and including trial n – 1.

$$L = -\sum_{i=1}^{n} \ln P(R_i) \tag{10}$$

$P(R_i)$ was the model probability for the observed response on trial *i*. Three possible responses were available to participants on each trial – they could predict outcome X, outcome Y, or outcome Z and $P(R_i)$ was a softmax function of the associative strengths of the cues present on trial *i* and a sensitivity parameter g as shown in Equation (11)(c.f. Ahn et al., 2008; Wikipedia, 2020; Yechiam & Busemeyer, 2005).

$$P(R_i) = \frac{exp(gV_{ir})}{\sum_{k=x}^{z} exp(gV_{ik})} \tag{11}$$

V_{ir} in the numerator of (11) is the associative strength for the outcome corresponding to the observed response summed over all cues present on the trial and the denominator includes the associative strength summed over all outcomes and all cues present on the trial. When g → 0 (11) results in guessing behaviour with the response probabilities approaching $\frac{1}{n}$ where n is the number of response options (n = 3 in this case). When g → inf (11) results in maximisation with the probability of the response for which the associative strength of the cues present on that trial is highest approaching 1.

The optimisations included some constraints on the parameter values in order to provide numerical stability and in order to preserve the psychological sense of the parameters in the current modelling context (e.g. although some analyses have suggested a modification of the Rescorla-Wagner model which allows negative learning rates (Dickinson & Burke, 1996; Van Hamme & Wasserman, 1994) these were not used here). All learning rates were constrained to the range [0.0001 . . . 0.75], *g* was constrained to the range [0.0001 . . . 15], and *d* was constrained to the range [0.0001 . . . 20]. In addition all optimisations were run with three initial values of θ. One value came from an initial exploratory optimisation, one value consisted of all parameters set to 0.1 except for g which was set to 2, and the third initial value vector was set to a selection of random values.

4.3 Model evaluation

Average maximum likelihood parameter estimates and *L* values are shown in **Table 33**, **Table 34**, and **Table 35** for the Rescorla-Wagner model, the configural Rescorla-Wagner model, and the Pearce configural model respectively for each experimental condition and overall.

As can be seen in **Table 33Table 35** the average *L* values were in the range 74 . . . 90, indicating that the average model probabilities for the observed responses were in the range 0.66 . . . 0.6. The average *L* values were larger in the deepened extinction group which could be due to the second stage extinction procedure used for this group. When this group went from extinction stage 1 into extinction stage 2 a second cue was added which could inflate the average *L* values, as participants in practice only seemed to react to the new cue for one trial, after which performance reverted back (Figure 25).

Table 33

Mean Maximum Likelihood Parameters and *L* for Rescorla-Wagner Model (standard error).

group	*L*	α_{ctx}	α_{cue}	β_{us}	$\beta_{\sim us}$	*g*
c	93.232 (4.6)	0.214 (0.024)	0.463 (0.026)	0.415 (0.026)	0.299 (0.028)	6.051 (0.398)
de	95.065 (4.289)	0.248 (0.035)	0.492 (0.026)	0.336 (0.028)	0.319 (0.034)	6.05 (0.449)
se	88.324 (3.075)	0.132 (0.019)	0.521 (0.02)	0.365 (0.023)	0.426 (0.026)	5.231 (0.301)
all	91.767 (2.308)	0.19 (0.015)	0.493 (0.014)	0.376 (0.015)	0.354 (0.017)	5.73 (0.217)

Table 34

Mean Maximum Likelihood Parameters and L for Configural Rescorla-Wagner model (standard error).

group	L	α_{ctx}	α_{cue}	β_{us}	$\beta_{\sim us}$	α_{cfg}	g
c	90.655	0.226	0.254	0.287	0.204	0.277	7.014
	(4.634)	(0.026)	(0.023)	(0.025)	(0.025)	(0.029)	(0.382)
de	92.519	0.21	0.292	0.305	0.224	0.272	6.64
	(4.398)	(0.035)	(0.03)	(0.034)	(0.034)	(0.032)	(0.483)
se	84.323	0.128	0.227	0.261	0.262	0.381	6.787
	(3.189)	(0.017)	(0.019)	(0.022)	(0.03)	(0.027)	(0.349)
all	88.641	0.184	0.253	0.281	0.231	0.316	6.833
	(2.357)	(0.015)	(0.013)	(0.015)	(0.017)	(0.017)	(0.227)

Table 35

Mean Maximum Likelihood Parameters and *L* for Pearce configural model (standard error).

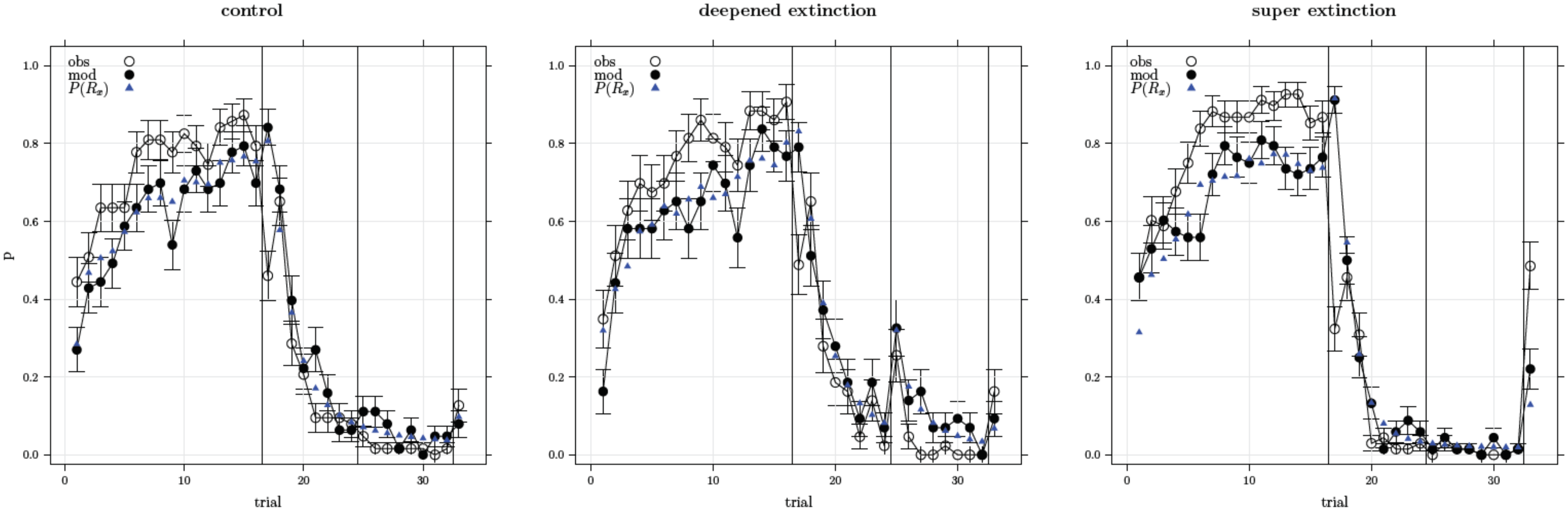
group	L	α_{pat}	β_{us}	$\beta_{\sim us}$	d	g
c	91.159	0.531	0.482	0.309	2.069	6.933
	(4.522)	(0.024)	(0.03)	(0.033)	(0.15)	(0.469)
de	93.16	0.578	0.427	0.261	2.564	6.391
	(4.433)	(0.028)	(0.032)	(0.034)	(0.456)	(0.513)
se	83.962	0.553	0.482	0.316	2.57	6.111
	(3.237)	(0.021)	(0.026)	(0.029)	(0.16)	(0.369)
all	88.841	0.551	0.468	0.3	2.387	6.478
	(2.347)	(0.014)	(0.017)	(0.018)	(0.14)	(0.256)

4.3.1 Simulations

Simulations of the experimental design shown in **Table 29** were carried out for each model and participant using maximum likelihood parameters. Figure 29-Figure **31** show the observed responses for each experimental condition and model alongside the model predicted responses. Data are shown for trials with cue A present and for outcome X responses. Participant responses were coded 1 if an outcome X response was observed and 0 otherwise and the plotted data is averaged across participants. The model predicted responses were generated from random Bernoulli deviates obtained for each trial and participant (1 coding the model predicting an X response and 0 otherwise) with the distribution for each trial parameterised by $P(R_x)$ for that trial with plotted data showing the model predicted responses averaged across participants. The simple Rescorla-Wagner model predictions for the control and deepened extinction were relatively accurate, however the recovery test predictions for the super-extinction group were not, the model predicting significantly less recovery than the observed levels. The configural Rescorla-Wagner model had better predictions for the recovery test of the super-extinction group, but the predictions for the control and deepened extinction seemed worse compared to the traditional model. The predictions of the Pearce configural model were very similar to the predictions of the configural Rescorla-Wagner model.

Figure 29

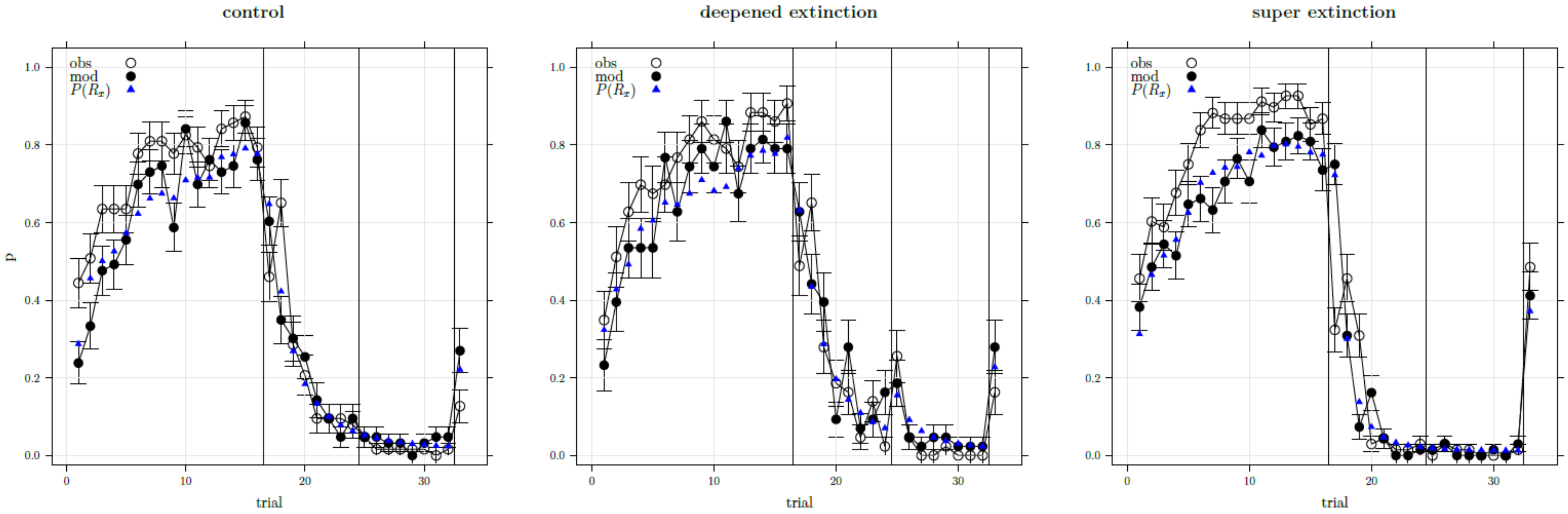
Average proportion of x-responses observed and expected for the Rescorla-Wagner model



Note. Simulations used maximum likelihood parameters on trials involving cue A by experimental condition (± 1 s.e.). Vertical lines separate acquisition, extinction 1, extinction 2, and recovery test phases. $P(R_x)$ is the average probability of an x-response used to parameterise the binomial distribution for generating random deviates for the model responses.

Figure 30

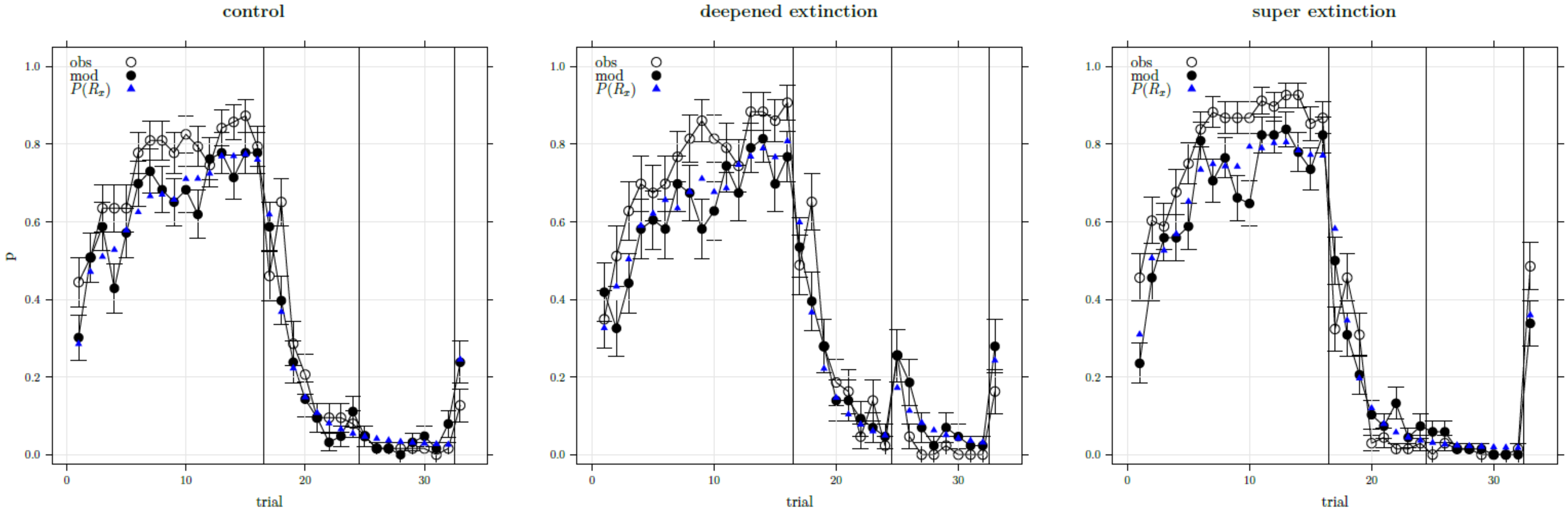
Average proportion of x-responses observed and expected for the configural Rescorla-Wagner model



Note. Simulations used maximum likelihood parameters on trials involving cue A by experimental condition (± 1 s.e.). Vertical lines separate acquisition, extinction 1, extinction 2, and recovery test phases. $P(R_x)$ is the average probability of an x-response used to parameterise the binomial distribution for generating random deviates for the model responses.

Figure 31

Average proportion of x-responses observed and expected for the Pearce configural model



Note. Simulations used maximum likelihood parameters on trials involving cue A by experimental condition (± 1 s.e.). Vertical lines separate acquisition, extinction 1, extinction 2, and recovery test phases. $P(R_x)$ is the average probability of an x-response used to parameterise the binomial distribution for generating random deviates for the model responses.

4.3.2 Akaike weight analysis

Table 36 provides the results of overall Akaike weight analyses. Each of the models discussed above was evaluated in addition to a simple baseline guessing model in which it was assumed that for all trials and participants $P(R_x) = P(R_y) = P(R_z) = \frac{1}{3}$. The finite sample correction form of Akaike's Information Criterion (AIC_c) as given in (12) was used. In (12) V is the number of parameters and *n* is the number of data points over which L was computed.

$$AIC_C = 2L + 2V + \frac{2V(V+1)}{n - V - 1} \tag{12}$$

Table 36

Overall Akaike weight analyses using corrected AIC

Model	Parameters	$2L$	AIC_c	ΔAIC_c	$wAIC_c$
guessing	0	68434.8	68434.8	35728.2	< 0.000001
Rescorla-Wagner	5	31934.9	33724.9	1018.3	< 0.000001
configural Rescorla-Wagner	6	30847.2	33007.7	301	< 0.000001
configural model	5	30916.5	32706.6	0	→ 1

Note. The column 'Parameters' gives the number of parameters estimated for each participant for each model. There were 174 participants so therefore, for example, the number of parameters estimated for $L_{\text{Rescorla–Wagner}}$ was 5 × 174 = 870. L computed over 179 trials for each of 174 participants – i.e. over 31146 data points.

The best model has the lowest AIC_c value, and the column ΔAIC_c in **Table 36** provides the AIC_c difference between the best model, the Pearce configural model, and each model listed. AIC_c > 10 indicates that a model has 'essentially no support' in the context of the current data and competing models (Burnham & Anderson, 2002). The probability of each model being the best model in the context of the current data and competing models is given by the Akaike weights ($wAIC_c$) in **Table 36** computed as in Equation (13). In Equation (13) the ΔAIC_c value for each model i is normalised by dividing by the ΔAIC_c values summed over the *K* models.

$$w_i AIC_C = \frac{exp\left(\frac{-1}{2}\Delta_i AIC_c\right)}{\sum_{k=1}^{K} exp\left(\frac{-1}{2}\Delta_k AIC_C\right)}$$

(13)

Based on **Table 36**, the average L was smaller for the configural Rescorla-Wagner than for the Pearce configural model, however the latter was the overall better model with the smallest AIC_c, this is because the Pearce configural model had one less parameter, managing therefore to predict behaviour almost as well as the configural Rescorla-Wagner using a more parsimonious method. In contrast, as previously mentioned the Rescorla-Wagner model predictions for the super-extinction group recovery test were highly inaccurate based on participants observed responses. Although the model performed better that a guessing model, it performed worst compared to the other two models.

Although the Pearce configural model was the best model overall, it was not the best model for every individual. AIC_c and $wAIC_c$ values were computed for each participant and it was found that the Pearce configural model was the best model in 107 cases, with 40, and 27 cases best fit by the configural Rescorla-Wagner model and by the Rescorla-Wagner model, respectively. The methodology set by Cavagnaro et al. (2016) was followed to assess the evidence that each of the models could be the best model for all participants. The individual Akaike weights give the probability that each model is best for that individual and therefore the product of the weights across participants gives the joint probability that a model is best for all participants. In addition the ratio of two Akaike weights provides the weight of evidence in favour (or against) of one model versus another. Putting this together Cavagnaro et al. (2016) define the group Akaike Information Criterion (*gAIC*) for model i as in Equation

(14). In(14) the denominator is *wAIC* for the guessing model so the $gAIC_i$ is the weight of evidence in favour of model *i* being best for all participants (*j* = 1 . . . *n*) in comparison to the guessing model.

$$gAIC_i = \prod_{j=1}^{n} \frac{wAIC_{ij}}{wAIC0_j}$$

(14)

Furthermore, the Akaike weights can be used to parameterise a Dirichelet distribution with a parameter for each of the *i* models computed from (15).

$$\alpha_i = 1 + \sum_{j=1}^{n} wAIC_{ij}$$

(15)

Once the distribution parameters are calculated the probability that model *i* will be the best for a randomly chosen participant is given by Equation (16). Equation (16) sums over the *m* models to normalise α_i.

$$P(best_i) = \alpha_i \left(\sum_{i=1}^{m} \alpha_i \right)^{-1}$$

(16)

Table 37 provides the results of the analyses described above.

Table 37

Comparison of models on group AIC and probability of each model being the best model for a randomly chosen participant.

Model	log *gAIC*	P(best)
guessing	0	0.006
Rescorla-Wagner	17349.8	0.16
configural Rescorla-Wagner	17707.3	0.28
Pearce configural model	17858.9	0.554

In summary, three formal models of associative learning: the Rescorla-Wagner, the configural Rescorla-Wagner, and the Pearce configural models were assessed on their ability to predict learning, extinction, and recovery in an ABC design where extinction was carried out using cue alone extinction, super-extinction, and deepened extinction independently. Out of the three models the Rescorla-Wagner model performed worst, being inaccurate in the amount of recovery predicted for the super-extinction group. This model predicted more recovery for the super-extinction group compared to the control and deepened extinction groups, however the recovery observed in the data was significantly higher than the levels predicted (Figure 29). The configural Rescorla-Wagner model and Pearce configural model performed similarly in terms of their predictions, with the predictions of the former being slightly better (**Table 36**). Using an Akaike weight analysis, the Pearce configural model was found to be the overall best model out of the three as it had one less parameter than the configural Rescorla-Wagner model (**Table 36**). A further in depth analysis revealed that although the Pearce configural model was the best model overall, out of the 174 participants it was the best for 107, while the configural Rescorla-Wagner model was the best model for 40, and the Rescorla-Wagner model was the best for 27. These results show that although fairly accurate predictions can be made using some of the most widely acknowledged models of associative learning, these models do not hold perfect predictions due to what seems to be some underlying individual differences. These individual differences were then showcased in the final part of the analysis where the best overall model was found not to be the best model for each participant.

Chapter 5 Discussion

The current thesis set to address the following three main aims. First it was intended to assess the existence of a potential link between associative and non-associative inhibition. Associative learning was defined on the basis of conditioned inhibition, and extinction, more specifically the speed of learning along with performance in summation tests were used as measures of associative inhibition. For non-associative inhibition, four measures were consistently used which map onto the substructures of inhibition proposed by Bari and Robbins (2013): BIS11, and BIS/BAS for cognitive inhibition, monetary choice task for delayed discounting, and the stop signal reaction task for response inhibition.

Second, it was aimed to assess the effectiveness of compound extinction compared to cue alone extinction. Across two studies compound extinction was defined as super-extinction first and then as either super-extinction or deepened extinction. Using an ABC design the differences in extinction acquisition speed between the three groups were assessed. Additionally, the differences in context inhibition developed to the extinction context, and the differences in recovery observed when the target cue was tested outside the extinction context were also examined.

Last, the predictions of three formal associative learning models: Rescorla-Wagner, configural Rescorla-Wagner, and Pearce configural model, were compared with the aim of identifying the best model which made the most accurate predictions when compared with the observed data. For this purpose the last super extinction study data was used and models were compared using an Akaike weight analysis.

5.1 Associative and Non-associative Inhibition

In the field of associative learning, inhibition is a construct of particular importance in the context of changing behaviour and adapting to a dynamic environment. Conditioned inhibition is one of the most obvious associative process that comes to mind when considering inhibition. Conditioned inhibition allows an organism to change its normally expected behaviour due to the conditioned inhibitor which signals the absence of an otherwise expected outcome. In the wider field of Psychology, inhibition was defined in a very similar manner where inhibition is a construct that employs a set of mechanisms through which certain processes are stopped (inhibited). Given the large variety of processes that can be inhibited, a large variety of inhibition phenomena have been defined and studied mainly in isolation, and mainly through the lens of impulsivity. While it is widely accepted that inhibition/impulsivity is a multidimensional construct, an agreed upon structure for this construct was not defined. Bari and Robbins (2013) proposed an underlying structure consisting of two main factors: cognitive and behavioural inhibition, on the basis that the two were repeatedly shown to be uncorrelated in the literature (e.g. Broos et al., 2012; Reynolds et al., 2006), with behavioural inhibition assumed to consist of delayed gratification, response inhibition, and reversal learning. Despite the focus on the relationship between various non-associative inhibition constructs/measures, associative inhibition was rarely considered in the formulation of models or tested alongside other inhibition measures. To date, only a few studies explored the relationship between associative and non-associative inhibition and the results reported were inconsistent, therefore it is still unclear if the two types of inhibition share a common source or whether they are independent factors.

He et al. (2011) conducted an indirect investigation into this relationship using a group of control participants from the general population and a group of participants who had a history of offending and who were also characterised as having impulsive behaviour. The latter group was further divided into participants who fit the criteria for personality disorder or dangerous and severe personality disorder. The investigation revealed that when asked to take part in a learning task, the control group showed a conditioned inhibition effect in a summation test while the group with a history of offending showed weak or no conditioned inhibition. This difference was more extreme in participants with dangerous and severe personality disorder. This was an early indirect indication of a link between associative and non-associative inhibition, and in a later follow-up study He et al. (2013) investigated the same effect using a sample of university students. As part of this study non-associative inhibition was measured using the BIS/BAS scale, and the results revealed a negative relationship between BIS and conditioned inhibition. Migo et al. (2006) also examined the relationship between associative inhibition and non-associative inhibition using a conditioned inhibition task and the BIS/BAS scale. They reported a positive correlation between conditioned inhibition and the BAS-reward subscale of the BIS/BAS. Together these results highlight the fact that the relationship between the associative and non-associative inhibition is not fully understood. It could be the case that the two are independent inhibition subtypes, but it could also be the case that some degree of similarity exists and this is at least partly captured by the BIS/BAS scale.

The current thesis reported a series of two studies that aimed to investigate this link by using a conditioned inhibition learning task and four measures of non-associative inhibition in the form of: BIS11, BIS/BAS, delayed discounting, and response inhibition. For the first study the learning task was modified to contain features of the stop signal task on account of the similarities between the two tasks. It was hypothesised that by including overall time pressure and a delay to the presentation of the feature negative discrimination a potential relationship between conditioned inhibition and response inhibition would be easier to detect. Participants learnt a feature negative discrimination during training, and conditioned inhibition was assessed using a summation test. Based on the summation test performance participants were classified as inhibitors or occasion setters. The feature negative discrimination learning speed and the classification into inhibitors and occasion setters were taken as measures of conditioned inhibition.

Despite the a priori expectations no relationship was found between conditioned inhibition and the four measures of non-associative inhibition with the partial exception of BIS. BIS was found to be a significant predictor of the feature negative discrimination acquisition, participants who had high BIS learnt the feature negative discrimination faster. In this model a group interaction was also included and the test was repeated for the intercept which reflected the terminal performance at the end of training. All remaining effects involving BIS were not significant, but were approaching significance ($p < .07$), therefore this could be interpreted as a potential indication of a relationship between conditioned inhibition and BIS. Another significant interaction was found between BIS and group on the classification into inhibitors and occasion setters. According to this relationship participants in the no delay group with high BIS scores were more likely to be inhibitors, while the reverse was true for the delay group. These results could be interpreted as an indication of a relationship between conditioned inhibition and BIS, however the learning task used failed to show a clear conditioned inhibition effect casting doubt on the results. Consequently, the learning task was updated, the new design tested in a pilot study and the initial study was repeated using the updated learning task.

The second conditioned inhibition study followed the same methodology as the first study with the only exception being the design of the new learning task. The learning task was updated following the pilot study and a scoping review of other studies that trained conditioned inhibition. The task consisted of two summation tests, one predictive and one evaluative, and had both a novel and neutral control. As a result, based on the evaluative summation test, participants were classified into inhibitors and occasion setters and each were assigned an inhibition score based on the evaluative summation test. Additionally, two control cues were included in both the summation tests so the classification and inhibition scores were computed twice. Once again no significant relationships were found between associative and non-associative inhibition with two exceptions. The first exception was the relationship between BIS and the overall level of conditioned inhibition showed in the evaluative summation test. Similarly to the initial study this relationship indicated that participants with high BIS scores showed more conditioned inhibition, however this was not replicated in the predictive summation test. The second exception was the stop signal reaction time, a significant interaction between the inhibition classification and SSRT on the feature negative discrimination intercepts which were reflective of terminal performance at the end of acquisition. This relationship was however not replicated across both classifications and summation tests.

The results of the first series of experiments suggest there might be a link between conditioned inhibition and BIS, as BIS was found to be a significant predictor of conditioned inhibition on multiple occasions across the two studies suggesting that participants who had higher BIS scores showed better conditioned inhibition in the learning task. These results contradict the existing literature, Migo et al. (2006) found no relationship between conditioned inhibition and BIS, while He et al. (2013) found such a relationship but in the opposite direction. Based on the relative questionable robustness of the first study, the lack of consistency in the second, and the contradictory effects reported in the literature it would be too early to draw confident conclusions regarding the relationship between BIS and conditioned inhibition, however it can be concluded that together these results suggest the existence of a relationship that is not yet fully understood. Based on the remaining measures of non-associative inhibition, it can be concluded that associative and non-associative inhibition are independent, therefore associative inhibition should be included and considered a standalone factor when developing future models of inhibition.

There are several important implications arising from the results of the first series of experiments, the first referring to the overall structure of inhibition and the resulting behaviour: impulsivity. The general lack of statistically significant associations between the measures of inhibition supports the hypothesis that inhibition is a multidimensional construct comprising multiple independent factors. The results suggest that one of these independent factors could be associative inhibition, which is rarely, if ever considered when discussing inhibition and impulsivity. Both associative and non-associative learning were separately found to be linked to disorders such as ADHD, schizophrenia, and substance abuse (Bauer, 2001; Enticott et al., 2008; Fillmore & Rush, 2006; Fillmore & Rush, 2002; Hoptman et al., 2002; Porter et al., 2011; Schachar et al., 1993). As a result, understanding the concept of inhibition as a whole and the way in which the independent underlying components come together to influence behaviour is of vital importance. Future research should focus on exactly that, the interaction between the independent sub-factors of inhibition and their effect on clinical populations. Although the current results indicate a lack of association in the general population, it cannot be concluded that the sub-factors do not interact in other populations.

5.2 Extinction

Extinction represents one way in which a previously learnt association can be changed. Extinction relies on the non-reinforcement of a previously established association which results in the weakening of the relationship to the point it appear that the association no longer exists. Phenomena such as recovery and renewal demonstrate however, an association that went through extinction is not completely destroyed. Understanding extinction is of particular interest in the field of clinical Psychology where associative learning is used to understand maladaptive learnt behaviours such as substance abuse. The development of addiction can be understood by following simple associative learning principles which state that constant pairing of substance abuse with a desired internal state leads to automatic substance consumption (Everitt & Robbins, 2016). In the same way the development of addiction can be understood through associative learning, it is hoped that treatment solution can be developed by studying extinctions and the factors influencing it. Cue-exposure therapy aims to do just that, by using extinction principles it aims to extinguish the maladaptive behaviour such as addiction by non-reinforcing substance related cues in an attempt to reduce the risk of relapse. The previously mentioned renewal and recovery effects might however reduce the effectiveness of this procedure as extinctions appears to be highly susceptible to contextual changes (Conklin & Tiffany, 2002). One approach to counter renewal is to train extinction in multiple contexts which was shown to be effective in multiple studies (e.g. Glautier et al., 2013). An alternative approach which was studied as part of the current thesis is multiple cue extinction.

A series of experiments which focused on assessing the effectiveness of compound extinction was reported above. In extinction study one, two groups which received identical acquisition training were used. For the control/cue alone extinction group acquisition was followed by the extinction of a target cue by itself across a series of trials, for the super extinction group the target cue was extinguished in compound with another cue which received reinforced training during acquisition. The study had an ABC design, therefore acquisition and extinction took place in different contexts. To assess the level of inhibition gained by the extinction context as a result of the extinction procedure, a summation (context inhibition) test was carried out in this third novel context. The main aim of the study was to compare the two groups based on the speed of extinction, context inhibition, and recovery. The study had a secondary aim which was carried over from the first series of studies, and that was to once again assess the relationship between associative and non-associative inhibition. Associative inhibition was defined as the speed of extinction as inhibition is required to stop responding, and the level of context inhibition acquired by the context as a result of the extinction, given that the context could become a conditioned inhibitor following the design used.

The results of the study revealed that there were no differences between the two groups when considering the extinction rates, context inhibition, or recovery. The results also revealed no link between associative and non-associative inhibition with the exception of BIS which was found to be a significant predictor of extinction acquisition rates for the control group. According to this effect participants who scored higher on BIS had less steep slopes meaning that they have acquired extinction slower, however this effect was not replicated in the super-extinction group.

According to the Rescorla-Wagner model a summation effect was expected during the first trial of extinction along with a faster extinction and less recovery for the super-extinction group. The results only partially supported these predictions, extinction was found to be faster for this group, however no summation was observed in the first trial of extinction, and this group showed more recovery than the control group. The results of the study were not entirely unexpected as these can be explained by the Pearce configural model which predicts none of the above-mentioned effects. Additionally, within the literature conflicting results have been reported with some studies showing a faster extinction and less recovery for super-extinction, while others didn't (e.g. Griffiths et al., 2017; Leung et al., 2012; Rescorla, 2000). One possible explanation for a lack of summation is a ceiling effect as both cues reached an asymptotic level towards the end of training, meaning that participants could not show more expectation for the compound to be reinforced compared to a cue alone. Correspondingly, it could not be determined with certainty whether a summation effect occurred or not. As a result the design of the learning task was updated and tested using a pilot study.

To ensure that a summation test could be observed the target cues were updated to be reinforced with a rate of 75% in order to avoid a ceiling effect, therefore when the two partially reinforced cues were combined in compound a summation test could be observed and the expectation for the compound to be reinforced could go above the expectation for the target cue alone. An additivity demonstration was also included during training in order to maximize the likelihood of observing a summation effect. Finally, a deepened extinction group was added and the extinction phase was split into two stages. No changes were made to the control and super-extinction groups, the two extinction stages being identical within these groups. For the deepened extinction group, the first extinction stage cue alone extinction was used, and for the second stage the target cue was paired with a cue that received reinforced training during acquisition.

The second extinction study followed the same methodology as extinction study 1 with the addition of a new group, a deepened extinction group. For the comparison of extinction rates the first stage of extinction was used and the control and deepened extinction groups were aggregated together as they received identical extinction during the first stage. The results revealed that the super extinction group did acquire extinction faster compared to the aggregated control and deepened extinction group, however no summation effect was observed. The three groups did not show different levels of conditioned inhibition, but the super-extinction group showed significantly more recovery compared to the control and deepened extinction groups. The latter groups did not significantly differ in the levels of recovery observed. These results contradict the findings of Rescorla (2000, 2006) who showed that super-extinction and deepened extinction was less prone to recovery compared to cue alone extinction. The same pattern of results was reported by Griffiths et al. (2017) who found that super-extinction resulted in more recovery than cue alone and that deepened extinction was not different from cue alone in terms or recovery. The current results suggest that compound extinction does not lead to a longer lasting/more stable extinction compared to cue alone extinction, despite the super-extinction showing faster extinction, more recovery was observed in this group.

Together these results are of significance for addiction, and cue exposure therapy more specifically as this type of treatment aims to use extinction principles to provide patients with robust mechanisms to tackle addiction. Although in theory, as suggested by the Rescorla-Wagner model, compound extinction should lead to a faster more stable extinction, according to the results reported in the current thesis this is not entirely the case. A faster extinction was reported for super-extinction, however this group also showed the most recovery. For cue exposure therapy the levels of recovery are critical, while the speed of extinction is less important since the main goal is for the mechanisms developed as part of the treatment to be robust and long-lasting. Based on the current results the most obvious conclusion would be that cue alone extinction is the best form of treatment that should be used to model cue-exposure therapy, however the current experiments used a sample of the general population and learning tasks where the cues and outcomes were of little significance. In reality cue-exposure therapy uses substance related cues which evoke a strong response for the participant. Future research should focus on validating the current results using a variety of cues and outcomes which carry varying degree of importance for the participants. Similarly, as previously mentioned the current set of experiments used a sample of the general population, whose learning, and inhibitory mechanisms might be different that a clinical populations'. Because the most important implications of the current experiments are clinical it should be aimed for the results to be validated with different populations, including clinical.

5.3 Formal associative learning models

The final aim of the current thesis was to evaluate the predictions of three formal models of associative learning: Rescorla-Wagner, configural Rescola-Wagner, and Pearce configural model for the final extinction study where three extinction techniques were used. Prior to the data collection and analysis of this study assumptions were made based on the above mentioned models, many of which were contradictory. To better understand the data and find the best model predictions the three models were compared to each other as described below.

First, for each participant and each model maximum likelihood parameters were estimated, there were a total of 174 participants who responded to 178 trials. Next, simulations based on the experimental design and the maximum likelihood parameters were carried out. These simulations produced expected responses which were compared to the observed data from the study. In order to compare the models with each other an Akaike weight analysis was carried out and the best overall model was chosen, additionally it was also determined for each participant individually which model made the best predictions.

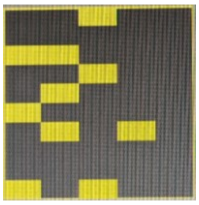
Upon an initial comparisons of the expected versus observed data for each model within each condition it was noted that the Rescorla -Wagner model performed worst, compared to configural Rescorla -Wagner and Pearce configural model. The latter two performed relatively similarly having quite accurate predictions, however the configural Rescorla-Wagner model was found to have prediction that were slightly better aligned with the data. When assessing the Akaike weights the Pearce configural model was concluded to be the best overall model overtaking the configural Rescorla -Wagner model as a result of having less parameters (Pearce configrual model had 5 parameters and the configural Rescorla-Wagner model had 6). Although the best model overall was the Pearce configural model, this was not the best model for every participant, the likelihood that this model would be the best for a randomly chosen participant was 55%. A further in depth exploration revealed that out of the 174 participants the Pearce configural model was the best model for 107 participants, the configural Rescorla-Wagner model was the best model for 40 participants, and the Rescorla-Wagner model was best for 27 participants.

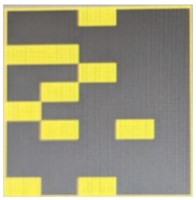
This model comparison shows that although an overall “superior” model can be identified there is quite a large degree of variability within the data, different models being better for certain participants. It could be the case that the current models are not complex enough to account for various types of behaviours and human behaviour might be far too complex to be summarised within a single model. This further highlights the need to understand all factors that influence associative learning with the aim of developing better learning models. The fact that the Pearce configural model was the best overall model to explain the data also sheds light on why the extinction studies did not show a summation effect and why the compound extinction did not produce more stable long lasting extinction as these predictions were made based on the Rescorla-Wagner model.

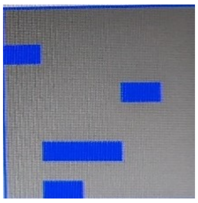
Models of associative learning are meant to provide an informative and accurate representation of how organisms might behave in certain situations, allowing for a priori hypotheses to be formulated. As a result, accuracy and versatility is of upmost importance, and the three models evaluated as part of the current thesis were shown to perform better than a random model. Additionally, a clear “winner” was chosen as part of the analysis based on having the most accurate overall predictions for the sample of participants used. In spite of all this as showcased by the last part of the analysis individual differences play a very important role in behaviour, and these were not adequately captured by any of the models. The current thesis provides a good example of how important these individual differences are. In the first series of experiments participants could be classified into inhibitors or occasion setters after being exposed to the same training. Additionally, this classification was found to be a significant predictor of feature negative discrimination learning highlighting the importance of individual differences in learning. As a result, future models should take these into account to increase accuracy and versatility.

Appendix A

Examples of stimuli and combinations of stimuli used in conditioned inhibition Study 1 and pilot study (Chapter 2).

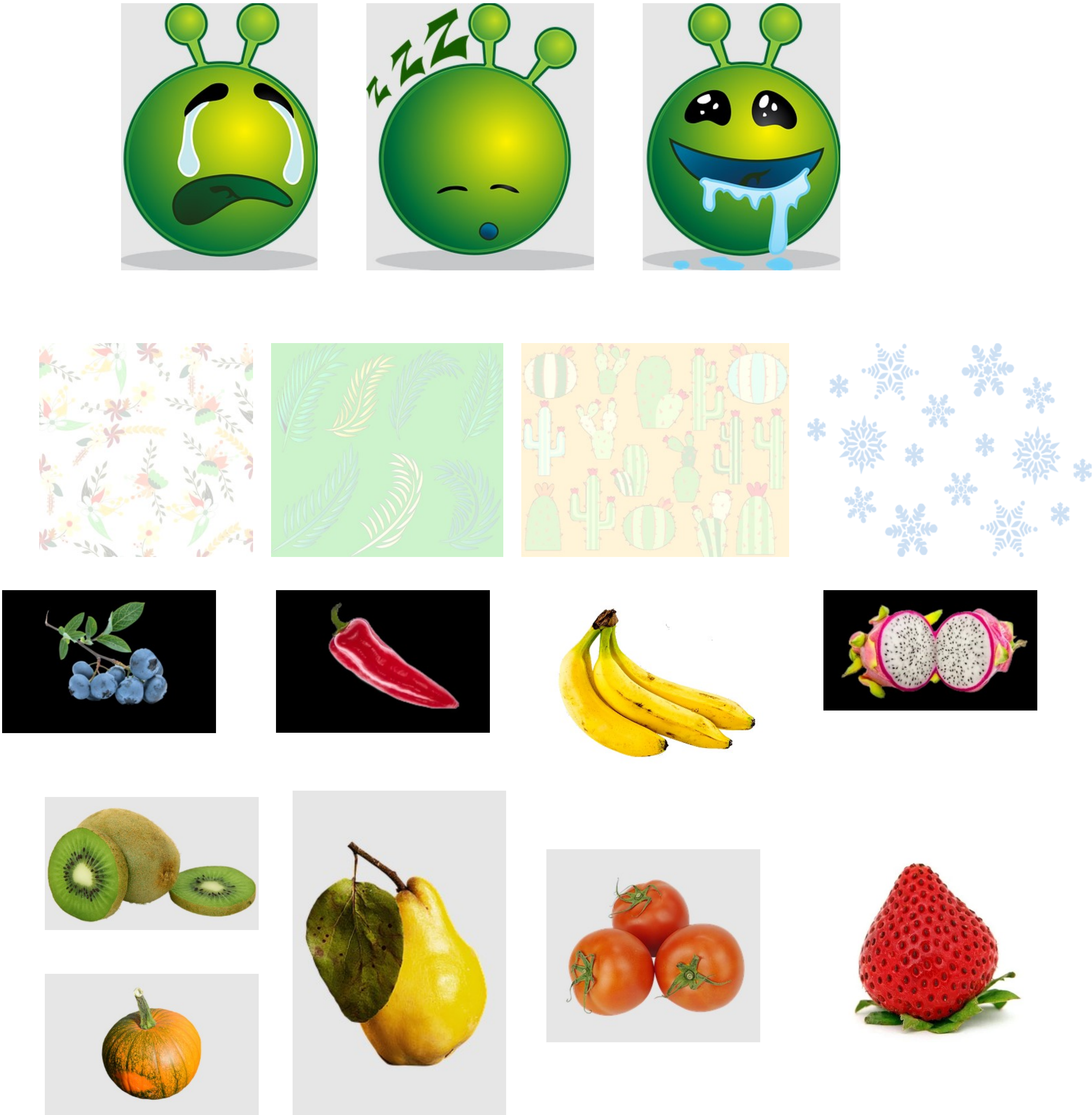






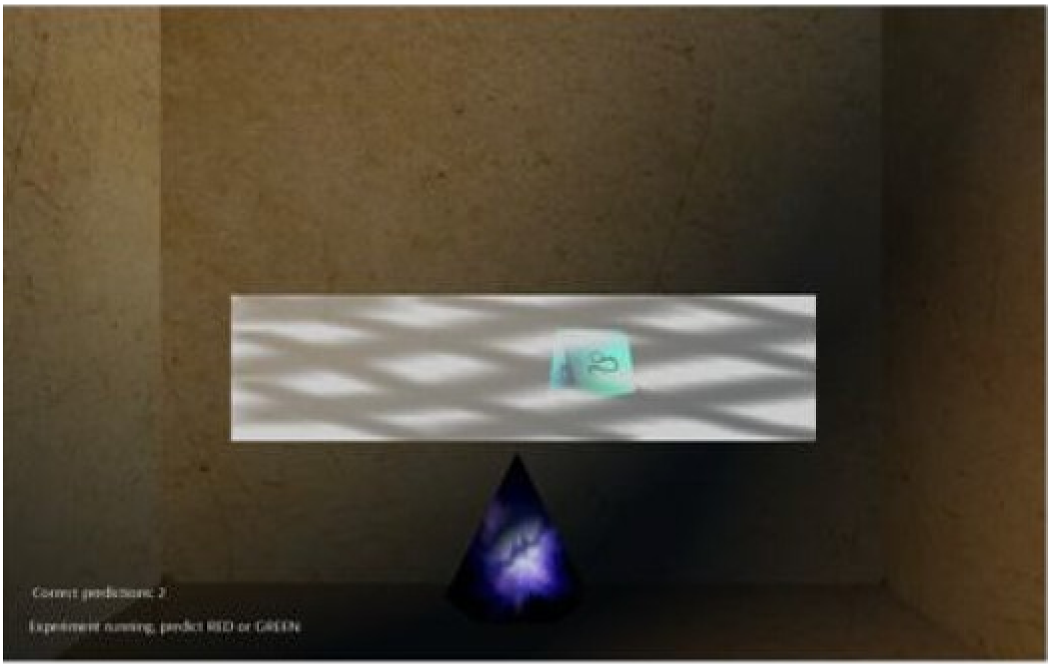
Appendix B

Stimuli used in conditioned inhibition Study 2 and extinction study 2. Contexts and neutral outcome were used in extinction Study 2 only.



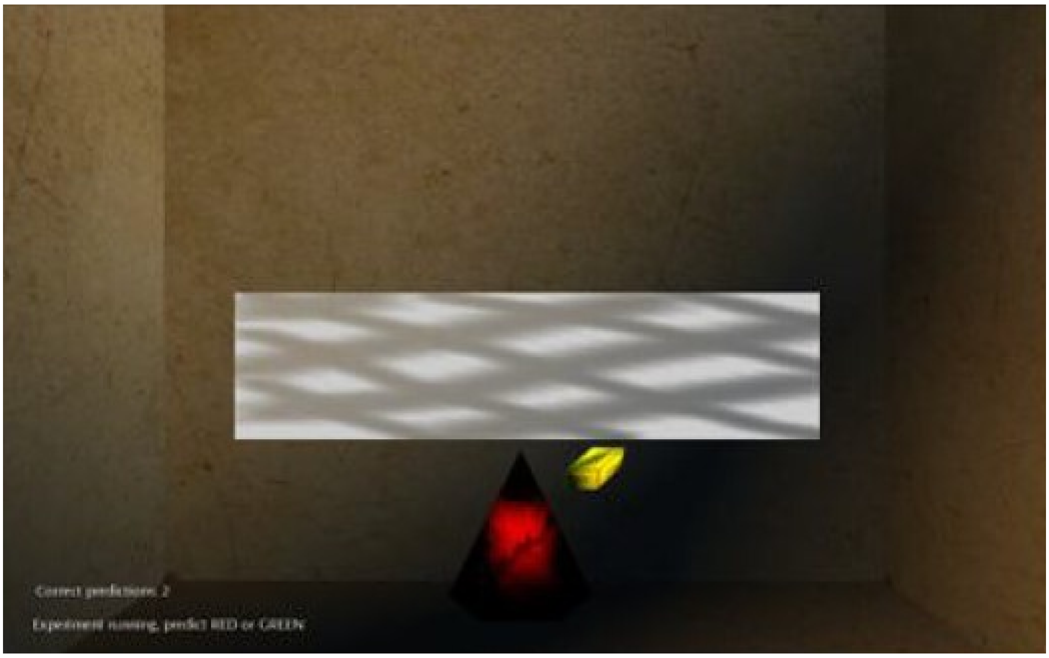
Appendix C

Examples of trials from Extinction Study 1



Correct predictions: 2

Experiment running, predict RED or GREEN



Correct predictions: 2

Experiment running, predict RED or GREEN

List of References

Ahn, W.-Y., Busemeyer, J. R., Wagenmakers, E.-J., & Stout, J. C. (2008). Comparison of Decision Learning Models Using the Generalization Criterion Method. *Cognitive Science*, *32*, 1376–1402. https://doi.org/10.1080/03640210802352992

Aichert, D. S., Wstmann, N. M., Costa, A., MacAre, C., Wenig, J. R., Mller, H. J., … Ettinger, U. (2012). Associations between trait impulsivity and prepotent response inhibition. *Http://Dx.Doi.Org/10.1080/13803395.2012.706261*, *34*(10), 1016–1032. https://doi.org/10.1080/13803395.2012.706261

Ainslie, G. (1975). Psychological Bulletin Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control, *82*(4).

Alarcón, D., & Bonardi, C. (2015). The Effect of Conditioned Inhibition on the Specific Pavlovian-Instrumental Transfer Effect. https://doi.org/10.1037/xan0000087.supp

Amundson, J. C., Wheeler, D. S., & Miller, R. R. (2005). Enhancement of Pavlovian conditioned inhibition achieved by posttraining inflation of the training excitor. *Learning and Motivation*, *36*(3), 331–352. https://doi.org/10.1016/j.lmot.2004.11.006

Anton, R. F., O'Malley, S. S., Ciraulo, D. A., Cisler, R. A., Couper, D., Donovan, D. M., … Zweben, A. (2006). Combined Pharmacotherapies and Behavioral Interventions for Alcohol Dependence: The COMBINE Study: A Randomized Controlled Trial. *JAMA*, *295*(17), 2003–2017. https://doi.org/10.1001/JAMA.295.17.2003

Aron, A. R., & Poldrack, R. A. (2005). The cognitive neuroscience of response inhibition: Relevance for genetic research in attention-deficit/hyperactivity disorder. *Biological Psychiatry*, *57*(11), 1285–1292. https://doi.org/10.1016/j.biopsych.2004.10.026

Baetu, I., & Baker, A. G. (2010). Extinction and blocking of conditioned inhibition in human causal learning. *Learning & Behavior*, *38*(4), 394–407. https://doi.org/10.3758/LB.38.4.394

Bari, A., & Robbins, T. W. (2013). Inhibition and impulsivity: Behavioral and neural basis of response control. *Progress in Neurobiology*, *108*, 44–79. https://doi.org/10.1016/j.pneurobio.2013.06.005

Bari, A., Robbins, T. W., & Dalley, J. W. (2011). Impulsivity. *Neuromethods*, *53*, 379–401. https://doi.org/10.1007/978-1-60761-934-5_14/FIGURES/2_14

Bauer, L. O. (2001). Antisocial personality disorder and cocaine dependence: their effects on behavioral and electroencephalographic measures of time estimation. *Drug and Alcohol Dependence*, *63*(1), 87–95. https://doi.org/10.1016/S0376-8716(00)00195-2

Bickel, W. K., Jones, B. A., Landes, R. D., Christensen, D. R., Jackson, L., & Mancino, M. (2010). Hypothetical Intertemporal Choice and Real Economic Behavior: Delay Discounting Predicts Voucher Redemptions During Contingency-Management Procedures. *Experimental and Clinical Psychopharmacology*, *18*(6), 546. https://doi.org/10.1037/A0021739

Bickel, W. K., & Marsch, L. A. (2001). Toward a behavioral economic understanding of drug dependence: delay discounting processes. *Addiction*, *96*, 73–86. https://doi.org/10.1080/09652140020016978

Bonardi, C., & Hall, G. (1994). Occasion-Setting Training Renders Stimuli More Similar: Acquired Equivalence between the Targets of Feature-Positive Discriminations. *Https://Doi.Org/10.1080/14640749408401348*, *47*(1), 63–81. https://doi.org/10.1080/14640749408401348

Bonardi, C., Robinson, J., & Jennings, D. (2017). Can existing associative principles explain occasion setting? Some old ideas and some new data. *Behavioural Processes*, *137*, 5–18. https://doi.org/10.1016/j.beproc.2016.07.007

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of pavlovian learning. *Psychological Bulletin*, *114*(1), 80–99. https://doi.org/10.1037/0033-2909.114.1.80

Bouton, M. E. (1994). Conditioning, remembering, and forgetting. *Journal of Experimental Psychology: Animal Behavior Processes*, *20*(3), 219–231. https://doi.org/10.1037//0097-7403.20.3.219

Bouton, M. E. (1997). Signals for whether versus when an event will occur. *Learning, Motivation, and Cognition: The Functional Behaviorism of Robert C. Bolles.*, 385–409. https://doi.org/10.1037/10223-019

Bouton, M. E. (2000). A learning theory perspective on lapse, relapse, and the maintenance of behavior change. *Health Psychology*, *19*(1, Suppl), 57–63. https://doi.org/10.1037/0278-6133.19.suppl1.57

Bouton, M. E., & Bolles, R. C. (1979). Contextual control of the extinction of conditioned fear. *Learning and Motivation*, *10*(4), 445–466. https://doi.org/10.1016/0023-9690(79)90057-2

Bouton, M. E., & Nelson, J. B. (1994). Context-Specificity of Target Versus Feature Inhibition in a Feature-Negative Discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, *20*(1), 51–65.

Bouton, M. E., & Swartzentruber, D. (1986). Analysis of the Associative and Occasion-Setting Properties of Contexts Participating in a Pavlovian Discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, *12*(4), 333–350. https://doi.org/10.1037/0097-7403.12.4.333

Brooks, D. C., & Bouton, M. E. (1993). A Retrieval Cue for Extinction Attenuates Spontaneous Recovery. *Journal of Experimental Psychology: Animal Behavior Processes*, *19*(1), 77–89. https://doi.org/10.1037/0097-7403.19.1.77

Broos, N., Schmaal, L., Wiskerke, J., Kostelijk, L., Lam, T., Stoop, N., … Goudriaan, A. E. (2012). The relationship between impulsive choice and impulsive action: A cross-species translational study. *PLoS ONE*, *7*(5), 1–9. https://doi.org/10.1371/journal.pone.0036781

Burger, D. C., Denniston, J. C., & Miller, R. R. (2001). Temporal coding in conditioned inhibition: Retardation tests. *Animal Learning and Behavior*, *29*(3), 281–290. https://doi.org/10.3758/BF03192893

Buss, A., & Plomin, R. (1975). *A temperament theory of personality development.* Retrieved from https://psycnet.apa.org/record/1975-29681-000

Bustamante, J., Uengoer, M., Thorwart, A., & Lachnit, H. (2016). Extinction in multiple contexts: Effects on the rate of extinction and the strength of response recovery. *Learning and Behavior*, *44*(3), 283–294. https://doi.org/10.3758/s13420-016-0212-7

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology*, *67*(2), 319–333. Retrieved from http://www.psy.miami.edu/faculty/ccarver/sclBISBAS.html%5Cnhttp://www.scribd.com/doc/8760706/Carver1994?secret_password=zkhvhvj8dknjmpqjutf

Caswell, A. J., Bond, R., Duka, T., & Morgan, M. J. (2015). Further evidence of the heterogeneous nature of impulsivity. *Personality and Individual Differences*, *76*, 68–74. https://doi.org/10.1016/J.PAID.2014.11.059

Cavagnaro, D. R., Aranovich, G. J., McClure, S. M., Pitt, M. A., & Myung, J. I. (2016). On the Functional Form of Temporal Discounting: An Optimized AdaptiveTest. *Journal of Risk and Uncertainty*, *52*(3), 233. https://doi.org/10.1007/S11166-016-9242-Y

Choy, Y., Fyer, A. J., & Lipsitz, J. D. (2007). Treatment of specific phobia in adults. *Clinical Psychology Review*, *27*(3), 266–286. https://doi.org/10.1016/j.cpr.2006.10.002

Coelho, C. A. O., Dunsmoor, J. E., & Phelps, E. A. (2015). Compound stimulus extinction reduces spontaneous recovery in humans. *Learning and Memory*, *22*(12), 589–593. https://doi.org/10.1101/lm.039479.115

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*(3), 323–341. https://doi.org/10.1006/lmot.1997.0971

Conklin, C. A., & Tiffany, S. T. (2002). Applying extinction research and theory to cue-exposure addiction treatments. *Addiction*, *97*(2), 155–167. https://doi.org/10.1046/j.1360-0443.2002.00014.x

Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy*. https://doi.org/10.1016/j.brat.2014.04.006

Crombag, H. S., Bossert, J. M., Koya, E., & Shaham, Y. (2008). Context-induced relapse to drug seeking: A review. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1507), 3233–3243. https://doi.org/10.1098/rstb.2008.0090

Culver, N. C., Vervliet, B., & Craske, M. G. (2015). Compound Extinction: Using the Rescorla–Wagner Model to Maximize Exposure Therapy Effects for Anxiety Disorders. *Clinical Psychological Science*. https://doi.org/10.1177/2167702614542103

Dickinson, A., & Burke, J. (1996). Within compound Associations Mediate the Retrospective Revaluation of Causality Judgements. *The Quarterly Journal of Experimental Psychology: Section B*, *49*(1), 60–80. https://doi.org/10.1080/713932614

Du, W., Green, L., & Myerson, J. (2002). CROSS-CULTURAL COMPARISONS OF DISCOUNTING DELAYED AND PROBABILISTIC REWARDS The major goal of the present study was to assess the cross-cultural generality of monetary decision-making processes by comparing the discounting of delayed and probabilistic rew, 479–492.

Enticott, P. G., Ogloff, J. R. P., & Bradshaw, J. L. (2006). Associations between laboratory measures of executive inhibitory control and self-reported impulsivity. *Personality and Individual Differences*, *41*(2), 285–294. https://doi.org/10.1016/j.paid.2006.01.011

Enticott, P. G., Ogloff, J. R. P., & Bradshaw, J. L. (2008). Response inhibition and impulsivity in schizophrenia. *Psychiatry Research*, *157*(1–3), 251–254. https://doi.org/10.1016/j.psychres.2007.04.007

Evenden, J. L. (1999). Varieties of impulsivity. *Psychopharmacology*, *146*(4), 348–361.

https://doi.org/10.1007/PL00005481

Everitt, B. J., & Robbins, T. W. (2016). Drug addiction: Updating actions to habits to compulsions ten years on. *Annual Review of Psychology*, *67*, 23–50. https://doi.org/10.1146/annurev-psych-122414-033457

Eysenck, H. J. (1990). Genetic and Environmental Contributions to Individual Differences: The Three Major Dimensions of Personality. *Journal of Personality*, *58*(1), 245–261. https://doi.org/10.1111/J.1467-6494.1990.TB00915.X

Eysenck, S. B. G., Eysenck, H. J., & Barrett, P. (1985). A REVISED VERSION OF THE PSYCHOTICISM SCALE. *Person. Individ. Difl*, *6*(1), 21–29.

Fillmore, M., & Rush, C. R. (2006). Polydrug abusers display impaired discrimination-reversal learning in a model of behavioural control Sensitivity to the disinhibiting effect of alcohol: The role of trait impulsivity and sex differences View project PDMP by provider View project. *Article in Journal of Psychopharmacology*. https://doi.org/10.1177/0269881105057000

Fillmore, M. T., & Rush, C. R. (2002). Impaired inhibitory control of behavior in chronic cocaine users. *Drug and Alcohol Dependence*, *66*(3), 265–273. https://doi.org/10.1016/S0376-8716(01)00206-X

Fortmann, S. P., & Killen, J. D. (1995). Nicotine Gum and Self-Help Behavioral Treatment for Smoking Relapse Prevention: Results From a Trial Using Population-Based Recruitment. *Journal of Consulting and Clinical Psychology*, *63*(3), 460–468. https://doi.org/10.1037//0022-006x.63.3.460

Franken, I. H. A., van Strien, J. W., Nijs, I., & Muris, P. (2008). Impulsivity is associated with behavioral decision-making deficits. *Psychiatry Research*, *158*(2), 155–163. https://doi.org/10.1016/J.PSYCHRES.2007.06.002

Glautier, S., & Brudan, O. (2019). Stable Individual Differences in Occasion Setting. *Experimental Psychology*, *66*(4), 281–295. https://doi.org/10.1027/1618-3169/a000453

Glautier, S., Elgueta, T., & Nelson, J. B. (2013). Extinction produces context inhibition and multiple-context extinction reduces response recovery in human predictive learning. https://doi.org/10.3758/s13420-013-0109-7

González, G., Alcalá, J. A., Callejas-Aguilera, J. E., & Rosas, J. M. (2019). Experiencing extinction with a non-target cue facilitates reversal of a target conditioned inhibitor in human predictive learning. *Behavioural Processes*, *166*(December 2018), 103898. https://doi.org/10.1016/j.beproc.2019.103898

Gray, J. A. (1987). Perspectives on anxiety and impulsivity: A commentary. *Journal of Research in Personality*, *21*(4), 493–509. https://doi.org/10.1016/0092-6566(87)90036-5

Gray, Jeffrey A. (1982). The neuropsychology of anxiety: An enquiry into the functions of septo-hippocampal theories. *Behavioral and Brain Sciences*, *5*(3), 492–493. https://doi.org/10.1017/S0140525X00013170

Griffiths, O., Holmes, N., & Westbrook, R. F. (2017). Compound Stimulus Presentation Does Not Deepen Extinction in Human Causal Learning. *Frontiers in Psychology*, *8*(FEB), 120. https://doi.org/10.3389/fpsyg.2017.00120

Grillon, C., & Ameli, R. (2001). Conditioned inhibition of fear-potentiated startle and skin conductance in humans. *Psychophysiology*, *38*(5), 807–815. https://doi.org/10.1111/1469-8986.3850807

Gullo, M. J., Jackson, C. J., & Dawe, S. (2010). Impulsivity and reversal learning in hazardous alcohol use. *Personality and Individual Differences*, *48*(2), 123–127. https://doi.org/10.1016/J.PAID.2009.09.006

Harris, J. A., Jones, M. L., Bailey, G. K., & Frederick Westbrook, R. (2000). Contextual Control Over Conditioned Responding in an Extinction Paradigm, *26*(2), 174–185. https://doi.org/10.1037/0097-7403.26.2.174

Harris, J. A., Kwok, D. W. S., Andrew, B. J., & Harris, J. (2014). Conditioned inhibition and reinforcement rate, *40*(3), 335–354. https://doi.org/10.1037/xan0000023

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*(MAY), 1–12. https://doi.org/10.3389/fpsyg.2013.00245

He, Z., Cassaday, H. J., Howard, R. C., Khalifa, N., & Bonardi, C. (2011). Impaired pavlovian conditioned inhibition in offenders with personality disorders. *Quarterly Journal of Experimental Psychology*, *64*(12), 2334–2351. https://doi.org/10.1080/17470218.2011.616933

Hermans, D., Craske, M. G., Mineka, S., & Lovibond, P. F. (2006). Extinction in Human Fear Conditioning. *Biological Psychiatry*, *60*(4), 361–368. https://doi.org/10.1016/J.BIOPSYCH.2005.10.006

Ho, M. Y., Mobini, S., Chiang, T. J., Bradshaw, C. M., & Szabadi, E. (1999). Theory and method in the quantitative analysis of ‘impulsive choice’ behaviour: implications for psychopharmacology. *Psychopharmacology (Berl)*, *146*(4), 362–372. https://doi.org/91460362.213 [pii]

Hofmann, W., Friese, M., & Strack, F. (2009). Impulse and Self-Control From a Dual-Systems Perspective.

Holland, P. C. (1989). Transfer of Negative Occasion Setting and Conditioned Inhibition Across Conditioned and Unconditioned Stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, *15*(4), 311–328. https://doi.org/10.1037/0097-7403.15.4.311

Holland, P. C. (1992). Occasion Setting in Pavlovian Conditioning. *Psychology of Learning and Motivation - Advances in Research and Theory*, *28*(C), 69–125. https://doi.org/10.1016/S0079-7421(08)60488-0

Holland, P. C., & Lamarre, J. (1984). Transfer of Inhibition after Serial and Simultaneous Feature Negative Discrimination Training. *LEARNING AND MOTIVATION*, *15*, 219–243.

Honey, R. C., & Hall, G. (1989). Acquired Equivalence and Distinctiveness of Cues. *Journal* *of Experimental Psychology: Animal Behavior Processes*, *15*(4), 338–346. https://doi.org/10.1037/0097-7403.15.4.338

Hoptman, M. J., Volavka, J., Johnson, G., Weiss, E., Bilder, R. M., & Lim, K. O. (2002). Frontal white matter microstructure, aggression, and impulsivity in men with schizophrenia: A preliminary study. *Biological Psychiatry*, *52*(1), 9–14. https://doi.org/10.1016/S0006-3223(02)01311-2

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*(3), 381–394. https://doi.org/10.1037/a0017837

Jacoby, R. J., & Abramowitz, J. S. (2016). Inhibitory learning approaches to exposure therapy: A critical review and translation to obsessive-compulsive disorder. *Clinical Psychology Review*, *49*, 28–40. https://doi.org/10.1016/j.cpr.2016.07.001

Janak, P. H., Bowers, M. S., & Corbit, L. H. (2012). Compound stimulus presentation and the norepinephrine reuptake inhibitor atomoxetine enhance long-term extinction of cocaine-seeking behavior. *Neuropsychopharmacology*, *37*(4), 975–985. https://doi.org/10.1038/npp.2011.281

Janak, P. H., & Corbit, L. H. (2011). Deepened extinction following compound stimulus presentation: Noradrenergic modulation. *Learning and Memory*, *18*(1), 1–10. https://doi.org/10.1101/lm.1923211

Johnson, M. W., & Bickel, W. K. (2002). WITHIN-SUBJECT COMPARISON OF REAL AND HYPOTHETICAL MONEY REWARDS IN DELAY DISCOUNTING. *JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR*, *77*, 129–146.

Jorm, A. F., Christensen, H., Henderson, A. S., Jacomb, P. A., Körten, A. E., & Rodgers, B. (1998). Using the BIS/BAS scales to measure behavioural inhibition and behavioural activation: Factor structure, validity and norms in a large community sample. *Personality and Individual Differences*, *26*(1), 49–58. https://doi.org/10.1016/S0191-8869(98)00143-

3

Kamin, L. (1969). PREDICTABILITY, SURPRISE, ATTENTION, AND CONDITIONING.

Kaplan, B. A., Amlung, M., Reed, D. D., Jarmolowicz, D. P., McKerchar, T. L., & Lemley, S. M. (2016). Automating Scoring of Delay Discounting for the 21- and 27-Item Monetary Choice Questionnaires. *Behavior Analyst*, *39*(2), 293–304. https://doi.org/10.1007/s40614-016-0070-9

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 Section B: Comparative and Physiological Psychology*, *57*(2), 153–178. https://doi.org/10.1080/02724990344000033

Karazinov, D. M., & Boakes, R. A. (2007). Second-order conditioning in human predictive judgements when there is little time to think, *60*(3), 448–460. https://doi.org/10.1080/17470210601002488

Kearns, D. N., Tunstall, B. J., & Weiss, S. J. (2012). Deepened extinction of cocaine cues. *Drug and Alcohol Dependence*, *124*(3), 283–287. https://doi.org/10.1016/j.drugalcdep.2012.01.024

Kirby, K. N. (1997). Bidding on the Future: Evidence Against Normative Discounting of Delayed Rewards. *Journal of Experimental Psychology: Genera*, *126*(1), 54–70.

Kirby, K. N., & Herrnstein, R. J. (1995). Preference Reversals Due to Myopic Discounting of Delayed Reward. *Https://Doi.Org/10.1111/j.1467-9280.1995.Tb00311.X*, *6*(2), 83–89. https://doi.org/10.1111/J.1467-9280.1995.TB00311.X

Kiyak, C., Simonetti, M. E., Norton, S., & Deluca, P. (2023). The efficacy of cue exposure therapy on alcohol use disorders: A quantitative meta-analysis and systematic review. *Addictive Behaviors*, *139*(August 2022), 107578. https://doi.org/10.1016/j.addbeh.2022.107578

Klein, S. D., Collins, P. F., & Luciana, M. (2022). Developmental trajectories of delay discounting from childhood to young adulthood: longitudinal associations and test-retest reliability. *Cognitive Psychology*, *139*(September 2021), 101518. https://doi.org/10.1016/j.cogpsych.2022.101518

Krypotos, A. M., & Engelhard, I. M. (2019). Targeting avoidance via compound extinction. *Cognition and Emotion*, *33*(7), 1523–1530. https://doi.org/10.1080/02699931.2019.1573718

Lagorio, C. H., & Madden, G. J. (2005). Delay discounting of real and hypothetical rewards III: Steady-state assessments, forced-choice trials, and all real rewards. *Behavioural Processes*, *69*, 173–187. https://doi.org/10.1016/j.beproc.2005.02.003

Laing, P. A. F., Burns, N., & Baetu, I. (2019). Individual differences in anxiety and fear learning: The role of working memory capacity. *Acta Psychologica*, *193*, 42–54. https://doi.org/10.1016/j.actpsy.2018.12.006

Laing, P. A. F., Vervliet, B., Angel, M., Savage, H. S., Davey, C. G., Felmingham, K. L., & Harrison, B. J. (2021). Behaviour Research and Therapy Characterizing human safety learning via Pavlovian conditioned inhibition. *Behaviour Research and Therapy*, *137*(January), 103800. https://doi.org/10.1016/j.brat.2020.103800

Larrauri, J. A., & Schmajuk, N. A. (2008). Attentional, Associative, and Configural Mechanisms in Extinction. *Psychological Review*, *115*(3), 640–676. https://doi.org/10.1037/0033-295X.115.3.640

Lee, J. C., & Livesey, E. J. (2012). Second-Order Conditioning and Conditioned Inhibition: Influences of Speed versus Accuracy on Human Causal Learning. *PLoS ONE*, *7*(11), e49899. https://doi.org/10.1371/journal.pone.0049899

Lee, J. C., & Lovibond, P. F. (2021). Individual differences in causal structures inferred during feature negative learning. *Quarterly Journal of Experimental Psychology*, *74*(1), 150–165. https://doi.org/10.1177/1747021820959286

Leung, H. T., Reeks, L. M., & Westbrook, R. F. (2012). Two ways to deepen extinction and the difference between them. *Journal of Experimental Psychology: Animal Behavior Processes*, *38*(4), 394–406. https://doi.org/10.1037/a0030201

Lipszyc, J., & Schachar, R. (2010). Inhibitory control and psychopathology: A meta-analysis of studies using the stop signal task. *Journal of the International Neuropsychological Society*, *16*(6), 1064–1076. https://doi.org/10.1017/S1355617710000895

Logan, G. D. (1994). On the ability to inhibit thought and action: A users' guide to the stop signal paradigm. In *Inhibitory processes in attention, memory, and language.* (pp. 189–239). San Diego, CA, US: Academic Press.

Logan, G. D., & Cowan, W. B. (1984). *On the Ability to Inhibit Thought and Action: A Theory of an Act of Control. Psychological Review* (Vol. 91). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.417.8257&rep=rep1&type=pdf

Logan, G. D., Schachar, R. J., & Tannock, R. (1997). Impulsivity and Inhibitoey Control. *Psychological Science*, *8*(1), 60–64.

Lotz, A., & Lachnit, H. (2009). Extinction of conditioned inhibition: Effects of different outcome continua. *Learning and Behavior*, *37*(1), 85–94. https://doi.org/10.3758/LB.37.1.85

Lotz, A., Vervliet, B., & Lachnit, H. (2009). Blocking of conditioned inhibition in human causal learning: No learning about the absence of outcomes. *Experimental Psychology*, *56*(6), 381–385. https://doi.org/10.1027/1618-3169.56.6.381

Lovibond, P. F., Davis, N. R., & O'Flaherty, A. S. (2000). Protection from extinction in human fear conditioning. *Behaviour Research and Therapy*, *38*(10), 967–983. https://doi.org/10.1016/S0005-7967(99)00121-7

Loy, I., Carnero-Sierra, S., Acebes, F., Muñiz-Moreno, J., Muñiz-Diez, C., & Sánchez-González, J. C. (2021). Where Association Ends. A Review of Associative Learning in

Invertebrates,Plants and Protista,and a Reflection on Its Limits. *Journal of Experimental Psychology: Animal Learning and Cognition*, *47*(3), 234–251. https://doi.org/10.1037/xan0000306

McConnell, B. L., Miguez, G., & Miller, R. R. (2013). Extinction with multiple excitors. *Learning & Behavior*, *41*(2), 119–137. https://doi.org/10.3758/s13420-012-0090-6

McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological Review*. https://doi.org/10.1037/h0069819

Melchers, K. G., Wolff, S., & Lachnit, H. (2006). Extinction of conditioned inhibition through nonreinforced presentation of the inhibitor. *Psychonomic Bulletin and Review*, *13*(4), 662–667. https://doi.org/10.3758/BF03193978

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. *Behavioural Brain Research*, *168*(2), 299–306. https://doi.org/10.1016/j.bbr.2005.11.021

Miguez, G., McConnell, B., Polack, C. W., & Miller, R. R. (2018). Proactive interference by cues presented without outcomes: Differences in context specificity of latent inhibition and conditioned inhibition. *Learning and Behavior*, *46*(3), 265–280. https://doi.org/10.3758/s13420-017-0306-x

Miguez, G., Soares, J. S., & Miller, R. R. (2015). The role of test context in latent inhibition of conditioned inhibition: Part of a search for general principles of associative interference. *Learning and Behavior*, *43*(3), 228–242. https://doi.org/10.3758/s13420-015-0175-0

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin*, *117*(3), 363–386. https://doi.org/10.1037/0033-2909.117.3.363

Nash, J. C., & Varadhan, R. (2011). Unifying Optimization Algorithms to Aid Software

System Users: optimx for R. *Journal of Statistical Software*, *43*(9), 1–14. https://doi.org/10.18637/JSS.V043.I09

Nelson, J. B. (2002). Context Specificity of Excitation and Inhibition in Ambiguous Stimuli. https://doi.org/10.1006/lmot.2001.1112

Neumann, D. L., Lipp, O. V., & Siddle, D. A. T. (1997). Conditioned inhibition of autonomic Pavlovian conditioning in humans. *Biological Psychology*, *46*(3), 223–233. https://doi.org/10.1016/S0301-0511(97)05248-4

Northrup, T. F., Stotts, A. L., Green, C., Potter, J. S., Marino, E. N., Walker, R., … Trivedi, M. (2015). Opioid withdrawal, craving, and use during and after outpatient buprenorphine stabilization and taper: A discrete survival and growth mixture model. *Addictive Behaviors*, *41*, 20. https://doi.org/10.1016/J.ADDBEH.2014.09.021

O’Donnell, B. F., Skosnik, P. D., Hetrick, W. P., & Fridberg, D. J. (2021). Decision Making and Impulsivity in Young Adult Cannabis Users. *Frontiers in Psychology*, *12*, 679904. https://doi.org/10.3389/FPSYG.2021.679904/BIBTEX

Patterson, C. M., & Newman, J. P. (1993). Reflectivity and Learning From Aversive Events: Toward a Psychological Mechanism for the Syndromes of Disinhibition. *Psychological Review*, *100*(4), 716–736. https://doi.org/10.1037/0033-295x.100.4.716

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the barratt impulsiveness scale. *Journal of Clinical Psychology*, *51*(6), 768–774. https://doi.org/10.1002/1097-4679(199511)51:6<768::AID-JCLP2270510607>3.0.CO;2-1

Paulsen, K., & Johnson, M. (1980). Impulsivity: A multidimensional concept with developmental aspects. *Journal of Abnormal Child Psychology*, *8*(2), 269–277. https://doi.org/10.1007/BF00919070

Pearce, J. M. (1987). A Model for Stimulus Generalization in Pavlovian Conditioning. *Psychological Review*, *94*(1), 61–73. https://doi.org/10.1037/0033-295X.94.1.61

Peters, J., & Büchel, C. (2011). The neural mechanisms of inter-temporal decision-making: Understanding variability. *Trends in Cognitive Sciences*, *15*(5), 227–239. https://doi.org/10.1016/j.tics.2011.03.002

Pineño, O. (2007). *Protection from extinction by concurrent presentation of an excitor or an extensively extinguished CS* (Vol. 28). Retrieved from www.opineno.com.

Polack, C. W., Laborda, M. A., & Miller, R. R. (2012). Extinction context as a conditioned inhibitor. *Learning and Behavior*, *40*(1), 24–33. https://doi.org/10.3758/s13420-011-0039-1

Porter, J. N., Olsen, A. S., Gurnsey, K., Dugan, B. P., Jedema, H. P., & Bradberry, C. W. (2011). Chronic Cocaine Self-Administration in Rhesus Monkeys: Impact on Associative Learning, Cognitive Control, and Working Memory. https://doi.org/10.1523/JNEUROSCI.5426-10.2011

Quirk, G. J. (2002). Memory for Extinction of Conditioned Fear Is Long-lasting and Persists Following Spontaneous Recovery. *Learning & Memory*, *9*(6), 402–407. https://doi.org/10.1101/LM.49602

Rachlin, H., & Green, L. (1972). *JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR COMMITMENT, CHOICE AND SELF-CONTROL'*. Retrieved from https://onlinelibrary.wiley.com/doi/pdf/10.1901/jeab.1972.17-15

Reberg, D. (1972). Compound tests for excitation in early acquisition and after prolonged extinction of conditioned suppression. *Learning and Motivation*, *3*(3), 246–258. https://doi.org/10.1016/0023-9690(72)90021-5

Redhead, E. S., & Chan, W. (2017). Conditioned inhibition in the spatial domain in humans and rats. *Learning and Motivation*, *59*(August), 27–37. https://doi.org/10.1016/j.lmot.2017.08.001

Rescorla, R. A. (2000). Extinction can be enhanced by a concurrent excitor [In Process Citation]. *Journal of Experimental Psychology: Animal Behavior Processes*, *26*(3), 251–

260.

Rescorla, R. A. (1969). Pavlovian conditioned inhibition. *Psychological Bulletin*, *72*(2), 77–94. https://doi.org/10.1037/h0027760

Rescorla, R. A. (1973). Evidence for ‘unique stimulus’ account of configural conditioning. *Journal of Comparative and Physiological Psychology*, *85*(2), 331–338. https://doi.org/10.1037/h0035046

Rescorla, R. A. (1986). Extinction of Facilitation. *Journal of Experimental Psychology: Animal Behavior Processes*, *12*(1), 16–24. https://doi.org/10.1037/0097-7403.12.1.16

Rescorla, R. A. (1987). Facilitation and Inhibition. *Journal of Experimental Psychology: Animal Behavior Processes*, *13*(3), 250–259. https://doi.org/10.1037/0097-7403.13.3.250

Rescorla, R. A. (2003). Protection from extinction. *Learning and Behavior*, *31*(2), 124–132. https://doi.org/10.3758/bf03195975

Rescorla, R. A. (2006). Deepened extinction from compound stimulus presentation. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(2), 135–144. https://doi.org/10.1037/0097-7403.32.2.135

Rescorla, R. A., & Holland, P. C. (1977). Associations in Pavlovian conditioned inhibition. *Learning and Motivation*, *8*(4), 429–447. https://doi.org/10.1016/0023-9690(77)90044-3

Rescorla, R. A. (1973). EVIDENCE FOR ‘UNIQUE STIMULUS’ ACCOUNT OF CONFIGURAL CONDITIONING 1. *Journal of Comparative and Physiological Psychology*, *85*(2), 331–338.

Rescorla, R. A., & Wagner, A. R. (1972). *A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement*. Retrieved from https://pdfs.semanticscholar.org/afaf/65883ff75cc19926f61f181a687927789ad1.pdf

Reynolds, B., Ortengren, A., Richards, J. B., & de Wit, H. (2006). Dimensions of impulsive behavior: Personality and behavioral measures. *Personality and Individual Differences*, *40*(2), 305–315. https://doi.org/10.1016/j.paid.2005.03.024

Richards, J B, Mitchell, S. H., de Wit, H., & Seiden, L. S. (1997). Determination of discount functions in rats with an adjusting-amount procedure. *Journal of the Experimental Analysis of Behavior*, *67*(3), 353–366. https://doi.org/10.1901/jeab.1997.67-353

Richards, Jerry B, Zhang, L., Mitchell, S. H., & De Wit, H. (1999). DELAY OR PROBABILITY DISCOUNTING IN A MODEL OF IMPULSIVE BEHAVIOR: EFFECT OF ALCOHOL. *JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR*, *71*, 121–143.

Richardson, R. A., Michener, P. N., Gann, C. L., North, I. M., & Schachtman, T. R. (2020). Summation and retardation test performance following extinction or Pavlovian conditioned inhibition training. *Learning and Motivation*, *71*, 101642. https://doi.org/10.1016/j.lmot.2020.101642

Rodriguez-Jimenez, R., Avila, C., Jimenez-Arriero, M. A., Ponce, G., Monasor, R., Jimenez, M., … Palomo, T. (2006). Impulsivity and sustained attention in pathological gamblers: Influence of childhood ADHD history. *Journal of Gambling Studies*, *22*(4), 451–461. https://doi.org/10.1007/s10899-006-9028-2

Ross, R. T., & Holland, P. C. (1981). *Conditioning of simultaneous and serial feature-positive discriminations*.

Rubio, G., Jiménez, M., Rodríguez-Jiménez, R., Martínez, I., Ávila, C., Ferre, F., … Palomo, T. (2008). The role of behavioral impulsivity in the development of alcohol dependence: A 4-year follow-up study. *Alcoholism: Clinical and Experimental Research*, *32*(9), 1681–1687. https://doi.org/10.1111/j.1530-0277.2008.00746.x

Sansa, J., Rodrigo, T., Santamaría, J. J., Manteiga, R. D., & Chamizo, V. D. (2009). Conditioned Inhibition in the Spatial Domain. *Journal of Experimental Psychology: Animal Behavior Processes*, *35*(4), 566–577. https://doi.org/10.1037/a0015630

Sato, T. (2005). The Eysenck Personality Questionnaire brief version: Factor structure and reliability. *Journal of Psychology: Interdisciplinary and Applied*, *139*(6), 545–552. https://doi.org/10.3200/JRLP.139.6.545-552

Schachar, R. J., Tannock, R., & Logan, G. (1993). INHIBITORY CONTROL, IMPULSIVENESS, AND ATTENTION DEFICIT HYPERACTIVITY DISORDER. *Clinical Po,Chology Review*, *13*, 721–739.

Sosa, R. (2022). Conditioned Inhibition , Inhibitory Learning , Response Inhibition , and Inhibitory Control : Outlining a Conceptual Clarification, (October).

Sosa, R., & Ramírez, M. N. (2019). Conditioned inhibition: Historical critiques and controversies in the light of recent advances. *Journal of Experimental Psychology: Animal Learning and Cognition*, *45*(1), 17–42. https://doi.org/10.1037/xan0000193

Stanford, M. S., Mathias, C. W., Dougherty, D. M., Lake, S. L., Anderson, N. E., & Patton, J. H. (2009). Fifty years of the Barratt Impulsiveness Scale: An update and review. *Personality and Individual Differences*, *47*(5), 385–395. https://doi.org/10.1016/J.PAID.2009.04.008

Stout, S., Escobar, M., & Miller, R. R. (2004). Trial number and compound stimuli temporal relationship as joint determinants of second-order conditioning and conditioned inhibition. *Learning and Behavior, 32*(2), 230–239. https://doi.org/10.3758/bf03196024

Swartzentruber, D. (1995). Modulatory mechanisms in Pavlovian conditioning. *Animal Learning & Behavior*, *23*(2), 123–143.

Taylor, J. R., Olausson, P., Quinn, J. J., & Torregrossa, M. M. (2009). Targeting extinction and reconsolidation mechanisms to combat the impact of drug cues on addiction. *Neuropharmacology*, *56*(SUPPL. 1), 186–195. https://doi.org/10.1016/j.neuropharm.2008.07.027

Thomas, B. L., & Ayres, J. J. . (2004). Use of the ABA fear renewal paradigm to assess the effects of extinction with co-present fear inhibitors or excitors: Implications for theories of extinction and for treating human fears and phobias. *Learning and Motivation*, *35*(1), 22–52. https://doi.org/10.1016/S0023-9690(03)00040-7

Urcelay, G. P., & Miller, R. R. (2008). Counteraction between two kinds of conditioned inhibition training. *Psychonomic Bulletin and Review*, *15*(1), 103–107. https://doi.org/10.3758/PBR.15.1.103

Urcelay, G. P., Perelmuter, O., & Miller, R. R. (2008). Pavlovian backward conditioned inhibition in humans: Summation and retardation tests. *Behavioural Processes*, *77*(3), 299–305. https://doi.org/10.1016/j.beproc.2007.07.003

Van Hamme, L. J., & Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation*, *25*(2), 127–151. https://doi.org/10.1006/lmot.1994.1008

Verbruggen, F., Aron, A. R., Band, G. P. H., Beste, C., Bissett, P. G., Brockett, A. T., … Boehler, C. N. (2019). A consensus guide to capturing the ability to inhibit actions and impulsive behaviors in the stop-signal task. *ELife*, *8*. https://doi.org/10.7554/ELIFE.46323

Vervliet, B., Craske, M. G., & Hermans, D. (2013). Fear Extinction and Relapse: State of the Art. *Annual Review of Clinical Psychology*, *9*(1), 215–248. https://doi.org/10.1146/annurev-clinpsy-050212-185542

Vervliet, B., Vansteenwegen, D., Hermans, D., & Eelen, P. (2007). Concurrent excitors limit the extinction of conditioned fear in humans. *Behaviour Research and Therapy*, *45*(2), 375–383. https://doi.org/10.1016/J.BRAT.2006.01.009

Williams, B. A., & Mcdevitt, M. A. (2002). *INHIBITION AND SUPERCONDITIONING*.

Williams, D. A. (1995). Forms of Inhibition in Animal and Human Learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *21*(2), 129–142. https://doi.org/10.1037/0097-7403.21.2.129

Woodbury, C. B. (1943). The learning of stimulus patterns by dogs. *Journal of Comparative*

Psychology, *35*(1), 29–40. https://doi.org/10.1037/H0054061

Yechiam, E., & Busemeyer, J. R. (2005). Comparison of basic assumptions embedded in learning models for experience-based decision making. *Psychonomic Bulletin and Review*, *12*(3), 387–402. https://doi.org/10.3758/BF03193783

Zaksaite, T., & Jones, P. M. (2019). The redundancy effect is related to a lack of conditioned inhibition: Evidence from a task in which excitation and inhibition are symmetrical. *Quarterly Journal of Experimental Psychology*, *2020*(2), 260–278. https://doi.org/10.1177/1747021819878430

Zou, A. R., Muñoz Lopez, D. E., Johnson, S. L., & Collins, A. G. E. (2022). Impulsivity Relates to Multi-Trial Choice Strategy in Probabilistic Reversal Learning. *Frontiers in Psychiatry*, *13*, 800290. https://doi.org/10.3389/FPSYT.2022.800290/BIBTEX

Bibliography

Model Comparison Code: https://osf.io/p59zu/

R Core Development Team. (2020). R: A language and environment for statistical computing.

Stop Signal Reaction Task: https://github.com/fredvbrug/STOP-IT

Wikipedia. (2020, May 5). *Softmax function*. https://en.wikipedia.org/wiki/Softmax_function