Judgement Bias in Hens

A thesis

submitted in partial fulfilment

of the requirements for the degree

of

Master of Applied Psychology (Behaviour Analysis)

at

The University of Waikato

by

STACEY (ACE) TER VEER-BURKE

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2016

Abstract

When an animal has learned a discrimination between two stimuli on the same dimension, exposing the animal to intermediate stimuli will produce a generalisation gradient. Events that that shift responding from this baseline are said to bias responding. Shifts in gradient which show more responding to intermediate stimuli are called positive judgement biases; while shifts in gradient showing reduced responding on intermediate stimuli are called negative judgement biases. While initial studies showed that exposure to poor conditions prior to testing produced negative biases, not all of the literature has supported this. This study aimed to clarify inconsistencies in the literature, using a within-subjects design and a short-term aversive event. Chickens were trained under two-component multiple schedules with each component associated with a stimulus location. Phase 1 involved 15 min exposure to white noise at 100dB, followed immediately by judgement bias testing. Phase 2 involved exposure to white noise at the same dB level, but for the duration of the test. Phase 3 involved judgment bias testing, interrupted by 15 min white noise at 100dB, following which the test resumed. According to early studies, chickens would show a negative bias in Phase 1 and in the second half of Phase 3. However, if release from stressful conditions produces a positive judgement bias, then a negative judgement bias would be expected in Phase 2 only. Results showed negative biases for some birds in Phase 1, and for all birds in Phase 2. No statistically significant bias was observed in Phase 3. The author argues that idiosyncratic differences may have influenced the direction of bias observed in Phase 1, and suggests further investigations be conducted.

Acknowledgements

I would like to thank my supervisors Professor T. Mary Foster, Dr Tim Edwards, and Dr James McEwan for their support and advice throughout this project. Many others have also helped me along the way, including Rob Bakker, Jennifer Chandler, Allan Eaddy, and fellow Master's and PhD students who ran the laboratory alongside myself.

Table of Contents

1. Introduction

Animals provide us with food, protection, companionship, entertainment and allow us to make scientific advances where human subjects are deemed inappropriate, along with many other uses. As the Western world has developed, so too have our moral and ethical expectations for the treatment of animals. It is now considered our moral and ethical duty to provide them with not only the necessary resources to survive, but also mental stimulation and enrichment to meet their 'wants'.

The quality of treatment an animal receives dictates their welfare, or current physical and mental state. With animals being used in a number of industries and under a variety of conditions, legislation for animal welfare is essential. It is also necessary that reliable and valid measures of welfare exist, in order to determine whether organisations are compliant. While there are currently accepted standards for the physical welfare of animals, mental states are more difficult to assess as animals are not able to report on affect.

1.1. Animal Emotion

Animals continue to be used as subjects in studies on addiction, mood disorders and other affective problems. This indicates some degree of acceptance amongst the scientific community that animals do experience affective states comparable to human emotion. Desire, Boissy and Veissier (2002) argue that affective states, and an organism's ability to perceive those affective states, may play an adaptive role in maintaining optimal conditions for survival. Put simply, seeking out rewarding events and avoiding events that cause distress is likely to improve chances of survival (Spruijt, van den Bos & Pijlman, 2001).

Considering this, it is perhaps unsurprising that animals express distinctly different behaviours in response to rewarding and aversive events. Correlations have been found between the frequency (kHz) of vocalisations emitted by rats and the reinforcing/aversive properties of events that rats were exposed to. Rats often emit 50-kHz vocalisation during positively valenced activities such as during copulation, play and when tickled by handlers (Panksepp, 2007). Conversely, vocalisations closer to 20-kHz have been associated with loss of reward (Burgdorf, Knutson & Panksepp, 2000) and exposure to aversive bright light (Knutson, Burgdorf & Panksepp, 1998). These vocalisations are also emitted in a similar fashion when anticipating events (Burgdorf et al., 2000; Knutson et al., 1998).

Similarly, hens behaved differently when exposed to sound cues that signalled a neutral (no treatment) event, a positive (delivery of mealworms) event and a negative (delivery of a spray of water) event (Zimmerman, Buijs, Bolhuis & Keeling, 2011). Significantly more 'stepping' and head movements were seen in the presence of the negative cue compared with the neutral or positive cue (Zimmerman et al., 2011). Time spent preening was also significantly greater in the presence of the positive cue when compared with the other two conditions (Zimmerman et al., 2011).

These changes in rate of behaviour suggest that the rats and hens were able to perceive the reinforcing/aversive qualities of signalled events. Particularly in the Zimmerman et al. (2011) study, which used a within-subjects design, the difference in rates of particular behaviours seen under negatively and positively valenced conditions indicates that hens may also experience differing levels of 'comfort' or 'pleasure'.

Not only do animals respond in distinct, predictable ways when faced with positively and negatively valenced cues and conditions; but associated cues and settings may elicit these behaviours even when there is no current observable biological pay-off. Starlings were trained to respond (key-peck) to two differently coloured keys on equal fixed-interval schedules, one under deprivation (H) conditions and the other under pre-fed (PF) conditions (Pompilio & Kacelnik, 2005). When given a concurrent choice between the two keys in a preference test, starlings chose the key colour associated with deprivation conditions. This preference was seen not only when schedules were equal, but also when the fixed interval on the H key was increased up to 50% longer than that on the PF key (Pompilio et al., 2005).

An establishing operation can be described as setting events or conditions "that affects an organism by momentarily altering (a) the reinforcing effectiveness of other events and (b) the frequency of occurrence of … [behaviours] relevant to those events as consequences" (Michael, 1993, p. 192). Here, it appears that the strong establishing operation of food deprivation altered the reinforcing value of the food on the H schedule (and so the motivation to work on the H schedule over the PF schedule). As the food provided was the same for both schedules, this suggests that perhaps reinforcement processes are more complex than just weighing up the quality (e.g. flavour) or quantity of a resource.

One explanation for the preference differences might be associative learning or classical conditioning. Classical conditioning involves the pairing of an unconditioned stimulus (US) such as food, with a conditioned stimulus (CS) such as a tone (Malaka, 1999). The result is that the CS comes to elicit the same response (called the conditioned response or CR) as the US, even in the absence of the US (Malaka, 1999). In the example, the tone would elicit salivation in the absence of food.

In the Pompilio et al. (2005) study, the food from the deprivation schedule may have been paired with release from distress (in form of deprivation). If this were the case, then it could be argued that release from distress may have a 'pleasurable' effect and so the food also becomes more 'pleasurable' in comparison to food from an alternative schedule. While this explanation remains speculative, it does appear that affective states play a functional role in behaviour.

Maintaining a balance of distressing and pleasurable affective states is essential for their welfare, allowing organisms to adapt quickly to new information and optimise outcomes (Spruijt et al., 2001). If an organism has never experienced distress, then they may not adapt well to unfamiliar events or contexts. On the reverse side, an animal that has experienced an abundance of distress may begin to show deterioration in condition. If animals are able to interpret and perceive differing events relevant to their welfare, it may be possible to assess if welfare is suffering before physical signs (for example, prolonged weight loss) appear.

1.2. Welfare Assessment

Animal emotion is commonly assessed in a number of ways. Physiological measurements often include hormone levels, heart rate, and processes within the hypothalamic-pituitary-adrenal (HPA) axis, which is implicated in stress (Paul, Harding, & Mendl, 2005). Although, these measures are sometimes unreliable as the same physiological changes (e.g. increase in heart rate) may be brought about by a number of events (Paul et al., 2005; Otovic, & Hutchinson, 2015).

Other measures of mental welfare include: rate of natural behaviours such as grooming, playing and exploration (Siegford, 2013); stress tests such as the novel object test; and behavioural indicators such as ear posture in sheep, or the vocalisations mentioned above (Siegford, 2013). While these may give some indication of welfare, many do not give a clear picture of the relative 'aversiveness' or 'pleasurableness' of differing conditions. In other words, they

may tell us that an animal is stressed by being transported in a truck, but not how that compares to being separated from conspecifics or being under no-stress conditions.

Judgement bias has been the focus of a number of recent studies on animal welfare, with particular interest in its potential for use in measuring welfare. When testing for a bias, animals are first trained to discriminate between two stimuli, which are paired with outcomes that differ to some degree. The S+ (positive stimulus) is reinforced on a rich schedule, while the S- (negative stimulus) will be on one of: a) a lean schedule of reinforcement; b) an extinction schedule; or c) punishment (normally via electric shock or puff of air).

One or more ambiguous (neutral) stimuli varying along the same dimension as the S+ and S- are presented, and the organisms' responses to these stimuli are measured. For example, three buckets located between the S+ and S- and situated equal distances apart from each other. These probe stimuli are paired with neither a positive nor a negative outcome. Once a baseline measure of responding across probes has been recorded, animals are exposed to a particular event or environment before a second judgement bias test is completed. Often, the resulting responses will be categorised. Positive biases are produced when responding on ambiguous probes increases following exposure (also referred to as 'optimistic' responding); while negative biases are produced when responding on ambiguous probes decreases following exposure (also referred to as 'pessimistic' responding). These patterns in responding, referred to as 'judgement biases', can be used to gauge the affective significance of events.

1.3. Stimulus Control

Central to the model of judgement bias is the concept of stimulus control. According to Terrace (1966), stimulus control can be defined as "the extent to which the value of an antecedent stimulus determines the probability of occurrence of a conditioned response" (p. 271). In other words, when stimuli are presented along some dimension (i.e. colour), those that signal opportunity for reinforcement (or appear similar to a signal) will elicit more responding than stimuli that are unfamiliar or signal no opportunity for reinforcement. Differential responding produces a generalisation gradient, where selective responding to the reinforced stimulus alone shows strong stimulus control and equal responding across all stimuli demonstrates a lack of stimulus control (Terrace, 1966).

Differential reinforcement has been identified as a requirement for stimulus control (Terrace, 1966). Differential reinforcement involves rewarding responses to one stimulus (the S+), while providing either punishment or no reward for responses to stimuli that are not the S+. In an investigation using pigeons, it was found that differential reinforcement of an S+ over an S- produced different patterns in responding as compared with intermittent or 'regular' reinforcement of an S+ alone (Jenkins, 1961). Pigeons in the discrimination group reduced responding on the S- while at the same time showing a decrease in latency to respond to the S+. This phenomenon was termed behavioural contrast (Reynolds, 1961). When one schedule is kept constant (S+) and the other is made leaner (S-), the increases in responding on S+ are referred to as positive behavioural contrast (McSweeny, 1983).

Interestingly, behavioural contrast is not seen with errorless learning procedures (Terrace, 1966). Errorless learning procedures were initially developed by Skinner (1938), and involved introducing the S- early on, and gradually narrowing the difference between the S+ and S-. This procedure prevents any experience of 'incorrect' responding (responding to S-). When chloropromazine or imipramine (drugs used in the treatment of mood disorders) were administered to pigeons who had learned a discrimination via error learning, responding on the S- was reinstated (Terrace, 1963). In contrast, pigeons who had learned the discrimination using an errorless learning procedure showed no change in performance. It was suggested that the extinction schedule of the S- was aversive for the pigeons, and that the drugs temporarily reduced the aversive consequences of the S- (Terrace, 1963).

Similarly, peak shift is only seen when a discrimination is taught through error learning methods (Terrace, 1964). Peak shift occurs when there is an interaction between the excitatory gradients (higher rates of responding) around the S+ and the inhibitory gradients (lower rates of responding) around the S- (Andrew, Perry, Barron, Berthon, Peralta, & Cheng, 2014). This shifts responding further from the S- (Andrew et al., 2014). Pigeons were split into three groups: group one was trained to respond to an S+ only (580 mu); group two was trained to respond to the S+ and to an S- (540 mu) under an errorless learning procedure; and group three was trained the same discrimination, except under error learning conditions (Terrace, 1964). It was found that for the first two

groups, the peak occurred on the S+; while the peak for the third group had shifted to 590 mu.

These findings would indicate that error learning is essential to the validity of judgement bias testing, which requires that the S- be aversive in order for a bias to be detected. The same effect can also be obtained where one schedule is lean and the alternative is made richer (Guttman, 1959). This means that animals do not need to be exposed to harm in order to find a stimulus aversive, and allows for simple designs in judgement bias testing.

Certain schedules of reinforcement also appear to attain steeper generalisation gradients than others. When differing variable-interval (VI) schedules were used (VI 30-s, VI 1-min, VI 2-min, VI 3-min, VI 4-min) it was found that VI 30-s and VI 1-min produced steeper gradients than the other three schedules (Hearst, Koresko and Poppen, 1964). Changes in rates of responding will more easily be detected with a steeper gradient. This is important for judgement bias testing, which assesses the changes in responding brought about by exposure to welfare conditions. Particular shifts in responding have been observed in a number of studies investigating judgement bias, when animals were exposed to aversive events.

1.4. Biases in Responding

Negative judgement biases have been reported following exposure to aversive events, across a number of investigations and subjects: including bees (Bateson, Desire, Gartside, & Wright, 2011); dairy calves (Daros, Costa, von Keyserlingk, Ho ̈tzel, & Weary, 2014); rats (Richter, Schick, Hoyer, Lankisch, Gass, & Vollmayr, 2012; Rygula, Pluta, & Popik, 2012; Papciak, Popik, Fuchs, & Rygula, 2013); starlings (Matheson, Asher, & Bateson, 2008); chicks (Salmeto, Hymel, Carpenter, Brilot, Bateson, & Sufka, 2011; Hymel, & Sufka, 2012); and mice (Kloke, Schrieber, Bodden, Mo ̈ llers, Ruhmann, Kaiser, Lesch, Sascher, & Lewejohann, 2014).

One of the first studies into animal judgement bias was carried out on rats, who were exposed to either unpredictable or predictable housing following training on a discrimination task (Harding, Paul, & Mendl, 2004). Unpredictable housing was intended to induce a “depression-like state” (p. 6972), by adding 1-2 aversive events at a random time each day. Once exposed to the housing conditions for 9 consecutive days, the rats were tested on a judgement bias task. Results showed that rats exposed to the unpredictable housing were less likely to

approach probes close to, and including, the known positive event (Harding et al., 2004). This was consistent with human studies into depression, which showed that those in depressive states were less likely to anticipate a positive event.

More recent investigations have supported this finding. Bees were trained to extend their proboscis to one set of odours (CS+) and avoid extending their proboscis to another set (CS-) (Bateson et al., 2011). Following this, bees were shaken for 60s in attempt to simulate attack. When faced with unfamiliar odour mixtures, which were varying combinations of the CS+ and CS-, shaken bees showed a pessimistic judgement bias (see Figure 1). In other words, they were less likely than non-shaken bees to extend their proboscis to all of the stimuli, though this change in behaviour was more significant for the CS- and those ambiguous probes closest in odour to the CS-. Interestingly, the researchers in this study also measured serotonin, dopamine, and octopamine [functions in reward learning] and found that levels of these hormones decreased following the simulation (Bateson et al., 2011).

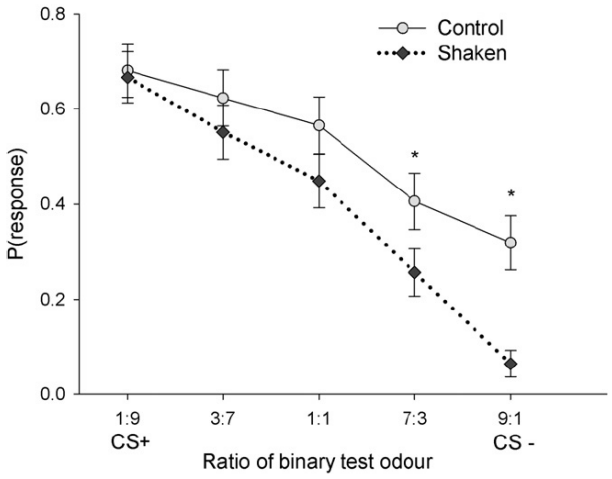

Figure 1. Proportion of responses (extension of proboscis) to a CS+ odour associated with reward, a CS- odour associated with no reward, and three novel odours (all containing differing ratios of 1-hexanol and 2-octanone) under control conditions and following an aversive event (shaken). Figure retrieved from "Agitated Honeybees Exhibit Pessimistic Cognitive Biases," by M. Bateson, S. Desire, S. E. Gartside, and G. A. Wright, 2011, Current Biology, 21, p. 1070-1073.

Positive biases have also been produced in research using starlings (Matheson et al., 2008). In this case, starlings housed in more enriched cages were more likely to respond to ambiguous cues as if they signalled a more positive

outcome, when compared with starlings housed in standard cages (Matheson et al., 2008).

1.5. Conflicting Results

Despite the evidence for use of judgement bias as a means of animal welfare testing, there have been some inconsistent results. Chickens housed in basic pens did not differ significantly in their responding to ambiguous probes compared to enriched birds, apart from being slightly (not statistically significantly) more likely to approach the middle probe (Wichman, Keeling, & Forkman, 2012). This was attributed to the fact that enriched birds received extra food and so were perhaps less motivated to approach. However, individual differences in anticipatory behaviour, relationship to other chickens and feeding motivation were correlated with judgement bias (Wichman et al., 2012). A number of other explanations have also been provided as to why some results depart from expectations; including habituation effects, satiation, history effects and repeated measures effects.

1.5.1. Habituation effects. Time is a factor in the effectiveness of an aversive stimulus, or poor welfare conditions, to produce the unconditioned response. When aversive stimuli are repeatedly presented, the unconditioned response (e.g. escape) is less likely to be emitted (Jordan, Todd, Bucci, & Leaton, 2015). This process is called habituation. Parker, Paul, Burman, Browne and Mendl (2014) investigated the effect of manipulating predictability of events on judgement bias in rats. Rats placed in unpredictable housing exhibited less stress behaviours than controls, who were more 'pessimistic' in their responding to ambiguous probes. These results contrast with Harding et al.'s (2004) study on rats, which found that unpredictable housing produced a negative judgement bias. One difference is that the Parker et al. (2014) study placed rats under 'unpredictable' conditions for 28 days; while Harding et al. (2004) only carried out this treatment for 9 days. Thus, the rats in the former study had more time to habituate to conditions.

1.5.2. Satiation. Burman, McGowan, Mendl, Norling, Paul, Rehn and Keeling (2011) explored effects of recent rewards on dogs' judgement of ambiguous probes in a go/no go judgement bias task. Those dogs who had recently experienced food rewards were more 'pessimistic' in their responding compared with non-rewarded dogs. Similarly, research into judgement bias with horses found that those trained using positive reinforcement – as opposed to

negative reinforcement – techniques were more pessimistic on a judgement bias task (Freymond, Briefer, Zollinger, Gindrat-von Allmen, Wyss, & Bachmann, 2014).

One justification made for these findings has been that those animals who had experienced receiving rewards had become 'tolerant' to the availability of food rewards and were less 'motivated' to work for food under uncertain conditions (Spruijt et al., 2001). Thus, the horses and dogs were unwilling to risk wasting energy to obtain a reinforcer that was abundant and may or may not be in that location. An alternative explanation given for the results found in the Freymond et al. (2014) study was that release from the condition of negative reinforcement training generated positive affect. In other words, history of welfare affected future welfare.

1.5.3. History effects. Previous experience of housing or conditions (while being raised/prior to experiment) may play a role in judgement bias (Wichman et al., 2012). If an animal has only ever experienced 'barren' housing, then exposing it to this environment may not evoke a shift in responding. In other words, the animal 'doesn't know what it's missing'. When rats never experienced a changeover in conditions from high to low light (poor to good conditions) or low to high light (good to poor conditions), no significant difference was found in judgement bias results between the high versus low light groups (Burman, Parker, Paul, & Mendl, 2009).

Interestingly, when animals are released from aversive conditions prior to testing, they often show positive judgement biases as opposed to the expected negative judgement bias. One example of this was observed in a study of neglected and non-neglected goats in a sanctuary (Briefer & McElligott, 2013). Goats neglected prior to arriving at the sanctuary were compared with those who had previously experienced adequate (according to legislation) housing and nutrition on a typical judgement bias task.

No differences were found in initial learning of the discrimination, although female neglected goats showed a positive judgement bias compared with controls. This suggests that welfare is relative and supports the idea that previous experience of welfare can influence judgement bias results. Similar results were found where: sheep were released from shearing (see Figure 2) (Sanger, Doyle, Hinch, & Lee, 2011); sheep were released from restraint/isolation (Doyle, Fisher,

Hinch, Boissy, & Lee, 2010); and when rats were moved from high light to low light (Burman et al., 2009).

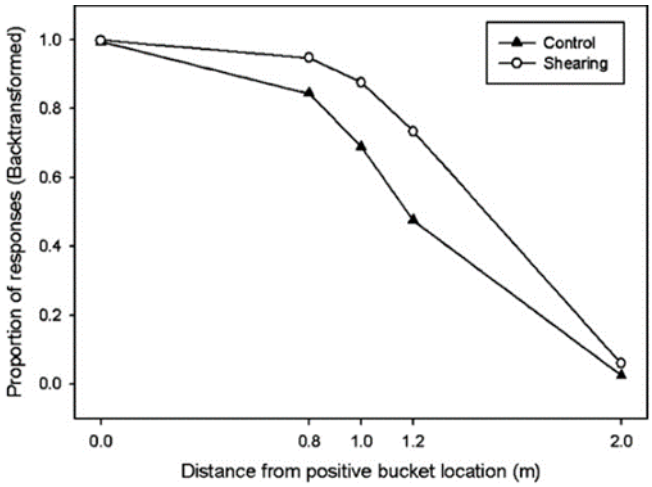

Figure 2. Proportion of responses (approach to bucket) to a CS+ bucket in a location associated with reward, a CS- bucket in a location associated with an aversive stimulus (dog), and three novel buckets placed in intermediate locations to the CS+ and CS- under control conditions and following an aversive event (shearing). Figure retrieved from "Sheep exhibit a positive judgement bias and stress-induced hyperthermia following shearing," by M. E. Sanger, R. E. Doyle, G. N. Hinch, and C. Lee, 2011, Applied Animal Behaviour Science, 131, p. 94-103.

The opposite effect is found when animals are removed from enriched conditions. In the Burman et al. (2009) study, rats moving from low light to high light showed the strongest negative judgement bias (stronger than rats who had only ever experienced high light). Pigs moved from enriched to barren conditions also showed a more negative judgement bias than pigs who were only ever placed in barren housing (Douglas, Bateson, Walsh, Bédué, & Edwards, 2012). Within the same study, enriched pigs showed a more positive judgement bias than barren pigs; however, the authors note that all pigs were in housing that met minimum requirements (barren) before the experiment began. Therefore, this effect may have been a result of the changeover from standard housing to enriched experimental housing (Douglas et al., 2012).

Where this changeover effect is not seen, the aversive events may have had long term impacts on welfare. For example, Daros et al. (2014) found that calves showed a negative judgement bias for a period after dehorning and separation from the dam.

1.5.4. Repeated measures effects. As probes in judgement bias testing must remain neutral, in order to avoid influencing responding towards or away from the positive or negative stimuli, they are often non-reinforced. One common limitation expressed in the judgement bias literature is that learning about ambiguous stimuli may have resulted in a change in responding to those stimuli, over multiple judgement bias tests. In other words, responding on probes may extinguish over time not due to exposure to poor welfare, but as a result of learning that responding on probes does not earn a reinforcer. Sheep trained on a go/no-go spatial discrimination task were tested over a period of 3 weeks on a judgement bias task using three ambiguous probes (Doyle, Vidal, Hinch, Fisher, Boissy, & Lee, 2010). No manipulation was used. It was found that sheep displayed an increasing bias away from ambiguous probes, in a pattern consistent with learning about the outcomes of those probes (Doyle et al., 2010).

A similar study investigated the influence of diazepam (an anti-anxiety drug) on judgement bias in lambs (Destrez, Deiss, Belzung, Lee, & Boissy, 2012). A similar go/no-go spatial discrimination task was used. Half of the lambs were injected with saline, while the other half received 0.10 mg/kg diazepam. All lambs were tested on a judgement bias test 10 minutes and then 3 hours following injections. While the control lambs showed an increased latency to approach ambiguous probes in the latter test, diazepam lambs showed no difference in test results. This suggests that the controls learned about the non-reinforcing consequences of the ambiguous probes, and that diazepam had an effect on learning in the treated lambs. Considering that diazepam is an anti-anxiety drug, it is possible that the diazepam diminished the aversive properties of non-reinforcement, as was observed by Terrace (1963).

While some influence from repeated measures effects may be expected when multiple tests are conducted, it is also important that the predicted effect of the welfare manipulations used in judgement bias research have some evidence base.

1.5.5. What is aversive or pleasurable? Some researchers have made the assumption that ‘barren’ housing, for example, is aversive. This assumption anthropomorphises the animal in question, rather than being based on scientific evidence (Stamp Dawkins, 2008). Similarly, differences between conditions may not be significant enough to produce a bias (Wichman et al., 2012). Current welfare standards dictate the ‘bare minimum’ for housing of animals; while

welfare is often assessed through observation of abnormal behaviours and analysis of each component implicating welfare (e.g. nutrition) (Mellor, 2014). These measurements do not provide details on what enrichments have the largest effect on welfare, nor whether absence of enrichments leads to negative affective states. Some knowledge of whether the stimulus/situation is aversive or preferred, that is independent to the judgement bias test, is required in order to support the claim that the effects seen are a result of the valence of the events rather than some external variable.

White noise as aversive stimulus. Some sounds have been shown to be aversive for certain animals, at certain amplitudes (dB). In a study on sound avoidance, a number of sounds were played only when chickens were standing on one side of the chamber (MacKenzie, Foster & Temple, 1993). This meant that the birds could 'turn off' those sounds that were aversive by moving to the alternate side. For some sounds, birds had the sound off for the majority of the session, including: a dog barking; a chicken stimulated by food; a chicken who had sighted a dog; chickens fighting; a commercial poultry shed; and a ventilator (MacKenzie et al., 1993). This suggests that these sounds were aversive at the 90dB level.

White noise has previously been used at the Learning Behaviour Welfare Research Unit as an aversive stimulus at the 90-105dB range. Here, chicks have shown a conditioned place preference for no noise/food as compared with white noise (Jones, Bizo, & Foster, 2012); while hens' demonstrated response biases away from white noise when it was available on one alternative in concurrent schedules (McAdie, 1998). There is some evidence to suggest that these effects are temporally mediated, as chickens exposed to noise at 100dB for 28 minutes showed significantly heightened biochemical stress indicators for only the first 14 minutes (Bedáňová, Chloupek, Vošmerová, Chloupek, & Večerek, 2010).

Not only does this evidence demonstrate that chickens can hear noise at the 100dB level, but also indicates that white noise may be an appropriate stimulus to allow the investigation of effects of aversive stimuli on judgement bias test results.

1.6. Tonic Immobility

Any idiosyncratic differences seen in judgement bias test results may have been influenced by external variables, other than the independent variable. Variation in traits between individuals is to be expected in nature and this

provides a functional purpose, as it allows species to adapt to change. In the context of judgement bias, this means that some individuals may be more 'emotional' or reactive than others.

Tonic immobility refers to a temporary state of "non-responsiveness", produced by gentle physical restraint (Jones & Faure, 1981), and is proposed to be a behavioural measure of 'fearfulness' or 'emotionality' in chickens (Gallup, Ledbetter, & Maser, 1976). An assessment of tonic immobility normally involves placing birds on their backs on a stable surface (for example, a table or U-shaped cradle) for a brief period of time until the bird remains still. A stopwatch measures the time taken until the bird moves or rights itself.

A study into the effect of exposure to different calls made by conspecifics showed that when birds were exposed to recorded calls associated with danger prior to tonic immobility testing, their righting times were longer than following calls unrelated to danger (Jones, 1986). Thus, Jones argues that such stressors prolong the tonic immobility response. Research using differing breeds of hens has shown that some breeds (White Leghorns, a layer bird) consistently present with longer tonic immobility response times than others (Production Red, used for both laying and meat), suggesting that a genetic component to 'emotionality' does exist (Gallup et al., 1976). As judgement bias is also purported to be a measure of animal emotion, we might expect the length of the tonic immobility response to be correlated with the degree and direction of bias seen during judgement bias testing. Thus, a tonic immobility test could be used to clarify judgement bias findings with chickens when aversive stimuli are used.

1.7. Research Question

Given the literature presented in the introduction, it is clear that more data are needed to help clarify the conflicting results. Thus, the aim of the current study was to reproduce both bias effects (positive and negative) seen in the literature, using three short-term white noise manipulations: exposing the subjects to white noise prior to the test; exposing subjects to white noise during the test; and interrupting the test with a period of white noise (during which time the test was paused). As white noise has not previously been used in the judgement bias literature, the current experiment also examined whether white noise would produce a bias in responding at the 100dB level.

Though food was used as a reinforcer, it was not used as an independent variable to produce a bias (for example, by pre-feeding prior to judgement bias

testing) and birds were not able to access more than their required daily intake of food within a session. This controlled for the possibility of satiation. Sessions where judgement bias testing occurred were kept to a minimum to reduce the influence of repeated-measures effects and the aversive event (exposure to white noise) was set at 15 minutes to avoid habituation to the noise within a session.

Predictions were that a negative bias would be seen when subjects were exposed to white noise for the duration of the test (Phase 2), while a positive bias would be seen when the white noise was turned off prior to testing (Phase 1). Prior to the noise exposure in Phase 3, it was expected that birds would respond to stimuli in a similar fashion to Baseline 1; while following the noise, birds would show a positive bias as with Phase 1. Birds who showed the longest tonic immobility response times were expected to show more negative (or less positive) judgement biases than birds with shorter tonic immobility response times.

2. Methods

2.1. Subjects

Subjects were five domestic chickens (*gallus gallus domestics)* of mixed breeds, including four Bantam cross (10.1-10.4) and one Buff Orpington hen (10.5). Of these, one was a rooster (10.1) and four were hens (10.2-10.5). Ages ranged from 24 months to 4 years. There were initially equal numbers of hens and roosters, though two of the roosters were not feeding effectively from the magazine during early training and so were removed from the experiment. All of the chickens had some experience with feeding from a magazine, with the rooster and one hen being experienced in key pecking and the remainder being experienced in screen pecking.

Chickens were kept at 80% of their free-feeding weight for the duration of the study, and were housed individually in wire cages at the Learning Behaviour Welfare Research Unit (LBWRU). Cages met industry standards for size and durability, and chickens had permanent access to water via tubing running along the top of cages. Ethics approval for the project was granted by the University of Waikato Animal Ethics Committee (Protocol # 957).

2.2. Apparatus

Experimental chambers measured approximately 800mm wide x 500mm high x 500mm deep. One side wall of the chamber was hinged, to allow access to the chamber. The front end of the chamber held a 17” monitor (1280x1024 pixels) with an infrared overlay, and a magazine below (see Figure 3). The infrared overlay allowed pecks to the screen to be detected, and coordinates of each peck to be recorded. The magazine was changed out partway through training, as one hen was able to access the reinforcer when the magazine was closed.

A speaker was placed at the centre rear of the chamber, so that chickens of different sizes would experience the aversive noise approximately to an equivalent volume. Plastic mats were placed on the floor of the chamber, to allow for easy cleaning. Three hooks were placed in three different locations on the ceiling of the chamber for testing the noise level in decibels (dB) close to the speaker, in the centre of the chamber and close to the screen. The experimental program was developed using Delphi version XE8.

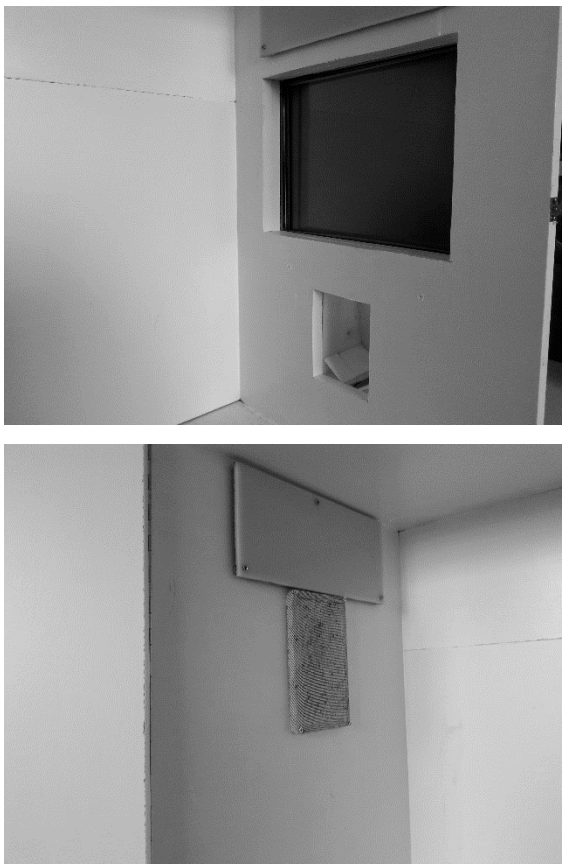

Figure 3. Top figure shows front of chamber, with air vents at top, monitor and infrared overlay in centre and magazine below. Bottom figure shows rear of chamber, with air vents at top and speaker below.

2.3. Procedure

2.3.1. Training. All chickens were magazine trained before being exposed to reinforcement on a rich variable-interval (VI) schedule for pecking two locations on an infrared overlay. Initially, the stimuli were red circles placed halfway down the screen, on the far left or the far right. Schedules began at VI 2-s VI 2-s and were gradually increased to VI 60-s VI 60-s. Once chickens were pecking both sides of the screen reliably, the circles were changed to white and discrimination training began. Three hens did not peck the circles initially, and were shaped to peck a single red key in a different chamber. When placed in the experimental chamber once again, one hen generalised the key-peck training. The remaining two hens were shaped in the experimental chamber, to peck the circles on the screen.

Discrimination training involved successive presentations of the S+ and the S- (locations for S+ and S- were counterbalanced across chickens). Here, the S- was initially reinforced under a variable-interval 80-s schedule and was gradually increased to VI 600-s. The S+ remained on a VI 60-s schedule. For

three of the chickens (10.1, 10.2 and 10.3) S+ was on the left hand key, and for the remainder of the chickens (10.4 and 10.5) S+ was on the right hand key. Component length was set at 30 seconds, with an inter-trial interval of 10 seconds. The degree of responding across the two stimuli was the measure of discrimination used. The criterion for achieving discrimination was five non-consecutive days of training where the proportion of responding on the rich schedule was greater than .8.

2.3.2. Baselines. Two baseline judgement bias testing sessions were conducted. The first baseline (Baseline 1) was carried out prior to the three manipulations, and the second baseline was conducted following the manipulations (Baseline 2). No white noise was played during either of the baseline sessions. Sessions ran for 45 components, consisting of 30 training components and 15 probe components. Training components were 15 presentations of each of the S+ and S-, reinforced on a VI 60-s schedule and VI 600-s schedule respectively. Pecks to probe stimuli had no programmed consequences (set on an extinction schedule) and were presented on every third component. Component length was set at 30 seconds, with an inter-trial interval of 10 seconds. Sessions ran for approximately 30 minutes.

2.3.3. Phase 1. Phase 1 involved 15 minutes exposure to white noise, played on a speaker at the rear of the cage. Exposure was limited to 15 minutes, as decided by the University of Waikato Animal Ethics Committee, based on evidence that stress indicators were only significant for approximately the first 15 minutes of exposure to noise at 100dB (see Bedáňová et al., 2010), and considering an approach where minimal exposure was used. Noise level was set at 100dB and was measured in three places (close to speaker, in the centre of the chamber, and close to the screen) using a dB meter at the beginning of each test day. During this time, the chicken was not able to receive reinforcement by pecking at the screen. Immediately following the noise exposure, a judgment bias test commenced. This followed the same procedure as outlined above for baselines. Sessions ran for approximately 45 minutes.

2.3.4. Phase 2. Phase 2 involved overlaying white noise at 100dB at the same time as the judgement bias test was conducted. Phase 2 testing was repeated three times, with each session containing 18 components (six presentations of each of the S+ and S-, and two presentations of each of the probe stimuli) in order

to meet ethics criteria. Component length and inter-trial intervals were set the same as in baselines and Phase 1. Sessions ran for approximately 11 minutes.

2.3.5. Phase 3. Phase 3 began with 24 components of the judgement bias test, as mentioned for baselines, and was interrupted before the 25th component with 15 minutes exposure to white noise at the 100dB level. During this time, the chicken was not able to receive reinforcement by pecking at the screen. A 10-second blackout followed the white noise exposure, before the judgement bias test resumed on the 25th component.

2.3.6. Maintenance. Between each judgement bias test, subjects were required to complete a minimum of three maintenance sessions under the same conditions as discrimination training. Here, the S+ was on a VI 60-s schedule and the S- was on a VI 600-s schedule as with judgement bias testing. Component length was set at 30 seconds, with an inter-trial interval of 10 seconds. Criterion for moving on to a new testing session was three non-consecutive days of maintenance where the proportion of responding on the rich schedule was greater than .8.

2.3.7. Tonic immobility testing. A tonic immobility test was also carried out once for each of the experimental subjects. A 'cradle' was made, using two rolled up towels, with another laid over the top and an extra towel to rest the chickens head at the end of the cradle (see Figure 4). Testing involved laying the chicken on its back in the cradle, with a hand cupped over the head and another hand restraining just above the legs. Restraint lasted for 20 seconds. If a chicken did not become immobile within this time, then inversion was repeated, up to five times. Chickens were considered immobile if they laid still for 10 seconds without restraint. Time was recorded from release to the time that birds righted themselves. Birds were left up to a maximum of 20 minutes in the immobile state.

Figure 4. Cradle used in tonic immobility testing. Left image shows cradle with no bird and right image shows cradle with 10.3 in an immobile state.

3. Results

3.1. Training

Response rates for the last four days of training where schedules were equal (VI 60-s) on the left and right stimuli are shown in Figure 5. Birds 10.1, 10.2 and 10.4 exhibited a greater rate of responding on the right hand stimulus for three of the final training days, while 10.3 and 10.5 had approximately equal rates of responding across training days. A two-way repeated measures ANOVA, conducted for response rate across left and right hand stimuli, showed no significant within-subjects effect of stimulus or day (see Table 1), and no significant interaction for stimulus by day (see Table 1).

Response rates on S+ and S- stimuli across variable-interval schedule changes are shown in Figure 6, across discrimination training days. The vertical dotted lines indicate a change in variable-interval schedule. Response rates on S+ remained fairly stable across schedule changes for 10.1 and 10.2; though rates for 10.3, 10.4 and 10.5 trended upwards across schedule changes. Responding on S- trended downwards for all birds. A separation between response rates on the S+ and S- began around VI 120-s for 10.3, VI 160-s for 10.1, VI 240-s for 10.4 and 10.5, and VI 360-s for 10.2. From here, responses rates on the S- dropped down to almost zero after the final schedule change for 10.1, 10.2, 10.3 and 10.5. Bird 10.4 maintained around five responses per minute after the final schedule change.

Figure 7 shows the proportion of responses made to S+ and S- across schedule changes on discrimination training days; again, the vertical dotted lines indicate a change in variable-interval schedule. Criterion for completion of discrimination training was five or more non-consecutive days where proportion of responses to S+ was greater than 0.8. Proportion of responding on S+ trended upwards for all birds, though birds reached criterion during different variable-interval schedules. Birds 10.1, 10.2 and 10.4 reached criterion at VI 600-s. Bird 10.3 reached criterion at VI 400-s and 10.5 reached criterion at VI 480-s (10.5 consistently responded to S+ for over 0.8 of all responses during VI 400-s, though the subject was only exposed to this condition for three days). Thus, 10.3 and 10.5 were the quickest to acquire the discrimination.

3.2. Judgement Bias Testing

3.2.1. Number of responses. Number of responses, for each stimulus, was the sum of responses emitted across all components for that stimulus. For S+ and S-, number of responses was derived from every third component as there were 15

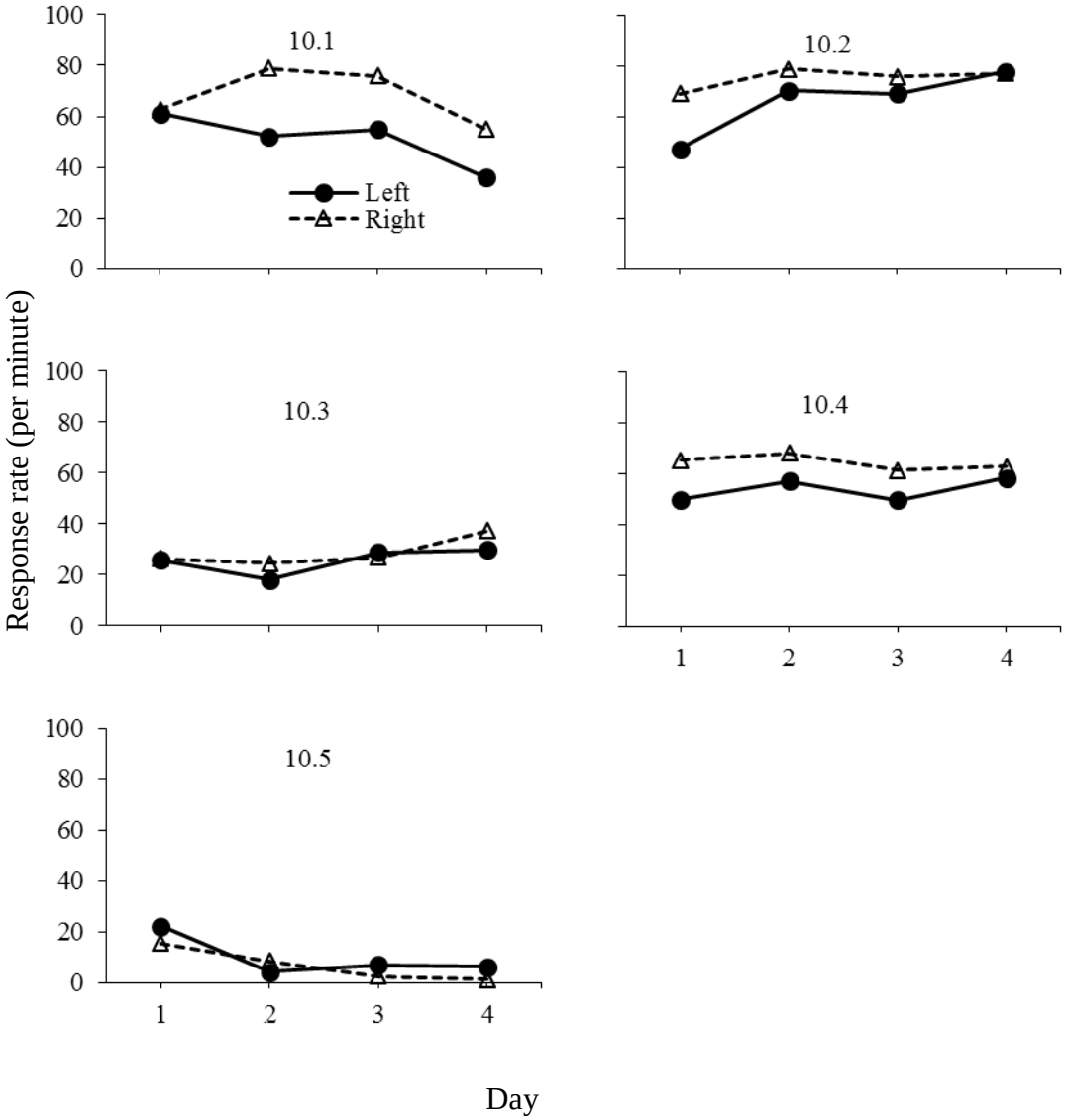

Figure 5. Response rates on the left and right stimuli for each of the last four days of training where both schedules were VI 60-s.

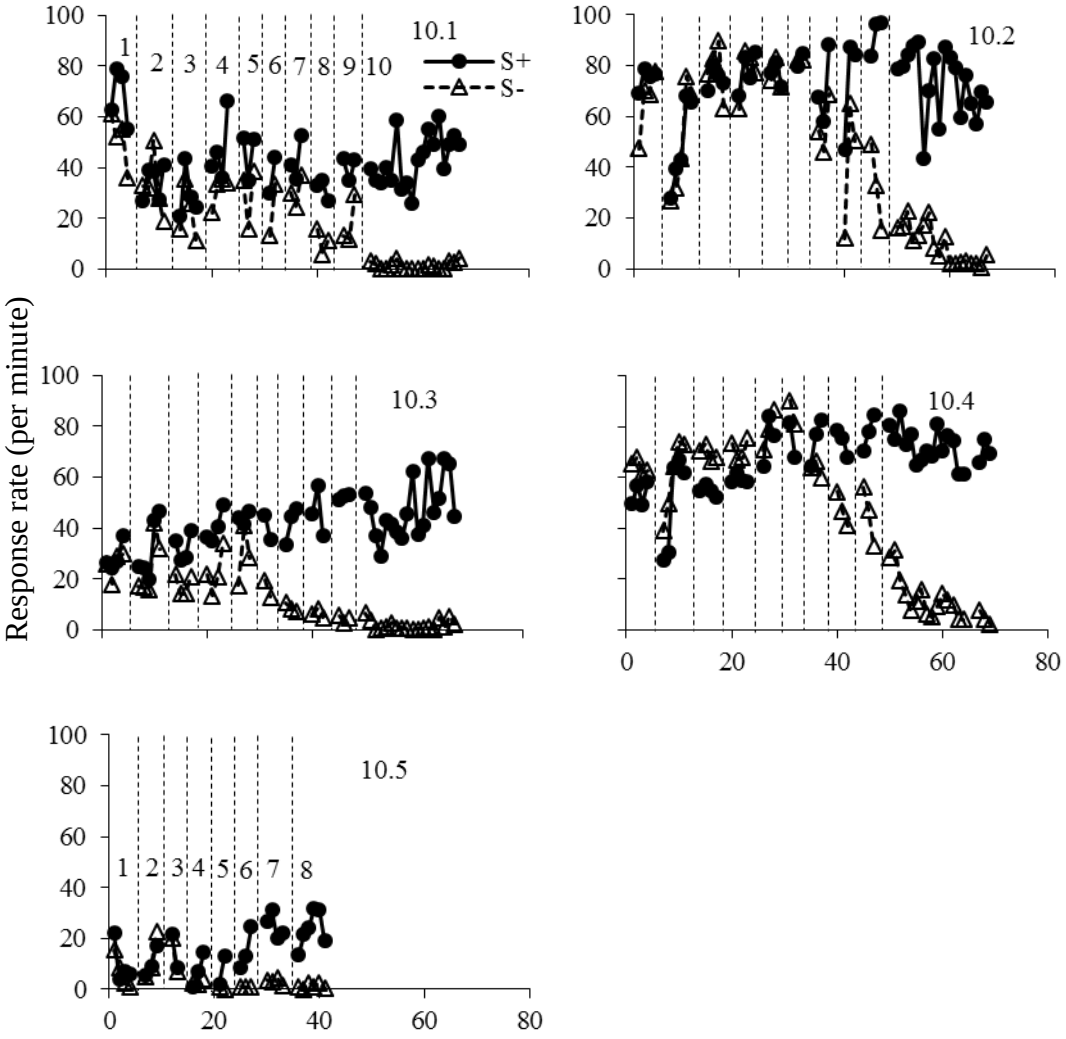

Figure 6. Response rates on S+ and S- during discrimination training days. Dotted lines indicate a change in variable-interval schedule on S- (located on the left side for 10.1 through 10.3, and on the right side for 10.4 and 10.5), beginning with VI 60-s and ending on VI 600-s. Refer to Table 2 and Table 3 to see variable-interval schedules for each number.

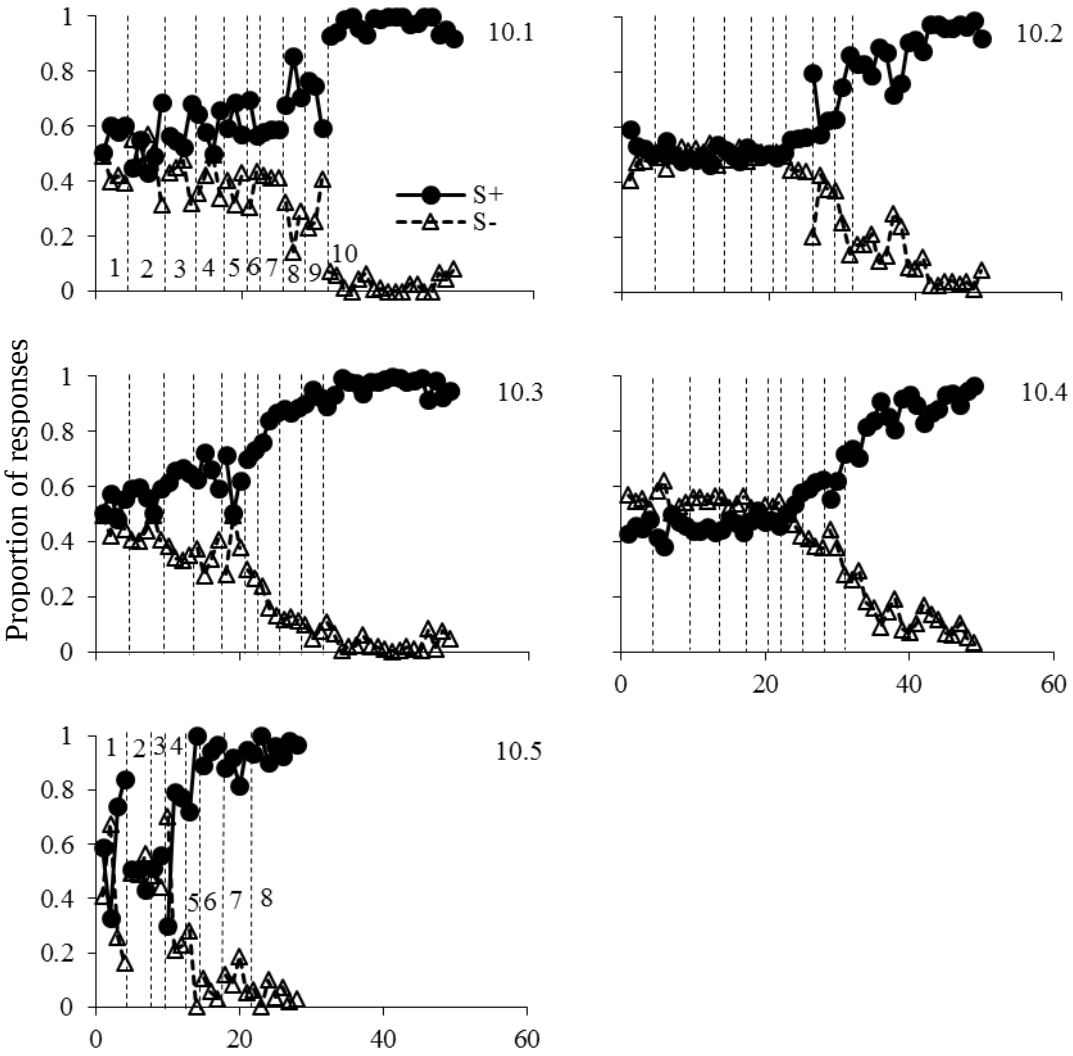

Figure 7. Proportion of responses on S+ and S- during discrimination training days. Dotted lines indicate a change in variable-interval schedule on S- (located on the left side for 10.1 through 10.3, and on the right side for 10.4 and 10.5), beginning with VI 60-s and ending on VI 600-s. Refer to Table 2 and Table 3 to see variable-interval schedules for each number.

Table 1

Two-way repeated-measures ANOVA results for Figure 5. Response rates on left and right stimuli for each of the last four days of training where both schedules were VI 60-s.

	df	F	*p*	partial η2
Response Rate				
Figure 5.				
Side	1, 4	4.73	.095	.54
Day	3, 12	0.06	.981	.01
Interaction	3, 12	0.68	.579	.15

Table 2

Variable-interval schedule changes for birds 10.1 through 10.4 in: a) Figure 6. Response rates on S+ and S- during discrimination training days; and b) Figure 7. Proportion of responses on S+ and S- during discrimination training days.

	Variable-Interval Schedule	
Period	Rich	Lean
1	60-s	60-s
2	60-s	80-s
3	60-s	120-s
4	60-s	160-s
5	60-s	200-s
6	60-s	240-s
7	60-s	360-s
8	60-s	400-s
9	60-s	480-s
10	60-s	600-s

Table 3

Variable-interval schedule changes for bird 10.5 in: a) Figure 6. Response rates on S+ and S- during discrimination training days; and b) Figure 7. Proportion of responses on S+ and S- during discrimination training days.

	Variable-Interval Schedule	
Period	Rich	Lean
1	60-s	60-s
2	60-s	80-s
3	60-s	160-s
4	60-s	240-s
5	60-s	360-s
6	60-s	400-s
7	60-s	480-s
8	60-s	600-s

components for each of the S+ and S- but only five presentations for each of the probes. Data for Phase 2 were derived from three separate sessions, each containing 18 components. Number of responses for each bird in Phase 2 was derived from the first two sessions' data, plus data from nine components from the final day, to make a total of 45 components. Data for Phase 3 across was derived from the entire session (both before and after the noise).

The mean across birds, and number of responses for each bird, was calculated for each of the stimuli under each of the experimental conditions (see Figure 8). A negative gradient was observed in the mean data across conditions, with the greatest number of responses on S+ and the least on S- in each phase. This suggests that number of responses emitted on each probe stimulus was moderated by its location relative to S+ and S-. As with the mean data, there was a negative slope for 10.2, 10.4 and 10.5 in all conditions. Birds 10.1 and 10.3 had negative slopes for Baseline 1, Phase 1 (noise before test), and Phase 3 (noise interrupting test), but not for Phase 2 (noise during test).

On average, Baseline 1 produced the greatest number of responses on S+ and Probe 1 stimuli. Responding on other probes and S- stimuli was similar across Baseline 1, Phase 1 and Phase 3. For 10.2, 10.4 and 10.5, more responses were emitted on each of the stimuli during Phase 1 than in Baseline 1, a contrast with the mean data.

For Phase 1 (noise before test), there were lower mean numbers of responses emitted on S+, Probe 1 and Probe 2 stimuli than during Baseline 1. Similarly, few responses were emitted by 10.1 during Phase 1; and Phase 1 resulted in fewer responses to each stimulus for 10.3, but particularly for Probe 2, Probe 3 and S- stimuli. Number of responses was greater for 10.1 during Phase 1b, and followed the mean data path more closely. The remaining three birds emitted a greater number of responses during Phase 1 than Baseline 1. During Phase 1, a greater number of responses were emitted by 10.2 to all three probe stimuli and S-. This produced a flatter gradient with more even distribution of responses than in Baseline 1. For 10.4, a greater number of responses were emitted on Probe 1 only. Two peaks are seen in responding for 10.5 at Probe 1 and Probe 3, where there were more responses than during Baseline 1.

The lowest mean number of responses on each stimulus, of all the conditions, was seen in Phase 2 (noise during test). The mean gradient in this phase was slightly flatter than for Baseline 1, Phase 1 and Phase 3 data. Birds 10.1

and 10.3 made almost no responses during this phase. Responses for 10.2 followed a similar data path to the Baseline 1 data, though there were fewer responses in Probe 2 than in Baseline 1. Thus, there was a shift in responding away from probe stimuli. The distribution of responses across stimuli for 10.4 and 10.5 was similar to Baseline 1, but fewer responses were emitted on each of the stimuli. Mean data was an accurate reflection of individual data for Phase 2, where all birds emitted the least responses to each stimulus under this condition.

Mean number of responses in Phase 3 (noise interrupting test) followed the Baseline 1 data path. This was also observed in individual data for 10.1, 10.4 and 10.5. Some shifts in responding were seen for 10.2 and 10.3. Bird 10.2 emitted a greater number of responses on the Probe 3 and S- stimuli than in Baseline 1, thus responding shifted towards probes, creating a flatter gradient. Fewer responses were emitted by 10.3 on S+, Probe 1 and Probe 3 stimuli but not Probe 2 or S- stimuli.

A two-way repeated measures ANOVA was conducted for number of responses made across all conditions, using the first exposure to Phase 1 for 10.1. Mauchly's test indicated that the assumption of sphericity had been violated for stimulus X^2 (9) = 23.77, p = .014. Degrees of freedom for this effect were corrected using Greenhouse-Geisser estimates (ε = .32). Results showed a significant within-subjects effect of stimulus location and condition (see Table 4). No significant interaction was found for stimulus by condition (see Table 4). Results were similar when the last exposure to Phase 1 for Bird 10.1 was used (see Table 4), and when only data from the probe stimuli were analysed (see Table 5). These findings indicate that subjects responded to stimuli differently and that the manipulation of noise exposure was correlated with a significant change in number of responses for at least one condition.

Paired samples t-tests were conducted for each bird, comparing means under Baseline 1 with means for each phase. A significant difference was found between means for Baseline 1 (M=111.40, SD=91.47) and Phase 1 (M=35.20, SD=42.29) for 10.3; t(4) = 3.27, p = .031, and between means for Baseline 1 (M=110.60, SD=70.62) and Phase 2 (M=74.40, SD=63.05) for 10.4; t(4) = 5.00, p = .007. All other comparisons showed no significant differences in means (see Table 6).

Cohen's d was calculated for each of the comparisons. According to Cohen (1992), the strength of an effect for Cohen's d is as follows: around .2 =

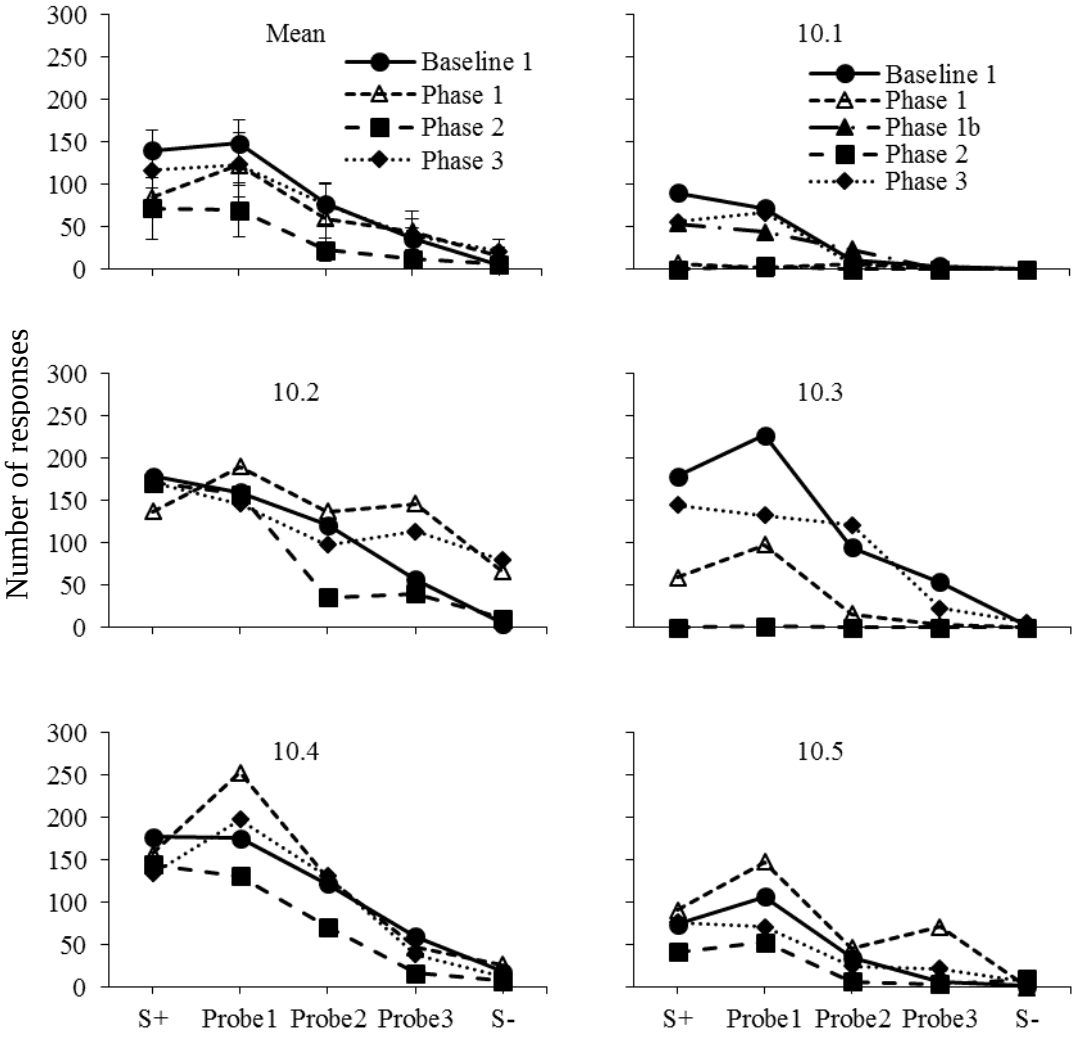

Figure 8. Number of responses on each of the stimulus locations, for each bird, during each of the experimental conditions (Phase 1 = noise before test, Phase 2 = noise during test, Phase 3 = noise interrupting test). Mean across birds is shown in top left. Bird 10.1 was exposed to Phase 1 twice, and thus has an extra condition (Phase 1b).

Table 4

Two-way repeated-measures ANOVA results for number of responses on each of the stimulus locations, across birds in: Figure 8. during each of the experimental conditions; Figure 9. during each of the three session days for Phase 2; and Figure 10. during baselines taken before and after manipulations.

	df	F	*p*	partial η2
Number of Responses				
Figure 8.				
First Exposure Phase 1				
Stimulus Location	1.28, 5.11	20.43	<.001	.84
Condition	3, 12	4.17	.031	.51
Interaction	12, 48	1.90	.058	.32
Second Exposure Phase 1				
Stimulus Location	4, 16	24.54	<.001	.86
Condition	3, 12	4.87	.019	.55
Interaction	12, 48	1.85	.066	.32
Figure 9.				
Stimulus Location	1.14, 4.56	4.07	.104	.50
Session	2, 8	2.91	.112	.42
Interaction	8, 32	0.58	.787	.13
Figure 10.				
Stimulus Location	1.13, 4.52	19.00	.009	.83
Baseline Time	1, 4	0.19	.688	.05
Interaction	4, 16	0.94	.465	.19

Table 5

Two-way repeated-measures ANOVA results for number of responses (only probes) on each of the stimulus locations, across birds in: Figure 8. during each of the experimental conditions; Figure 9. during each of the three session days for Phase 2; and Figure 10. during baselines taken before and after manipulations.

	df	F	*p*	partial η2
Number of Responses				
Figure 8.				
First Exposure Phase 1				
Stimulus Location	2, 8	18.68	.001	.82
Condition	3, 12	4.30	.028	.52
Interaction	6, 24	1.02	.434	.20
Second Exposure Phase 1				
Stimulus Location	2, 8	22.38	.001	.85
Condition	3, 12	5.04	.017	.56
Interaction	6, 24	1.13	.377	.22
Figure 9.				
Stimulus Location	2, 8	4.37	.052	.52
Session	2, 8	2.88	.114	.42
Interaction	4, 16	1.50	.248	.27
Figure 10.				
Stimulus Location	2, 8	32.35	<.001	.89
Baseline Time	1, 4	0.92	.392	.19
Interaction	2, 8	0.85	.463	.18

Table 6

Paired samples t-test for each bird in Figure 8. Number of responses on each of the stimulus locations, for each bird, during each of the experimental conditions.

		M	SD	*t*	*df*	*p*	*d*
Figure 8.							
10.1	Baseline	35.40	42.29				
	vs. Phase 1a	3.20	3.11	1.77	4	.152	.79
	vs. Phase 1b	24.40	24.75	1.21	4	.294	.54
	vs. Phase 2	0.80	1.79	1.87	4	.135	.83
	vs. Phase 3	26.40	32.35	1.42	4	.228	.64
10.2	Baseline	104.40	72.56				
	vs. Phase 1	135.20	44.48	-1.39	4	.237	.62
	vs. Phase 2	83.20	75.18	1.29	4	.265	.58
	vs. Phase 3	121.80	36.68	-0.86	4	.437	.39
10.3	Baseline	111.40	91.47				
	vs. Phase 1	35.20	42.29	3.27	4	.031	1.46
	vs. Phase 2	0.40	0.89	2.73	4	.052	1.22
	vs. Phase 3	85.60	65.73	1.27	4	.274	.57
10.4	Baseline	110.60	70.62				
	vs. Phase 1	123.00	91.06	-0.726	4	.508	.32
	vs. Phase 2	74.40	63.05	5.00	4	.007	2.24
	vs. Phase 3	102.80	76.07	0.69	4	.530	.31
10.5	Baseline	45.00	44.94				
	vs. Phase 1	71.2	54.76	-2.23	4	.090	1.00
	vs. Phase 2	23.20	22.62	1.98	4	.119	.89
	vs. Phase 3	40.20	31.76	0.55	4	.612	.25

small; around .5 = moderate; and around .8 or above = strong. Large effect sizes (> .8) were found for comparisons between Baseline 1 and Phase 2 for 10.1, 10.3, 10.4, and 10.5 (see Table 6), with a moderate effect size (d = .58) for 10.2. Moderate to large effect sizes were also seen for the comparison between Baseline and Phase 1 for 10.1, 10.2, 10.3, and 10.5 (see Table 6).

The mean number of responses across birds, and number of responses for each bird, were calculated for each of the stimuli across the three sessions for Phase 2 (noise during test) (see Figure 9). Across sessions, all mean data followed a negative slope. Slopes for 10.2 and 10.4 were steeper than were the mean data. Mean distribution of responses was similar between Session 1 and Session 3. This was also observed in the individual data with 10.2, 10.4 and 10.5. Mean number of responses on S+, Probe 1 and Probe 2 was greater in Session 2. Again, this was observed with 10.2, 10.4 and 10.5. Birds 10.1 and 10.3 made few responses during any session.

A two-way repeated measures ANOVA was conducted for number of responses made across Phase 2 sessions. Mauchly's test indicated that the assumption of sphericity had been violated for stimulus, X^2 (9) = 25.36, p = .008. Degrees of freedom for this effect were corrected using Greenhouse-Geisser estimates (ε = .29). Results showed no significant within-subjects effect of stimulus location or session (see Table 4), and no significant interaction for stimulus by session (see Table 4). Results were similar when only data from the probe stimuli were used in the analysis (see Table 5). These findings suggest that repeated exposures did not have a significant effect on responding in this phase.

The mean across birds, and mean number of responses for each bird, was calculated for each of the stimuli across two baseline phases taken prior to and following exposure to experimental manipulations (see Figure 10). Mean number of responses followed a similar data path across both baselines. Birds 10.2, 10.3 and 10.5 emitted fewer responses on probes in Baseline 2, though data followed a similar slope to Baseline 1 for 10.2 and 10.3. Slightly more responses were emitted on probes in Baseline 2 for 10.1.

A two-way repeated measures ANOVA was conducted for number of responses emitted across baselines. Mauchly's test indicated that the assumption of sphericity had been violated for stimulus X^2 (9) = 19.56, p = .048. Degrees of freedom for this effect were corrected using Greenhouse-Geisser estimates (ε = .28). Results showed a significant within-subjects effect of stimulus location (see

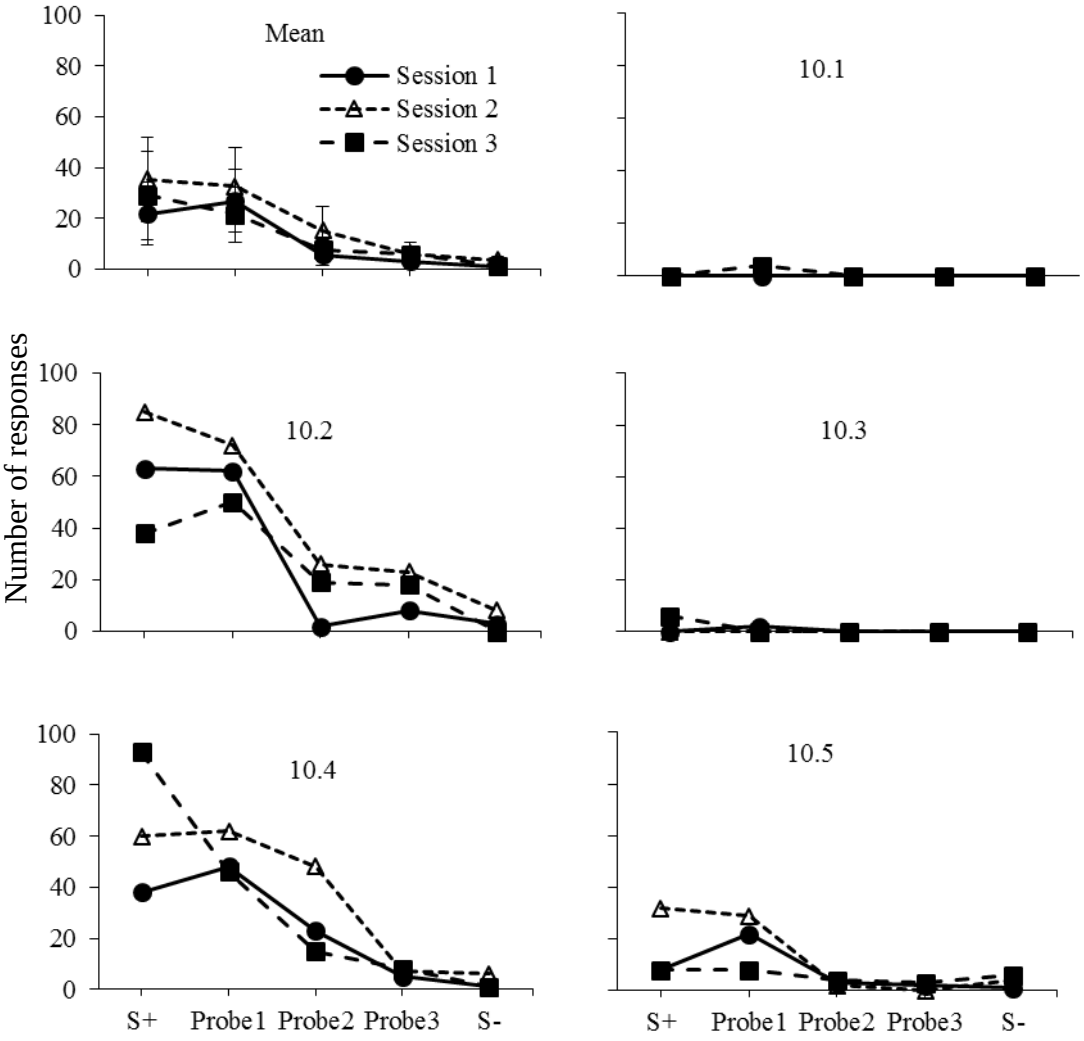

Figure 9. Number of responses on each of the stimulus locations, for each bird, during each of the three session days for Phase 2. Mean across birds is shown in top left.

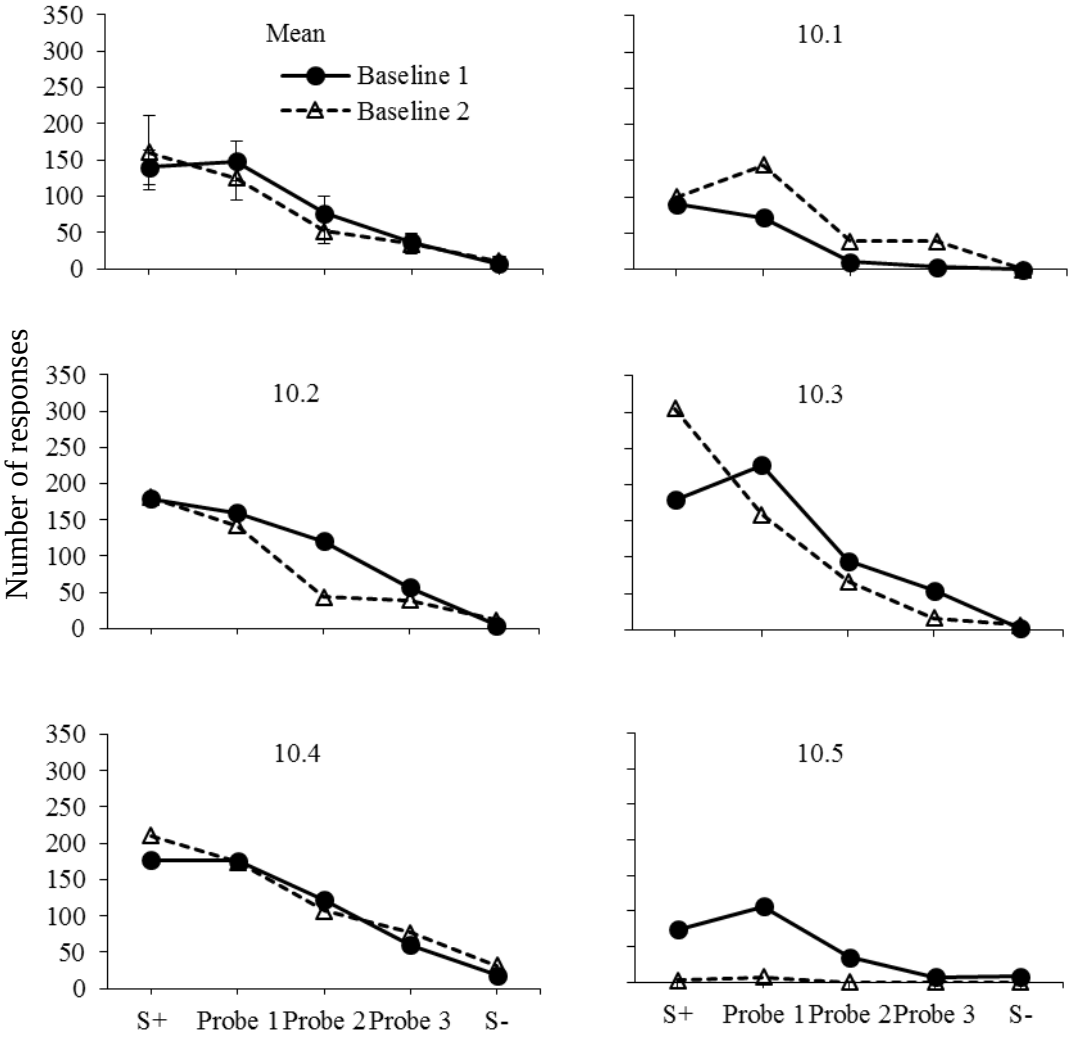

Figure 10. Number of responses on each of the stimulus locations, for each bird, during baselines taken before and after manipulations. Mean across birds is shown in top left.

Table 4). There was no significant within-subjects effect of baseline time and no significant interaction for stimulus location by baseline time (see Table 4). Results were similar when only probes were used in the analysis (see Table 5). Thus, exposure to the three manipulations (Phase 1, Phase 2, and Phase 3) did not have a significant effect on distribution or overall number of responses emitted during baselines.

3.2.2. Response rate. Response rates for Phase 3 (noise interrupting test) were calculated separately for the first 24 components (completed before the noise exposure) and the last 21 components (completed after the noise exposure). Figure 11 shows the mean response rates (per minute) across birds, and response rates for each bird, on each stimulus before and after the noise in Phase 3. Mean response rates followed almost identical data paths for before and after noise. Distributions followed a similar slope to the mean data for most birds. Prior to noise exposure, the gradient for 10.4 was steeper than mean data; while after the noise, 10.2 had approximately equal rates of responding across stimuli.

A two-way repeated measures ANOVA was conducted for response rates in Phase 3. Results showed a significant within-subjects effect of stimulus location (see Table 7). There was no significant within-subjects effect of test time and no significant interaction for stimulus location by test time (see Table 7). Results were similar when only data for probe stimuli were used in the analysis (see Table 8). These findings indicate that the exposure to noise did not have a significant impact on response rates between test times.

3.2.3. Proportion of responses. Proportion of responses, for each stimulus, was derived from the sum of responses emitted across all components for that stimulus divided by the total number of responses emitted within the session. For S+ and S-, the number of responses were derived from every third component within a session. Data for Phase 2 was derived from three separate sessions, each containing 18 components. Proportion of responses for each bird in Phase 2 was derived from the first two days data, plus nine components from the final day, to make a total of 45 components.

The mean proportion across birds, and proportions for each bird, were calculated for each of the stimuli under each of the experimental conditions (see Figure 12). Note that as means were calculated across stimuli, the sum of proportions for each data path did not always equal 1. Mean data paths for Baseline 1 and Phase 1 (noise before test) were almost identical. As with means,

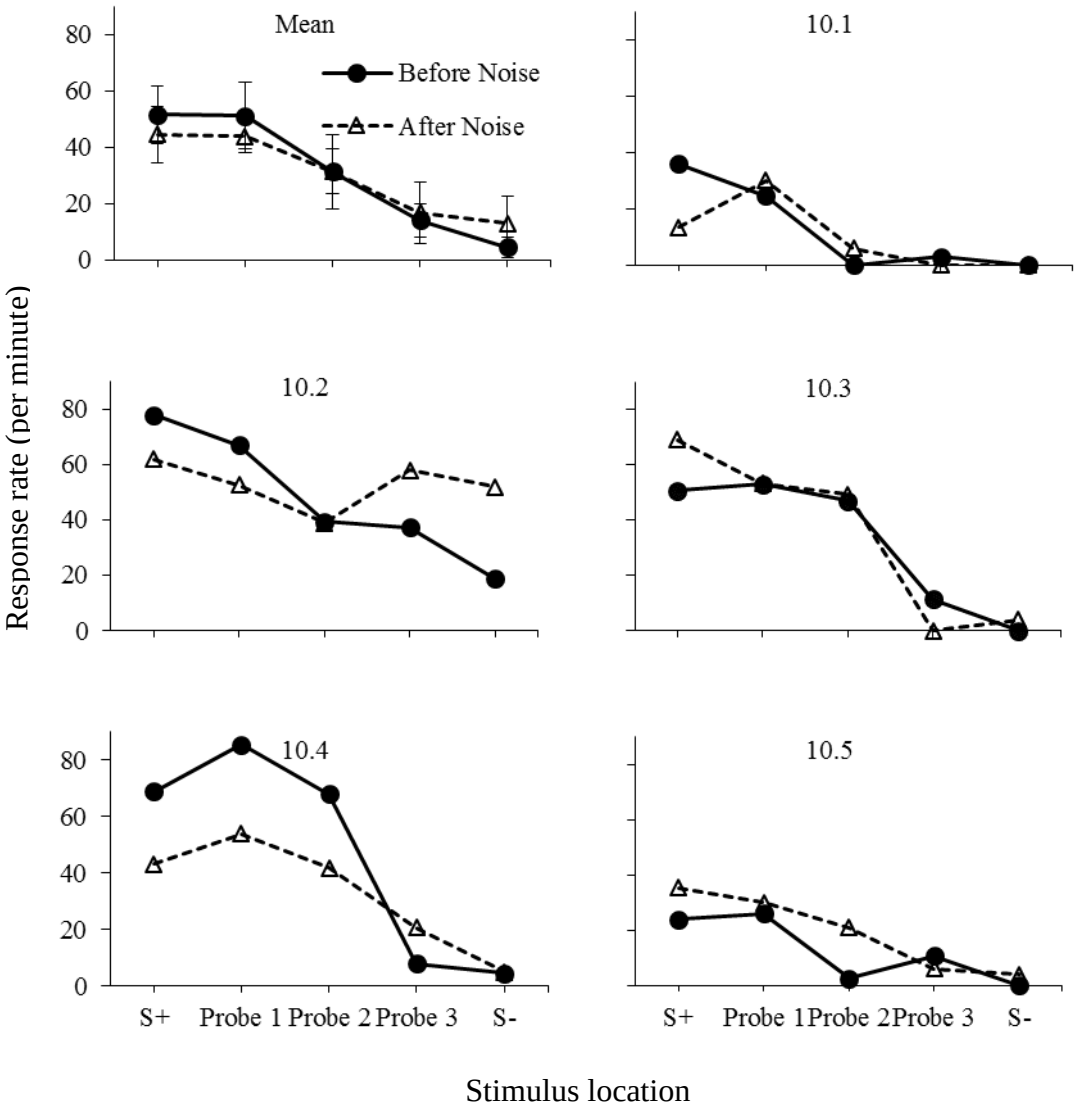

Figure 11. Response rates on each of the stimulus locations, for each bird, before and after the noise exposure in Phase 3. Mean across birds is shown in top left.

Table 7

Two-way repeated-measures ANOVA results for response rate in Figure 11. Response rates on each of the stimulus locations, for each bird, before and after the noise exposure in Phase 3.

Figure 12.	*df*	F	*p*	partial η2
Stimulus	4, 16	12.91	<.001	.76
TestTime	1, 4	0.03	.882	.01
Interaction	4, 16	0.92	.474	.19

Table 8

Two-way repeated-measures ANOVA results for response rate (only probes) in Figure 11. Response rates on each of the stimulus locations, for each bird, before and after the noise exposure in Phase 3.

	df	F	*p*	partial η2
Response Rate				
Figure 11.				
Stimulus Location	2, 8	8.02	.012	.67
Test Time	1, 4	0.16	.713	.04
Interaction	2, 8	0.56	.593	.12

birds 10.3, 10.4, and 10.5 showed little variation in the proportion of responses on each stimulus between Baseline 1 and Phase 1. During Phase 1, 10.2 had a flatter distribution of responses, with a higher proportion of responses on Probe 3 and S- stimuli than in Baseline 1. Bird 10.1 emitted proportionately more responses on Probe 2 and proportionately less responses on Probe 1, in Phase 1 than Baseline 1.

During Phase 2 (noise during test), mean proportions were shifted towards Probe 1, with a lower proportion of responses on S+, Probe 2, and Probe 3 stimuli than in Baseline 1. Similarly, Phase 2 gave the lowest proportion on Probe 2 and Probe 3 stimuli and the greatest proportion on S+ and Probe 1 of all conditions, for each bird. Proportions for 10.1 and 10.3 followed a similar data path to the mean data, with all responses being emitted to the Probe 1 stimulus in Phase 2. This produced a spike in proportion at the Probe 1 stimulus. Phase 3 (noise interrupting test) produced a mean data path similar to Baseline 1, and this was also observed for all birds in the individual data.

A two-way repeated measures ANOVA was conducted for proportion of responses across the probe stimuli, for each of the experimental conditions, using the first exposure to Phase 1 for 10.1. Mauchly's test indicated that the assumption of sphericity had been violated for condition X^2 (5) = 14.40, p = .019. Degrees of freedom were corrected for this effect using Greenhouse-Geisser estimates (ε = .37). Results showed a significant within-subjects effect of stimulus location but not for condition (see Table 9). Though a significant interaction for stimulus location by condition was found (see Table 9). Similar results were found when the last exposure to Phase 1 for 10.1 was used (see Table 9). Thus, there was no significant effect of the noise on proportion of responses.

Figure 13 shows the responses for each bird on each stimulus, in each phase, as a log proportion of Baseline 1. Values are missing where no responses were emitted on that stimulus during that phase. For 10.1, similar proportions of responding were observed under Phase 1b (noise before test) and Phase 3 (noise interrupting test). During Phase 1 (noise before test), 10.1 and 10.3 had lower proportions of responding on all stimuli compared to Baseline 1. Proportions for 10.5 were higher under Phase 1 compared to Baseline 1. For most birds, the lowest proportion of responses were emitted on most stimuli during Phase 2 (noise during test). During this phase, data paths for all of the birds fell below the x axis, meaning that a lower proportion of responses were made during this phase comparative to Baseline 1. Similarly, data paths for most birds

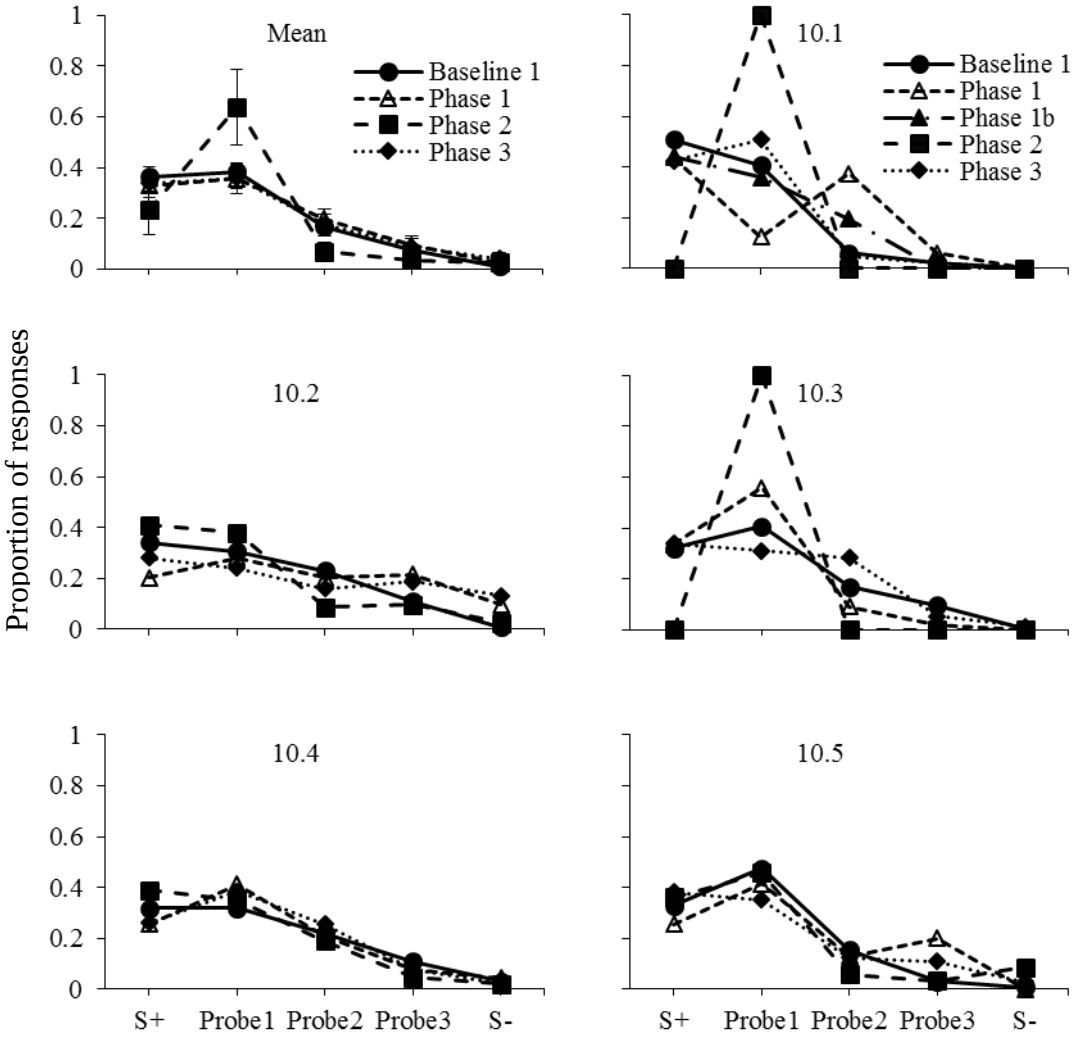

Figure 12. Proportion of responses on each of the stimulus locations, across birds, during each of the experimental conditions (Phase 1 = noise before test, Phase 2 = noise during test, Phase 3 = noise interrupting test). Mean across birds is shown in top left. Bird 10.1 was exposed to Phase 1 twice, and thus has an extra condition (Phase 1b).

Table 9

Two-way repeated-measures ANOVA results for proportion of responses (only probes) on each of the stimulus locations, across birds in: Figure 12. during each of the experimental conditions; Figure 18. during each of the three session days for Phase 2; Figure 19. during baselines taken before and after manipulations; and Figure 20. before and after noise exposure in Phase 3.

	df	F	*p*	partial η2
Proportion of Responses				
Figure 12.				
First Exposure Phase 1				
Stimulus Location	2, 8	23.54	<.001	0.86
Condition	1.10, 4.40	0.72	.452	.15
Interaction	6, 24	2.88	.029	.42
Second Exposure Phase 1				
Stimulus Location	2, 8	20.05	.001	.83
Condition	1.11, 4.43	0.75	.543	.16
Interaction	6, 24	3.81	.008	.49
Figure 18.				
Stimulus Location	2, 8	86.93	<.001	.96
Session	2, 8	0.56	.593	.12
Interaction	1.35, 5.39	0.66	.497	.14
Figure 19.				
Stimulus Location	1.03, 4.13	17.74	.013	.82
Baseline Time	1, 4	0.16	.714	.04
Interaction	1.05, 4.22	1.22	.33	.23
Figure 20.				
Stimulus Location	2, 8	8.23	.011	.67
Test Time	1, 4	0.20	.681	.05
Interaction	2, 8	0.43	.666	.10

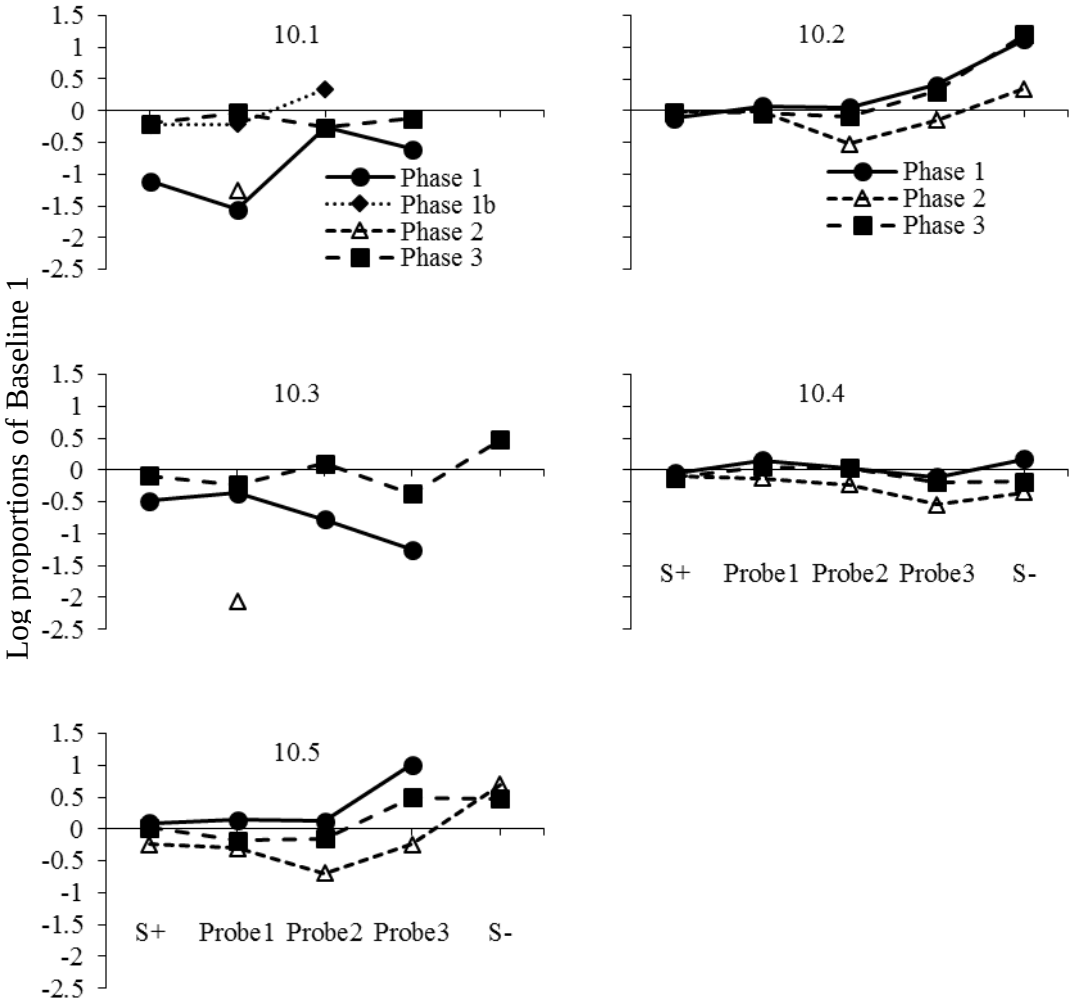

Figure 13. Responses as a log proportion of baseline on each of the stimulus locations, for each bird, over the three probe days (Phase 1 = noise before test, Phase 2 = noise during test, Phase 3 = noise interrupting test). Values are missing where no response was emitted. Bird 10.1 was exposed to Phase 1 twice, and thus has an extra condition (Phase 1b).

in Phase 3 approximately followed the proportion of responses emitted during Baseline 1. Birds 10.2 and 10.5 were the exception, where the data paths for Phase 3 followed Baseline for S+, Probe 1, and Probe 2 stimuli but were higher than Baseline 1 for Probe 3 and S- stimuli. With the exception of Phase 2, proportions approximately followed Baseline 1 for 10.4.

Figure 14 shows the proportion of responses made to the S+ across maintenance and judgement bias testing days, for each bird. Proportion of responses made to S+ remained high for all birds, with a few exceptions. Bird 10.1 made no responses for the first two sessions for Phase 2 (noise during test), and proportion of responses on S+ also dropped below 0.8 during maintenance following the final session for Phase 2. The proportion for 10.2 dropped below 0.8 during Phase 1 (noise before test) and Phase 3 (noise interrupting test). Bird 10.3 emitted no responses on the second session for Phase 2 but otherwise maintained a high proportion on S+. Proportion for 10.5 dropped below 0.8 on the first maintenance day following Phase 1 and again following the first session for Phase 2. Similarly, on the first and last Phase 2 sessions, responding on S+ was below 0.8. Bird 10.4 was the only subject to maintain a proportion above 0.8 across manipulations.

Proportion data were also analysed for the three sessions for Phase 2, the two baselines, and before and after noise exposure in Phase 3 (see Figure 18, Figure 19, and Figure 20 in appendices). Findings were similar to those described for number of responses and for response rate. Results for two-way repeated measures ANOVAs on proportion data are shown in Table 9.

3.2.4. Latency to first response. Latency to first response, for each stimulus, was the sum of all latencies in all components for that stimulus divided by the number of components presented for that stimulus. Where a bird made no response in a component, latency was set to 30 seconds. Data for Phase 2 was derived from three separate sessions, each containing 18 components. Latency to first response for each bird in Phase 2 was derived from the first two days data, plus nine components from the final day, to make a total of 45 components.

The mean latency across birds, and for each bird, was calculated for each of the stimuli in each of the experimental conditions (see Figure 15). Mean data followed a positive slope across conditions, with the shortest latencies on S+ and the longest on S- in each phase, supporting the evidence that a generalisation gradient was produced. Where there were fewer responses on a stimulus in a

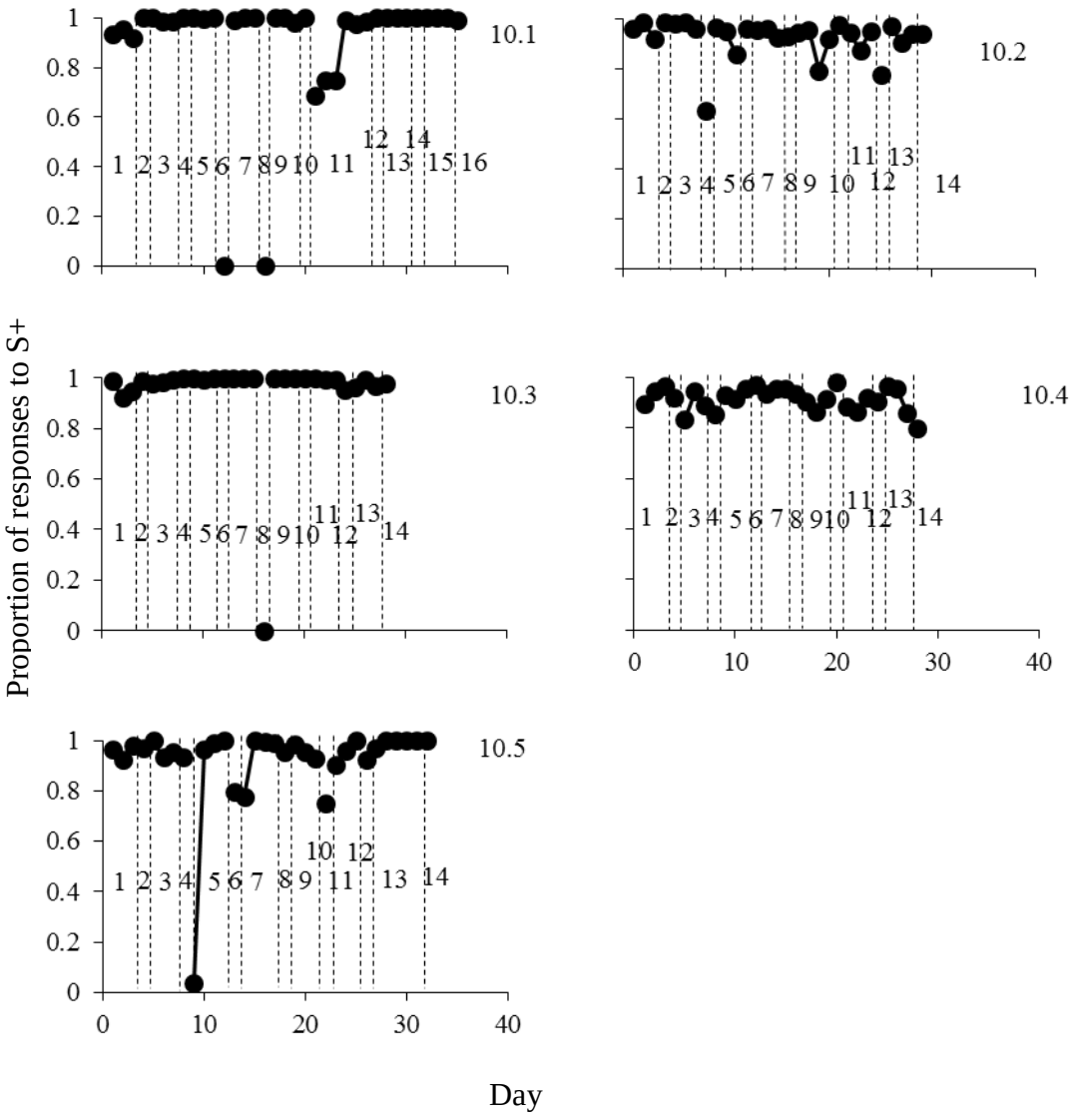

Figure 14. Proportion of responding on S+ (out of the total for S+ and S- response rates), for each bird, over probe and maintenance days. Dotted lines indicate a change in session type. Refer to Table 10 and Table 11 to see variable-interval schedules for each number.

Table 10

Session type for each period for 10.2 to 10.5 in Figure 14. Proportion of responding on S+ (out of the total for S+ and S- response rates), for each bird, over probe and maintenance days.

Period	Session Type
1	Training
2	Baseline 1
3	Maintenance
4	Phase 1
5	Maintenance
6	Phase 2
7	Maintenance
8	Phase 2
9	Maintenance
10	Phase 2
11	Maintenance
12	Phase 3
13	Maintenance
14	Baseline 2

Table 11

Session type for each period for Bird 10.1 in Figure 14. Proportion of responding on S+ (out of the total for S+ and S- response rates), for each bird, over probe and maintenance days.

Period	Session Type
1	Training
2	Baseline 1
3	Maintenance
4	Phase 1a
5	Maintenance
6	Phase 2
7	Maintenance
8	Phase 2
9	Maintenance
10	Phase 2
11	Maintenance
12	Phase 3
13	Maintenance
14	Phase 1b
15	Maintenance
16	Baseline 2

phase, it was found that there were also longer latencies. Conversely, if there were a greater number of responses made on a stimulus in a phase, there were shorter latencies. Mean Baseline 1 latencies were the shortest. Birds 10.1 and 10.3 also had the shortest latencies across stimuli in Baseline 1, though the remaining three birds were quicker to respond to probe stimuli during Phase 1 (noise before test) than any other condition.

During Phase 1, mean latencies across birds were longer for S+, Probe 1 and Probe 2 stimuli but latencies for Probe 3 and S- followed a similar data path to Baseline 1. During Phase 1, latencies for 10.1 were longer than Baseline 1 across all stimuli and there was less variation in length of latency between stimuli. This produced a flatter distribution than Baseline 1. Data paths for 10.1 in Phase 1b approximated data paths for Baseline 1. Bird 10.3 had longer latencies on all stimuli during Phase 1 but particularly on Probe 2, Probe 3 and S-.

The longest mean latencies, across birds and for all stimuli, were observed in Phase 2 (noise during test). Latencies were also longest in Phase 2 for each bird in the individual data. During Phase 3 (noise interrupting test), mean data paths followed Baseline 1. Similar findings were observed for 10.3, 10.4, and 10.5. Bird 10.1 had longer latencies on S+ and Probe 2 during Phase 3, and 10.2 had shorter latencies on all stimuli during Phase 3 than in Baseline 1.

A two-way repeated measures ANOVA was conducted for latency to first response across conditions, using the first exposure to Phase 1 for 10.1. Mauchly's test indicated that the assumption of sphericity had been violated for condition X^2 (5) = 14.73, p = .017. Degrees of freedom for this effect were corrected using Greenhouse-Geisser estimates (ε = .40). Results showed a significant within-subjects effect of stimulus location (see Table 12). Though there was no significant within-subjects effect of condition and no significant interaction for stimulus location by condition (see Table 12). Results were similar when the second exposure to Phase 1 for 10.1 was used (see Table 12) and when only probe data was used in the analysis (see Table 13). This contrasts with the finding that number of responses did differ significantly across conditions.

Latency data were also analysed for the three sessions for Phase 2, the two baselines, and before and after noise exposure in Phase 3 (see Figure 21, Figure 22, and Figure 23 in appendices). Findings were similar to those described for number of responses and for response rate. Results for two-way repeated measures ANOVAs on these latency data are shown in Table 12 and Table 13.

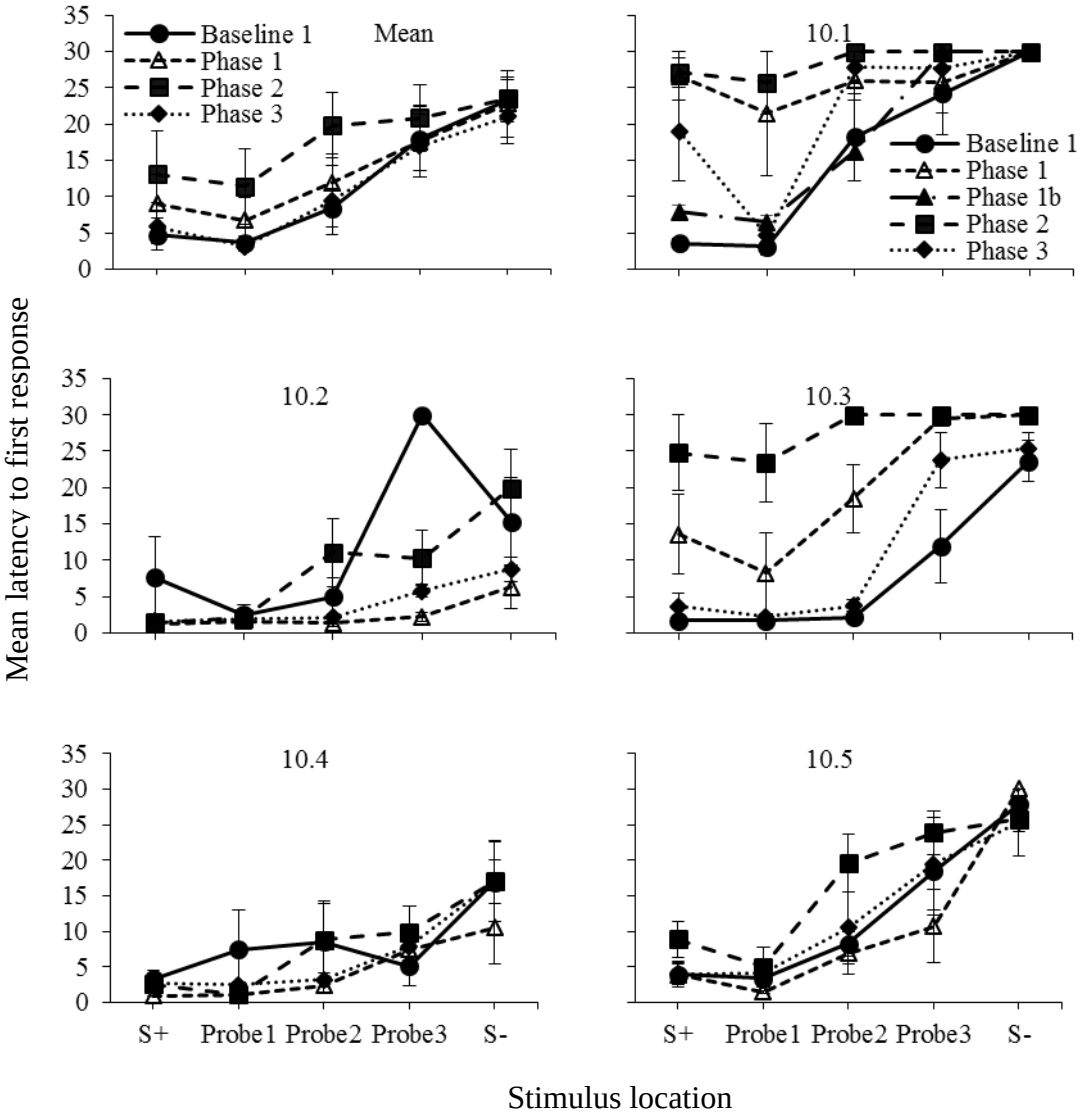

Figure 15. Mean latency to first response in a component for each of the stimulus locations, for each bird, during each of the experimental conditions (Phase 1 = noise before test, Phase 2 = noise during test, Phase 3 = noise interrupting test). Mean across birds is shown in top left. Bird 10.1 was exposed to Phase 1 twice, and thus has an extra condition (Phase 1b).

Table 12

Two-way repeated-measures ANOVA results for mean latency to first response in a component for each of the stimulus locations, across birds in: Figure 15. during each of the experimental conditions; Figure 21. during each of the three session days for Phase 2; Figure 22. during baselines taken before and after manipulations; and Figure 23. before and after the noise exposure in Phase 3.

	df	F	*p*	partial η2
Latency to First Peck				
Figure 15.				
First Exposure Phase 1				
Stimulus Location	4, 16	29.96	<.001	.88
Condition	1.21, 4.85	2.00	.224	.33
Interaction	12, 48	1.06	.414	.21
Second Exposure Phase 1				
Stimulus Location	2.20, 8.79	33.65	<.001	.89
Condition	3, 12	2.43	.116	.38
Interaction	12, 48	0.84	.607	.17
Figure 21.				
Stimulus Location	4, 16	8.88	.001	.69
Session	2, 8	0.52	.612	.12
Interaction	8, 32	1.04	.430	.21
Figure 22.				
Stimulus Location	4, 16	12.78	<.001	.76
Baseline Time	1, 4	0.02	.889	<.01
Interaction	4, 16	0.83	.526	.17
Figure 23.				
Stimulus Location	4, 16	12.08	<.001	.75
Test Time	1, 4	0.55	.498	.12
Interaction	4, 16	1.99	.145	.33

Table 13

Two-way repeated-measures ANOVA results for mean latency to first response in a component (only probes) for each of the stimulus locations, across birds in: Figure 15. during each of the experimental conditions; Figure 21. during each of the three session days for Phase 2; Figure 22. during baselines taken before and after manipulations; and Figure 23. before and after the noise exposure in Phase 3.

	df	F	*p*	partial η2
Latency to First Peck				
Figure 15.				
First Exposure Phase 1				
Stimulus Location	2, 8	20.54	.001	.84
Condition	3, 12	2.31	.128	.37
Interaction	6, 24	0.89	.518	.18
Second Exposure Phase 1				
Stimulus Location	2, 8	20.39	.001	.84
Condition	3, 12	2.73	.090	.41
Interaction	6, 24	0.67	.677	.14
Figure 21.				
Stimulus Location	2, 8	12.68	.003	.76
Session	2, 8	0.50	.627	.11
Interaction	4, 16	2.52	.082	.386
Figure 22.				
Stimulus Location	2, 8	11.11	.005	.74
Baseline Time	1, 4	0.03	.867	<.01
Interaction	2, 8	0.93	.432	.19
Figure 23.				
Stimulus Location	2, 8	6.24	.023	.61
Test Time	1, 4	2.39	.197	.37
Interaction	2, 8	0.92	.437	.19

3.3. Tonic Immobility Test

Tonic immobility data were analysed by latency to righting time and number of inversions required. Figure 16 illustrates the latency to righting time for each of the subjects. Bird 10.3 demonstrated the longest latency to righting time of all birds, while 10.2 showed the shortest latency. Figure 17 shows the number of inversions required for each bird to reach a tonically immobile state. Birds 10.1 and 10.2 required three and four inversions, respectively, while the remainder of the birds became immobile within one inversion. Birds 10.1 and 10.2 also showed the shortest latencies to righting time of all the birds. There was no difference in tonic immobility results attributable to a difference in breed for 10.5.

3.4. Summary

Shifts in number of responses emitted to the various stimuli were observed during Phase 1 (noise before test) and Phase 2 (noise during test) manipulations, but not during Phase 3 (noise interrupting test) (see Figure 8, Table 4, and Table 6). During Phase 1, 10.1 and 10.3 emitted few responses while the remaining three birds emitted more responses on probe stimuli compared to Baseline 1 (see Figure 8, Table 4, and Table 6). During Phase 2, all birds emitted fewer responses to each stimulus, but particularly probe stimuli close to and including the S- (see Figure 8). Results from a two-way repeated-measures ANOVA showed a statistically significant effect for stimulus location and for condition (see Table 4). Effect sizes were moderate to strong (see Table 4.). The interaction effect was not statistically significant and the effect size for the interaction was weak (see Table 4.). Results were similar for proportion of responses (see Figure 12 and Table 9).

Paired-samples t-tests conducted for each bird, comparing Baseline 1 to each phase, showed no significant difference in mean number of responses across stimuli for any comparison. Birds 10.3 and 10.4 were the exception, with significant differences observed for Phase 1 and Phase 2, respectively (see Table 6). Moderate to strong effect sizes were found for comparisons between Baseline 1 and Phase 1, and Baseline 1 and Phase 2 for 10.1, 10.2, 10.3, and 10.5 (see Table 6). Bird 10.4 had a strong effect size for Baseline versus Phase 2 but a weak effect size for Baseline 1 versus Phase 1 (see Table 6).

Latencies to the first response in a component were longest for probes in Phase 2, across birds (see Figure 15). For 10.1 and 10.3, latencies were longer on probes during Phase 1 than Baseline, while latencies for all other birds were shorter during Phase 1 than Baseline (see Figure 15). Phase 3 did not produce any

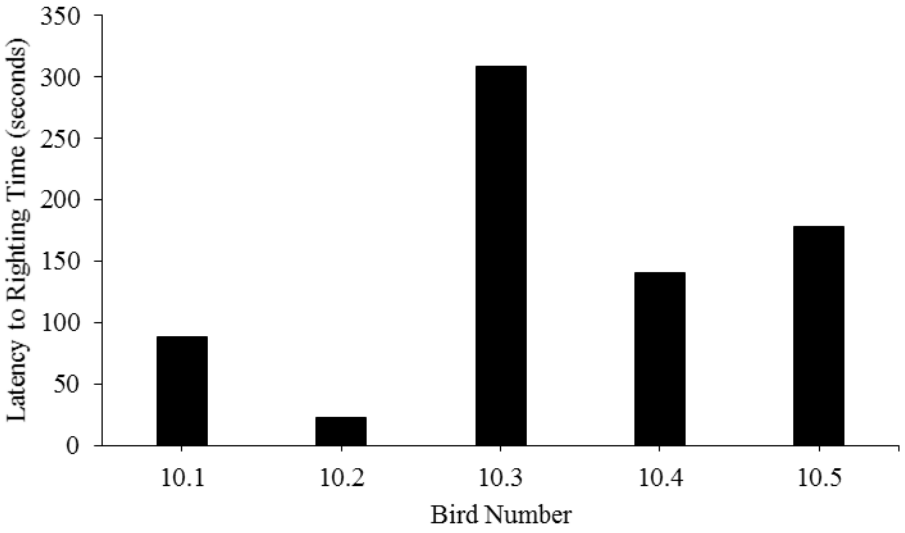

Figure 16. Latency to righting time in a tonic immobility test for each bird.

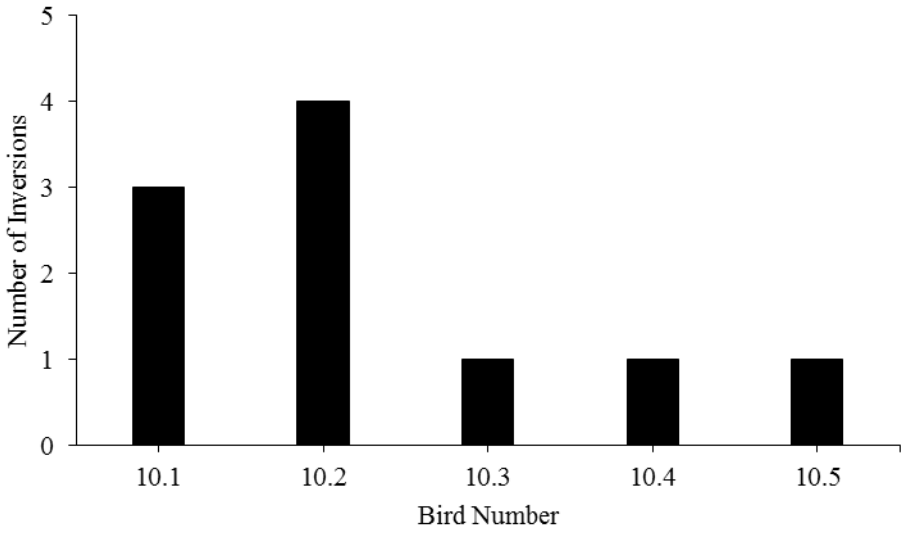

Figure 17. Number of inversions attempted in a tonic immobility test for each bird.

statistically significant changes in responding (see Figure 11 and Table 7); and no statistically significant change in responding was observed over session days for Phase 2 (see Figure 9 and Table 4), or over baseline measurements (refer to Figure 10 and Table 4).

4. Discussion

Chickens were successfully taught a spatial discrimination between two identical stimuli associated with rich (S+) and lean (S-) VI schedules, located on opposing sides of a screen. As expected, a generalisation gradient was established for all chickens during baseline measurements for assessing judgement bias. This indicated stimulus control by the location dimension. Following this discrimination training, subjects were exposed to three different manipulations with 100dB white noise played: prior to a judgement bias test (Phase 1); during a judgement bias test (Phase 2); and interrupting a judgement bias test (Phase 3).

It was hypothesised that the exposure to 100dB of white noise would be aversive to chickens and have an effect on current and future responding. When changes to responding on probe stimuli are observed, then a bias is present. An increase in responding on probe stimuli would constitute a positive bias, while a decrease in responding on probe stimuli would constitute a negative bias. Consistent with predictions, exposure to the noise did have a statistically significant effect on responding across conditions and birds (see Figure 8 and Table 4). Birds were expected to exhibit a positive bias following the noise in Phase 1. Three out of the five birds did show positive biases; though contrasting results were found for the remaining two birds, who showed negative biases in responding.

For Phase 2, it was hypothesised that a negative bias would be observed. As expected, when chickens were exposed to aversive conditions (white noise at 100dB) during testing, they showed a negative bias compared to when no noise was present (as with Baseline 1). Similar to Phase 1, birds were expected to maintain Baseline 1 levels of responding prior to the noise exposure in Phase 3, and show positive biases following the noise. After noise interrupted judgement bias testing (in Phase 3), birds did not show any biases in responding.

4.1. Multiple-Schedules in a Judgement Bias Task

Traditionally, judgement bias tests have been go/no go tasks, carried out using discrete trials with only one stimulus available at a time (e.g. Douglas et al., 2012; Doyle, et al. 2010; Burman et al., 2009). A go/no go procedure involves teaching animals to approach one stimulus (S+) for reward and not approach another (S-) to avoid punishment (or no reward), then assessing the proportion of the group that chose to approach each probe stimulus. Often, the measure of judgement bias in these designs is a) likelihood that animals will approach a

stimulus and/or b) latency to approach each stimulus. As with the current study, the stimulus dimension is often location.

As far as the author is aware, the current study was the first to employ a go/go multiple schedules procedure. Here, judgement bias was primarily analysed by number of responses emitted on each stimulus, rather than the proportion of animals that approached each stimulus. Testing was also conducted within-subjects in the current study, while the majority of the judgement bias literature examines between-subjects effects. Despite these differences to design, the positive and negative biases observed with group results in previous literature were able to be replicated (Bateson et al. 2011; Sanger et al., 2011).

Latency to approach stimuli has been a common measure of judgement bias in previous studies. Latency to first response in a component was also measured across components, for each stimulus in the current study. Just as latencies were shorter for some birds in the current study who moved from noise to no noise (as in Phase 1), latency to approach was also shorter for rats moved from high light to low light in a previous judgement bias investigation (Burman et al., 2009).

These similarities in findings to traditional approaches illustrate that a multiple schedules go/go procedure is appropriate for use in judgement bias testing. Furthermore, it may be preferable to use this procedure when testing bias within-subjects, as a traditional go/no go approach procedure only gives a degree of bias when testing across a group of animals (by proportion of animals that approached each stimulus).

4.2. Behavioural Contrast and S- as an Aversive Stimulus

Positive behavioural contrast, or an increase in the rate of responding on S+ as the schedule on S- was made leaner, was observed during discrimination training for 10.3 and 10.5 but not for other birds (see Figure 6). It has been argued that behavioural contrast is produced, at least in part, through error learning during the initial discrimination (Terrace, 1966). Error learning involves exposing the animal to both the reinforcing consequences of responding to S+ and to the non-reinforcing consequences of responding to S-. This results in the S- acting as an aversive stimulus. Similarly, it is important that in judgement bias testing, S+ is associated with reward while S- is aversive. Without these characteristics, the judgement bias procedure loses some external validity.

As behavioural contrast was observed for 10.3 and 10.5, this suggests that for at least these birds, the lean schedule of reinforcement on the S- was aversive.

McSweeny (1983) successfully produced a positive behavioural contrast using a multiple schedules VI 15-s EXT procedure with pigeons, where the operant response was a lever press. Though some previous studies using a similar design have failed to produce a positive behavioural contrast (McSweeny, 1978; Hemmes, 1973; Westbrook, 1973). McSweeny (1983) argued that for these studies, the differences between the rates of reinforcement on each schedule may not have been large enough to produce an effect, as they had used VI 60-s EXT and VI-120s EXT procedures. The current study used VI 60-s VI 600-s, so perhaps a positive behavioural contrast might have been observed for all chickens if S+ had been associated with a richer schedule of reinforcement.

Regardless of whether behavioural contrast was observed, all birds had low number of responses on S- during all conditions in the current study, suggesting that the discrimination was well established. As judgement biases were observed, this suggests that S- was aversive for all chickens but that this may not be the only requirement to produce a behavioural contrast in responding.

4.3. Judgement Bias Test

4.3.1. Phase 1 (noise before test). Birds 10.2, 10.4, and 10.5 increased their responding on probe stimuli following the noise in Phase 1 (see Figure 8), indicating a positive bias or 'optimistic' responding. The remaining two birds (10.1 and 10.3) emitted very few responses across any of the stimuli during Phase 1 (see Figure 8). Arguably, this demonstrates a negative bias or 'pessimistic' responding, in the sense that birds either a) were less likely to anticipate food delivery under the aversive conditions, and/or b) the significance of the aversive event was such that the emotional response produced by the noise interfered with the motivating operation of food deprivation.

Previous literature indicated that exposure to aversive conditions, prior to testing, produces negative judgement biases (Bateson et al., 2011; Harding et al., 2004). Results for 10.1 and 10.3 are consistent with this, though conflicting results were found for 10.2, 10.4, and 10.5. Other studies have found that a history of exposure to aversive conditions makes their absence more 'pleasurable' than if no aversive condition was experienced, resulting in 'optimistic' responding (Douglas et al., 2012; Burman et al., 2009). While this contradicts the previously described findings, it helps to explain results for 10.2, 10.4, and 10.5.

It is unlikely that repeated testing produced the negative biases observed in Phase 1, as this was the second judgement bias test to be conducted and the first

with noise exposure. Furthermore, no statistically significant difference was observed in responding between Baseline 1 and Baseline 2; indicating that repeated testing did not have a significant effect on the dependent measures in this study. Negative biases were not produced by satiation effects, as testing was always carried out in the morning before feeding and birds were never given more than their daily requirement of food in the chamber. As Phase 1 was the first test using noise exposure, there was no opportunity to habituate to the noise before this condition, meaning that habituation effects do not account for the positive biases seen with 10.2, 10.4, and 10.5.

An alternative explanation for these findings, is that differences between birds were a reflection of variations in responding that tend to occur from time to time, rather than as a result of the noise exposure. One test for this might involve comparing variability in responding between Baseline 1 and Baseline 2, with variation in responding between Baseline 1 and Phase 1. If the variation between Baseline 1 and Phase 1 was significantly greater than between Baseline 1 and Baseline 2, then it could be concluded that differences were likely caused by the noise exposure. Differences between birds may also be reflective of idiosyncratic traits.

Idiosyncratic differences. As mentioned before, the differences in findings may be explained in terms of idiosyncratic differences. It could be that the emotional response produced by the noise interfered with the motivating operation of food deprivation for 10.1 and 10.3, but not for other birds. In other words, idiosyncratic differences in 'emotionality' may have influenced the direction of judgement biases. Some evidence of individual differences was observed in a study that used multiple measures of affect (Wichman et al., 2012). While no significant differences in judgement bias were observed between treatment groups in the study, findings of heightened 'fearfulness' for some birds in a novel object test were correlated with a lack of trainability for testing (as they did not reliably feed from the bowl) (Wichman et al., 2012). Few studies into judgement bias have employed a within-subjects' design, where all subjects experienced all conditions in the same order. Thus, individual differences are not often reported. The current study included a test of tonic immobility in order to understand individual differences that might have influenced the effect of the noise exposures.

Tonic immobility. Tonic immobility tests are proposed to be a measure of animal 'emotionality' or 'fearfulness', with longer latencies to righting time

indicating a more ‘emotional’ bird. When tonic immobility test were conducted with experimental subjects, the latency to righting time was considerably higher for one bird (10.3) than the rest (see Figure 16). This bird also demonstrated negative biases for number of responses following the noise in Phase 1 (see Figure 8). This supports the idea that individual dispositions for ‘emotionality’ may influence judgement bias results. However, if this were true we would have expected the other bird (10.1) who showed a negative bias in Phase 1 to also demonstrate a long latency to righting time. In contrast, 10.1 demonstrated the second shortest latency of all the birds (see Figure 16). It is possible that 10.1 was not more ‘emotional’ than other birds and showed a difference in responding in Phase 1 due to some other intervening variable. However, there are a number of other possible explanations for this result.

A maximum of five inversions are conducted before a bird is considered ‘not susceptible’ to tonic immobility (Jones et al., 1981; Gallup et al., 1976). It is unclear whether this should be interpreted as meaning that a bird is less ‘emotional’ than birds who have reached an immobile state within five inversions, or whether tonic immobility is not a good test of ‘emotionality’ for that particular bird. As 10.1 required three inversions, this does indicate that this bird was at least less ‘susceptible’ to tonic immobility than the majority of the birds.

In studies examining breed differences in tonic immobility, it was found that while the latency to righting time differed significantly between White Leghorn and Production Red breeds, the number of inversions did not (Gallup et al., 1976). There was no apparent correlation between breed and tonic immobility in the current study, where Bantam cross and Buff Orpington chickens were used as subjects. The author argued that number of inversions may measure a separate component of tonic immobility to ‘emotionality’. In contrast, number of inversions in the current study did correlate with shorter latencies to righting time (see Figure 16 and Figure 17). Though sample sizes in the current study were small (N = 5), while Gallup et al. (1976) performed the test with a large sample of birds (N = 98). Thus, correlations between latency to righting and number of inversions in the current study may have been due to variation rather than a concrete effect. Birds in the current study were also handled on a daily basis, while the birds used in the study by Gallup et al. (1976) were likely not. Thus, the handling itself in the Gallup et al. (1976) study may have served as an aversive condition which may have produced different results than under ‘neutral’ conditions.

One study on behavioural measures of 'fearfulness', a construct supposedly also measured by tonic immobility, found inconsistent results both between-measures and with repeated testing. The authors argued that the construct itself, fearfulness, may not be stable (Miller, Garner, & Mench, 2006). Thus, there may have been some idiosyncratic 'emotionality' influence over effects of the noise; however, testing in the current study was conducted on a separate occasion from noise exposure, and individual differences in 'emotionality' may not be as easily detected when an animal is not in distress.

4.3.2. Phase 2 (noise during test). As predicted, fewer responses were emitted across stimuli in Phase 2, than during Baseline 1 (see Figure 8). Though t-tests were not significant for most birds, all effect sizes were above .8, indicating strong effects. Previous studies have also found reduced responding to probes when judgement bias was measured during exposure to differing aversive conditions. Both pigs currently placed in barren housing, and calves who were recently dehorned and separated from the dam, showed longer latencies to approach and a lower proportion of approach responses to probe stimuli than when in enriched or baseline conditions (Daros et al., 2014; Burman et al., 2009). In both studies, responses to positive (S+) and negative (S-) cues remained similar to baseline levels.

As fewer responses were emitted across all stimuli (not just probes) in the current study, it could be argued that differences in responding were a result of variation in number of responses from day to day, rather than a result of the noise exposure. Though if this were the case, we would not expect all of the hens to have consistently made fewer response under this condition. An alternative explanation for the lower numbers of responses during this condition might be that chickens would have made fewer responses, regardless of whether the noise was aversive or not. This could be tested by playing other sounds during judgement bias testing, which are not expected to be aversive, and comparing responding under these conditions with tests that were conducted whilst white noise was played. Decibel level may also play a part in whether a sound is aversive or not. To test this, the sounds could be played at differing decibel levels and judgement bias results compared.

4.3.3. Phase 3 (noise interrupting test). Phase 3 did not produce significantly different responding than Baseline 1, following the noise (see Figure 8, Figure 11 and Table 1). It was expected that a similar result would be found for

the second half of Phase 3, as was found for Phase 1. In contrast, noise did have an effect on responding during Phase 1 (see Figure 8 and Table 6). As there were a total of five separate sessions involving exposure to noise, with Phase 1 being the first and Phase 3 being the last, it is possible that chickens may have habituated to the noise over time.

During the first exposure to noise in Phase 1 for 10.1, the experiment was stopped seven components short of completion. This meant that Phase 1 had to be repeated for 10.1, following exposure to other phases (Phase 2 and Phase 3). Though there were seven extra components, Phase 1b gives an estimate of the effect of habituation to the noise for this chicken. Here, responding followed a similar data path to Baseline 1, indicating that at least some habituation to the noise may have taken place over time (see Figure 8). Thus, it is possible that the noise similarly had no statistically significant effect in Phase 3 due to habituation. Similar effects were observed in previous literature on rats (Parker et al., 2014). Just as repeated exposure to aversive stimuli can result in habituation, so too can repeated-testing affect the validity of results.

4.4. Repeated Measures Effects.

Baseline measurements of judgement bias were not statistically significant, when recorded prior to manipulations and following manipulations. This contrasts with studies that found with repeated testing, animals made fewer responses to probes, when controlling for all other variables (Doyle et al., 2010). One difference was that in the current study, there was a minimum of three days maintenance training between each judgement bias test, so birds were tested a maximum of two times per week. Doyle et al. (2010) tested three times per week, and two of these test days were consecutive. Thus, sheep in the Doyle et al. (2010) study had more recent experience with probe stimuli.

Another difference was that Doyle et al. (2010) used only one presentation of each stimulus on a test day. In the current study, there were multiple presentations of each stimulus, with each of the S+ and S- presented at a 3:1 ratio to probe stimuli components. Perhaps chickens were not exposed to the probe stimuli often enough in the current study, relative to other stimuli, for chickens to associate probe stimuli with non-reinforcement. Thus, responding on probes during the experiment may have been based, in large part, on chickens' judgement of the probability of reinforcement rather than consequences associated with probe stimuli. This would indicate that presenting S+ and S- on a greater number of trials,

relative to probes, adds to the validity of judgement bias tests. Importantly, this finding shows that repeated measures effects (from exposure to non-reinforced probe stimuli) did not have a statistically significant influence over the dependent measures in the current study.

4.5. Limitations

Some limitations need to be considered when analysing results of the current study. Significant differences in number of responses emitted across conditions were observed when analysed across birds, however no significant differences were detected when conducting t-tests on the individual level. This may have been a result of the small number of stimuli used in the test (N=5), which reduces the power of the t-test. In future studies, more stimuli may be used in order to increase the power and thus reduce the likeliness of making a type II error.

Chickens in the current study were mostly of mixed breeds, meaning that any underlying strain differences may not have been observed. Perhaps if all birds had been the same breed, then less variation might have been observed in biases between individuals. Idiosyncratic differences in hearing capacity may have also meant that different chickens experienced the noise differently. In future studies using noise, it is recommended that a hearing test be completed with each of the subjects. One procedure, based on signal detection theory, involved teaching chickens a discrimination between 'tone on' and 'tone off' conditions (O'Donnell, 1981). This was followed by presentation of tones at differing frequencies. If chickens correctly pecked the key associated with 'tone on' when a tone was played, it was assumed that the chicken was able to detect the noise (O'Donnell, 1981).

It appeared that chickens habituated to the noise and this had an influence on responding, at least in Phase 3. If this experiment were to be repeated then a simpler design may reduce the number of exposures needed. Phase 1 and Phase 2 from the current experiment could be combined, wherein noise would be played during the test for the first half, then turned off during the second half. This would remove the need for three extra sessions, as were conducted for Phase 2.

4.6. Conclusions

Since animals are unable to report on their welfare, it continues to be the responsibility of those caring for animals to estimate whether adequate housing and enrichment is being provided. The current study aimed to determine the appropriateness of a multiple schedules judgement bias procedure for the

assessment of animal welfare, and provide some clarity where previous literature has been inconsistent. As expected, white noise was effective as an aversive stimulus at the 100dB level. Result met hypotheses for Phase 2 and for the majority of birds in Phase 1. Where results did not meet expectations, it is argued idiosyncratic differences and habituation effects were at work. These findings indicate that judgement bias testing could be used to provide a good measure of whether the current welfare state for an animal, resulting from exposure to specific conditions, is satisfactory. It was also highlighted that conditions that may satisfy welfare needs for one individual may not satisfy welfare needs for another.

5. Future Research

There are a number of avenues that future research could take. To assess emotional reactivity, tonic immobility tests (or other behavioural measures) could be conducted immediately prior and following exposure to noise, with the relative change in tonic immobility response being the dependent measure. Results could be compared with judgement bias findings both within-subjects and between-subjects using different breeds of chickens, in order to further investigate whether a genetic 'emotionality' component might have an influence over judgement bias. Specific breeds of chicken are often selected for livestock in farming. Thus, any consistent findings of differences between breeds may help to better inform farmers about how to manage welfare for that particular breed.

6. References

Andrew, S. C., Perry, C. J., Barron, A. B., Berthon, K., Peralta, V., & Cheng, K. (2014). Peak shift in honey bee olfactory learning. *Animal Cognition, 17*(5), 1177-1186. DOI: 10.1007/s10071-014-0750-3

Bateson, M., Desire, S., Gartside, S. E., & Wright, G. A. (2011). Agitated honeybees exhibit pessimistic cognitive biases. *Current Biology, 21*(12), 1070–1073. DOI: 10.1016/j.cub.2011.05.017

Bedáňová, I., Chloupek, P., Vošmerová, P., Chloupek, J., & Večerek, V. (2010). Time course changes in selected biochemical stress indices in broilers exposed to short-term noise. *Acta veterinaria Brno, 79*(9), S40. DOI: 10.2754/avb201079S9S035

Briefer, E. F., & McElligott, A. G. (2013). Rescued goats at a sanctuary display positive mood after former neglect. *Applied Animal Behaviour Science, 146*(1-4), 45– 55. DOI: 10.1016/j.applanim.2013.03.007

Burgdorf, J., Knutson, B., & Panksepp, J. (2000). Anticipation of rewarding electrical brain stimulation evokes ultrasonic vocalization in rats. *Behavioral Neuroscience, 114*(2), 320-327. DOI: 10.1037/0735-7044.114.2.320

Burman, O. H. P., Parker, R. M. A., Paul, E. S., & Mendl, M. T. (2009). Anxiety-induced cognitive bias in non-human animals. *Physiology & Behavior, 98*(3), 345–350. DOI: 10.1016/j.physbeh.2009.06.012

Burman, O., McGowan, R., Mendl, M., Norling, Y., Paul, E., Rehn, T., & Keeling, L. (2011). Using judgement bias to measure positive affective state in dogs. *Applied Animal Behaviour Science, 132*(3), 160-168. DOI: 10.1016/j.applanim.2011.04.001

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159. DOI: 10.1037/0033-2909.112.1.155

Daros, R. R., Costa, J. H. C., von Keyserlingk, M. A. G., Ho¨ tzel, M. J., & Weary, D. M. (2014). Separation from the dam causes negative judgement bias in dairy calves. *PLoS ONE, 9*(5), 1-5. DOI: 10.6084/m9.figshare.1014331

Désiré, L., Boissy, A., & Vessier, I. (2002). Emotions in farm animals: A new approach to animal welfare in applied ethology. *Behavioural Processes, 60*(2), 165-180. DOI: 10.1016/S0376-6357(02)00081-5

Destrez, A., Deiss, V., Belzung, C., Lee, C., & Boissy, A. (2012). Does reduction

of fearfulness tend to reduce pessimistic-like judgment in lambs? *Applied Animal Behaviour Science, 139*(3-4), 233-241. DOI: 10.1016/j.applanim.2012.04.006

Douglas, C., Bateson, M., Walsh, C., Bédué, A., & Edwards, S. A. (2012). Environmental enrichment induces optimistic cognitive biases in pigs. *Applied Animal Behaviour Science, 139* (1-2), 65-73. DOI: 10.1016/j.applanim.2012.02.018

Doyle, R. E., Fisher, A. D., Hinch, G. N., Boissy, A., & Lee, C. (2010). Release from restraint generates a positive judgement bias in sheep. *Applied Animal Behaviour Science, 122*(1), 28–34. DOI: 10.1016/j.applanim.2009.11.003

Doyle, R. E., Vidal, S., Hinch, G. N., Fisher, A. D., Boissy, A., & Lee, C. (2010). The effect of repeated testing on judgement biases in sheep. *Behavioural Processes, 83*(3), 340-352. DOI: 10.1016/j.beproc.2010.01.019

Freymond, S. B., Briefer, E. F., Zollinger, A., Gindrat-von Allmenc, Y., Wyss, C., & Bachmann, I. (2014). Behaviour of horses in a judgment bias test associated with positive or negative reinforcement. *Applied Animal Behaviour Science, 158*, 34-45. DOI: 10.1016/j.applanim.2014.06.006

Gallup, G. G., Ledbetter, D. H., & Maser, J. D. (1976). Strain differences among chickens in tonic immobility: Evidence for an emotionality component. *Journal of Comparative and Physiological Psychology, 90*(11), 1075-1081. DOI: 10.1037/h0078662

Guttman, N. (1959). Generalisation gradients around stimuli associated with different reinforcement schedules. *Journal of Experimental Psychology, 58*(5), 335-340. DOI: 10.1037/h0045679

Gygax, L. (2014). The A to Z of statistics for testing cognitive judgement bias. *Animal Behaviour, 95*, 59-69. DOI: 10.1016/j.anbehav.2014.06.013

Harding, E. J., Paul, E. S., & Mendl, M. (2004). Cognitive bias and affective state. *Nature, 427*(6972), 312. Retrived from ProQuest online database.

Hearst, E., Koresko, M. B., & Poppen, R. (1964). Stimulus generalization and the response-reinforcement contingency. *Journal of the Experimental Analysis of Behavior, 7*(5), 369-380. DOI: 10.1901/jeab.1964.7-369

Hymel, K. A., & Sufka, K. J. (2012). Pharmacological reversal of cognitive bias in the chick anxiety-depression model. *Neuropharmacology, 62*(1), 161-166. DOI: 10.1016/j.neuropharm.2011.06.009

Jenkins, H. M. (1961). The effect of discrimination training on extinction. *Journal of Experimental Psychology, 61*(2), 111-121. DOI: 10.1037/h0047606

Jones, A. R., Bizo, L. A., & Foster, T. M. (2012). Domestic hen chicks' conditioned place preferences for sound. *Behavioural Processes, 89*(1), 30-35. DOI: 10.1016/j.beproc.2011.10.007

Jones, R. B. (1986). Conspecific vocalisations, tonic immobility and fearfulness in the domestic fowl. *Behavioural Processes, 13*(3), 217-225. DOI: 10.1016/0376-6357(86)90085-9

Jones, R. B., & Faure, J. M. (1981). Sex and strain comparisons of tonic immobility ("righting time") in the domestic fowl and the effects of various methods of induction. *Behavioural Processes, 6*(1), 47-55. DOI: 10.1016/0376-6357(81)90015-2

Jordan, W. P., Todd, T. P., Bucci, D. J., & Leaton, R. N. (2015). Habituation, latent inhibition, and extinction. *Neurobiology of Learning and Memory, 92*(2), 215-224. DOI: 10.1016/j.nlm.2008.07.001

Kloke, V., Schreiber, R. S., Bodden, C., Mo¨ llers, J., Ruhmann, H., Kaiser, S., Lesch, K., Sachser, N., & Lewejohann, L. (2014). Hope for the best or prepare for the worst? Towards a spatial cognitive bias test for mice. *PLoS ONE, 9*(8), 1-12. DOI: 10.1371/journal.pone.0105431

Knutson, B., Burgdorf, J., & Panksepp, J. (1998). Anticipation of play elicits high-frequency ultrasonic vocalizations in young rats. *Journal of Comparative Psychology, 112*(1), 65-73. DOI: 10.1037/0735-7036.112.1.65

MacKenzie, J. G., Foster, T. M., & Temple, W. (1993). Sound avoidance by hens. *Behavioural Processes, 30*(2), 143-156. DOI: 10.1016/0376-6357(93)90004-B

Malaka, R. (1999). Models of classical conditioning. *Bulletin of Mathematical Biology, 61*(1), 33-83. DOI: 10.1006/bulm.1998.9998

Matheson, S. M., Asher, L., & Bateson, M. (2008). Larger, enriched cages are associated with'optimistic' response biases in captive European starlings (Sturnus vulgaris). *Applied Animal Behaviour Science, 109*(2-4), 374–383. DOI: 10.1016/j.applanim.2007.03.007

McAdie, T. M. (1998). *The effects of white noise on the operant behaviour of domestic hens.* (Unpublished doctoral thesis). University of Waikato, Hamilton, New Zealand.

McSweeny, F. K. (1983). Positive behavioural contrast when pigeons press treadles

during multiple schedules. *Journal of the Experimental Analysis of Behavior, 39*(1), 149-156. DOI: 10.1901/jeab.1983.39-149

Mellor, D. J. (2014). Positive animal welfare states and reference standards for welfare assessment. *New Zealand Veterinary Journal, 63*(1), 17-23. DOI: 10.1080/00480169.2014.926802

Michael, J. (1993). Establishing operations. *The Behavior Analyst, 16*(2), 191-206. Retrieved from NCBI online database.

Miller, K. A., Garner, J. P., & Mench, J. A. (2006). Is fearfulness a trait that can be measured with behavioural tests? A validation of four fear tests for Japanese quail. *Animal Behaviour, 71*(6), 1323-1334. DOI: 10.1016/j.anbehav.2005.08.018

O'Donnell, C. S. (1981). *Detection of auditory stimuli by hens* (Unpublished Master's thesis). University of Waikato, Hamilton, New Zealand.

Otovic, P., & Hutchinson, E. (2015). Limits to using HPA axis activity as an indication of animal welfare. *Altex, 32*(1), 41-50. DOI: 10.14573/altex.1406161

Panksepp, J. (2007). Neuroevolutionary sources of laughter and social joy: Modeling primal human laughter in laboratory rats. *Behavioural Brain Research, 182*(2), 231-244. DOI: 10.1016/j.bbr.2007.02.015

Papciak, J., Popik, P., Fuchs, E., & Rygula, R. (2013). Chronic psychosocial stress makes rats more 'pessimistic' in the ambiguous-cue interpretation paradigm. *Behavioural Brain Research, 256,* 305– 310. DOI: 10.1016/j.bbr.2013.08.036

Parker, R. M. A., Paul, E. S., Burman, O. H. P., Browne, W. J., & Mendl, M. (2014). Housing conditions affect rat responses to two types of ambiguity in a reward–reward discrimination cognitive bias task. *Behavioural Brain Research, 274*, 73–83. DOI: 10.1016/j.bbr.2014.07.048

Paul, E. S., Harding, E. J., & Mendl, M. (2005). Measuring emotional processes in animals: the utility of a cognitive approach. *Neuroscience and Biobehavioral Reviews, 29*(3), 469–491. DOI: 10.1016/j.neubiorev.2005.01.002

Pompilio, L. & Kacelnik, A. (2005). State-dependent learning and suboptimal choice: When starlings prefer long over short delays to food. *Animal Behaviour, 70*(3), 571–578. DOI: 10.1016/j.anbehav.2004.12.009

Reynolds, G. S. (1961). Behavioral contrast. *Journal of the Experimental Analysis of Behavior, 4*(1), 57–71. DOI: 10.1901/jeab.1961.4-57

Richter, S. H., Schick, A., Hoyer, C., Lankisch, K., Gass, P., & Vollmayr, B. (2012). A glass full of optimism: Enrichment effects on cognitive bias in a rat model of depression. *Cognitive, Affective, & Behavioral Neuroscience, 12*(3), 527–542. DOI: 10.3758/s13415-012-0101-2

Rygula, R., Pluta, H., & Popik, P. (2012). Laughing rats are optimistic. *PLoS ONE, 7*(12), 1-6. DOI: 10.1371/journal.pone.0051959

Salmeto, A. L., Hymel, K. A., Carpenter, E. C., Brilot, B. O., Bateson, M., & Sufka, K. J. (2011). Cognitive bias in the chick anxiety–depression model. *Brain Research, 1373*, 124-130. DOI: 10.1016/j.brainres.2010.12.007

Sanger, M. E., Doyle, R. E., Hinch, G. N., & Lee, C. (2011). Sheep exhibit a positive judgement bias and stress-induced hyperthermia following shearing. *Applied Animal Behaviour Science, 131*(3), 94–103. DOI: 10.1016/j.applanim.2011.02.001

Siegford, J. M. (2013). Multidisciplinary approaches and assessment techniques to better understand and enhance zoo nonhuman animal welfare. *Journal of Applied Animal Welfare Science, 16*(4), 300–318. DOI: 10.1080/10888705.2013.827914

Skinner, B. F. (1938). *The behaviour of organisms.* New York: Appleton-Century-Crofts.

Spruijt, B. M., van den Bos, R., & Pijlman, F. T. A. (2001). A concept of welfare based on reward evaluating mechanisms in the brain: anticipatory behaviour as an indicator for the state of reward systems. *Applied Animal Behaviour Science, 72*(2), 145-171. DOI: 10.1016/S0168-1591(00)00204-5

Stamp Dawkins, M. (2008). The science of animal suffering. *Ethology, 114*(10), 937-945. DOI: 10.1111/j.1439-0310.2008.01557.x

Terrace, H. S. (1966). Stimulus control. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application* (pp. 271-344). New York, NY: Meredith Publishing Company.

Terrace, H. S. (1964). Wavelength generalization after discrimination learning with and without errors. *Science, 144*(3614), 78-80. Retrieved from JSTOR online database.

Terrace, H. S. (1963). Errorless discrimination learning in the pigeon: Effects of chlorpromazine and imipramine. *Science, 140*(3564), 318-319. Retrieved from JSTOR online database.

Wichman, A., Keeling, L. J., & Forkman, B. (2012). Cognitive bias and anticipatory behaviour of laying hens housed in basic and enriched pens. *Applied Animal Behaviour Science, 140*(1-2), 62– 69. DOI: 10.1016/j.applanim.2012.05.006

Zimmerman, P.H., Buijs, S.A.F., Bolhuis, J.E., & Keeling, L.J. (2011). Behaviour of domestic fowl in anticipation of positive and negative stimuli. *Animal Behaviour, 81*(3), 569-577. DOI: 10.1016/j.anbehav.2010.11.028

7. Appendices

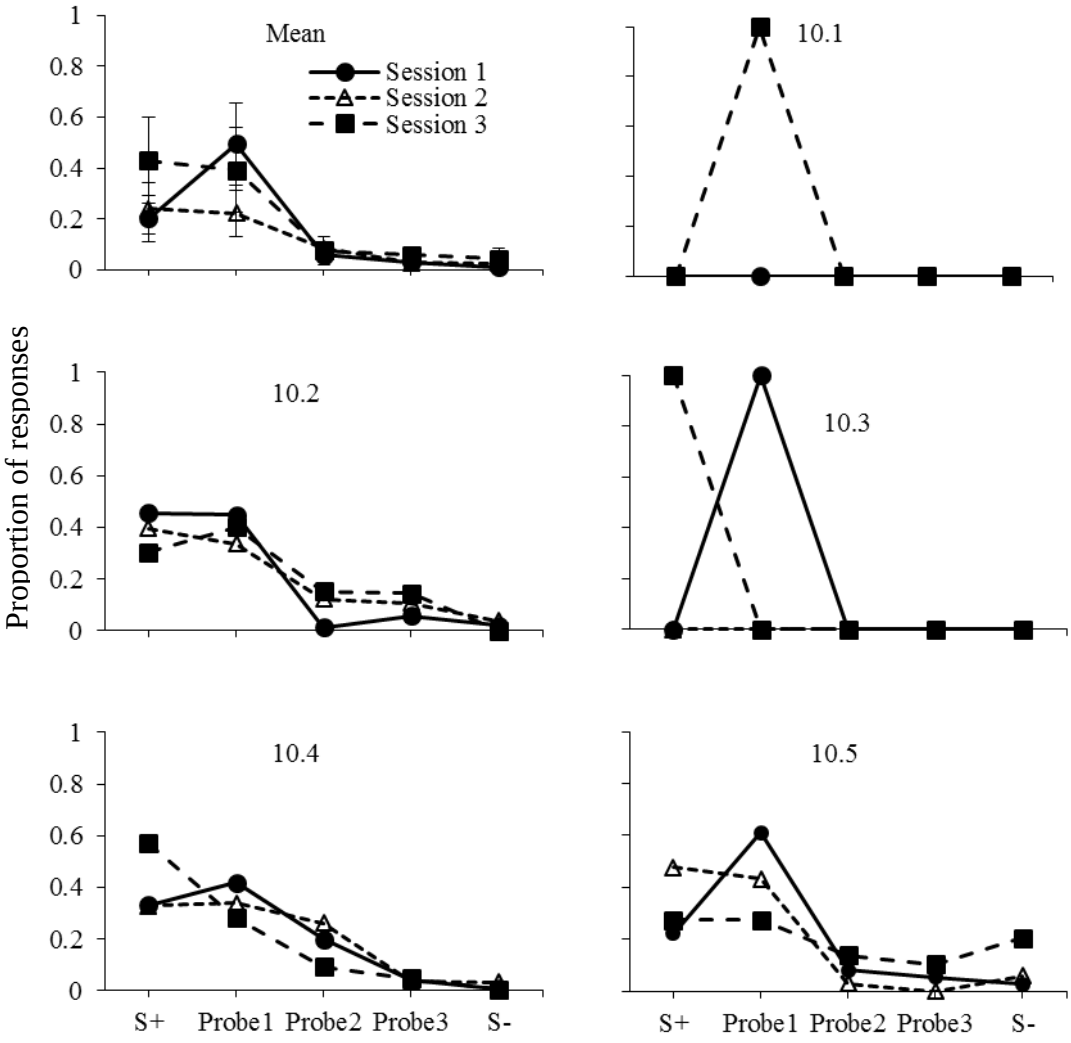

Figure 18. Proportion of responses on each of the stimulus locations, for each bird, during each of the three session days for Phase 2. Mean across birds is shown in top left.

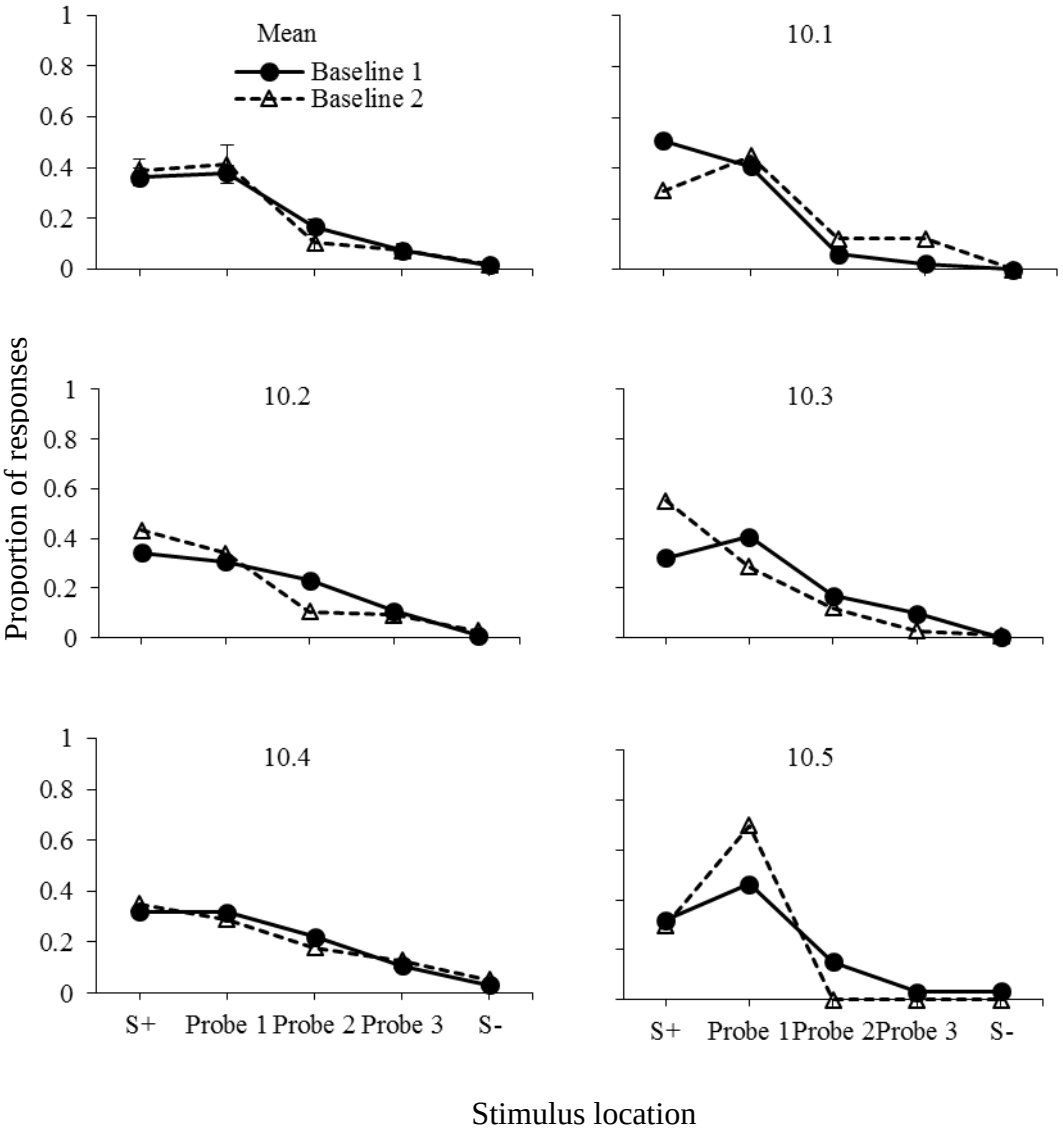

Figure 19. Proportion of responses on each of the stimulus locations, for each bird, during baselines taken before and after manipulations. Mean across birds is shown in top left.

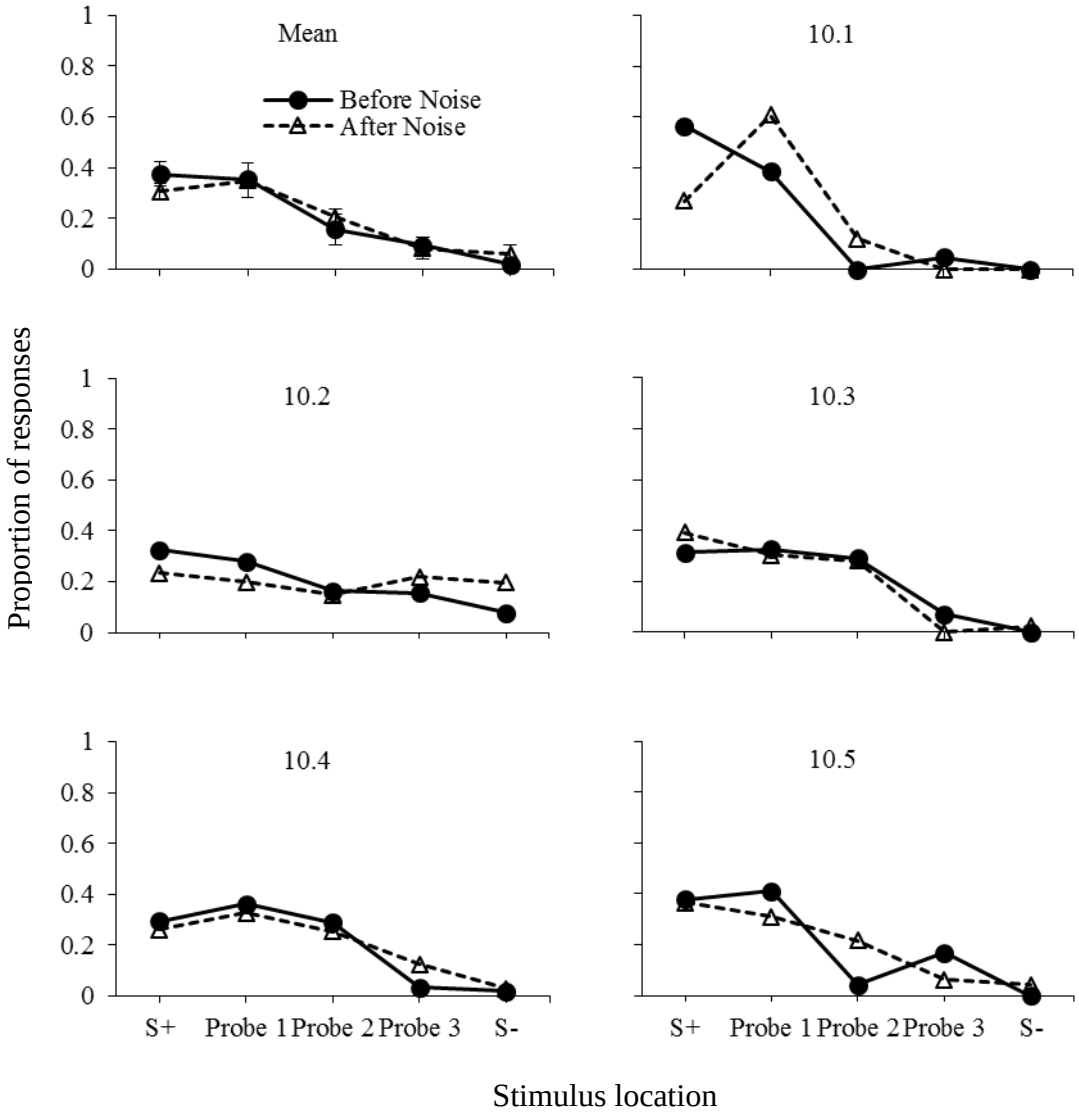

Figure 20. Proportion of responses on each of the stimulus locations, for each bird, before and after noise exposure in Phase 3. Mean across birds is shown in top left.

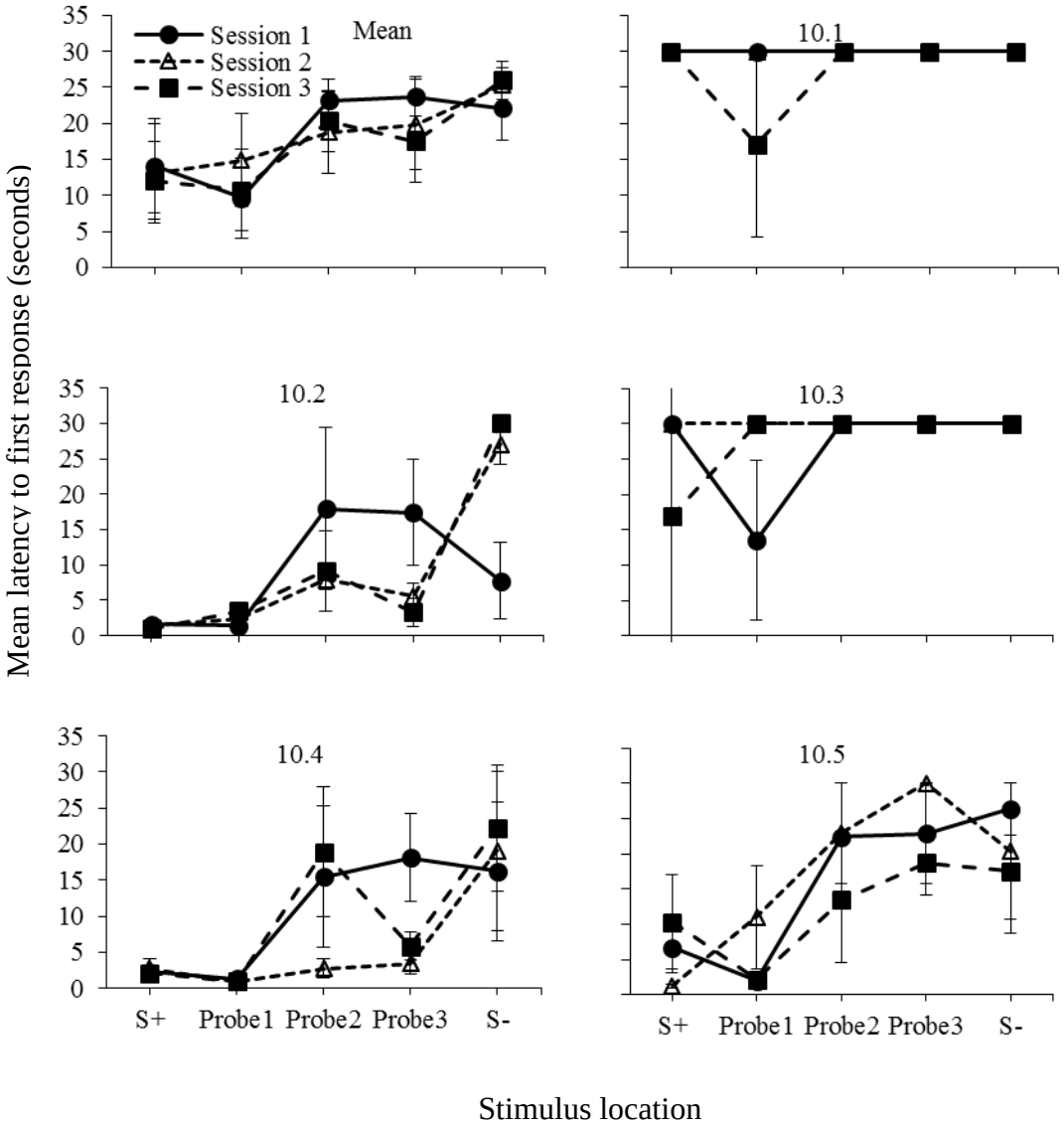

Figure 21. Mean latency to first response in a component for each of the stimulus locations, for each bird, during each of the three session days for Phase 2. Mean across birds is shown in top left.

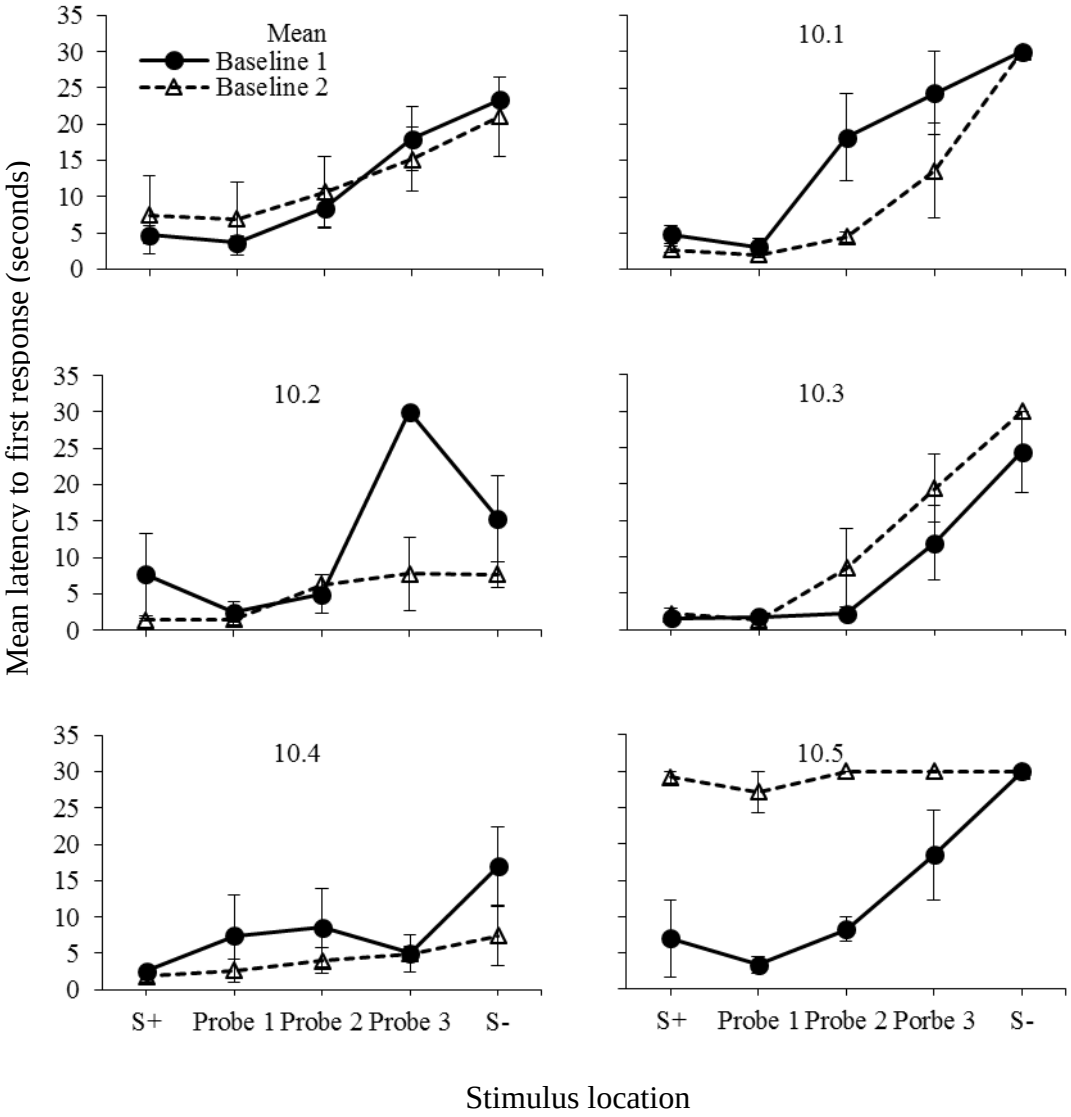

Figure 22. Mean latency to first response in a component for each of the stimulus locations, for each bird, during baselines taken before and after manipulations. Mean across birds is shown in top left.

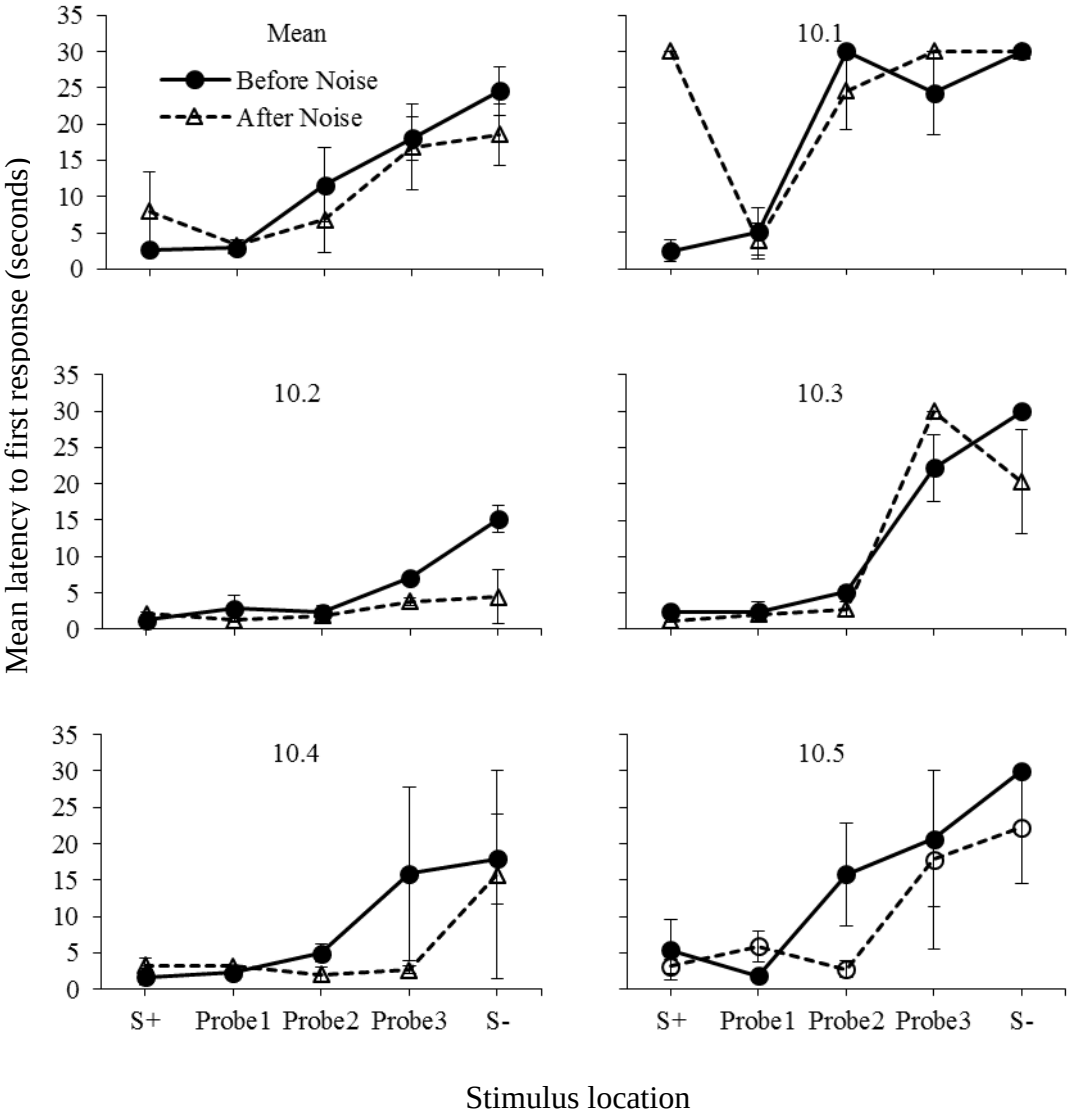

Figure 23. Mean latency to first response in a component for each of the stimulus locations, for each bird, before and after the noise exposure in Phase 3. Mean across birds is shown in top left.