Inferring affective state through biased actions in rats

Haris Organtzidis

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Life Sciences.

School of Physiology, Pharmacology and Neuroscience.

March 2022

Word count: 39793

Abstract

Affective state is an integral component of the way animals, including humans, perceive and interact with their environment. Animals can be biased by their affective state in the way they act to receive rewards or avoid punishments.

The Judgement Bias Task (JBT) is a decision-making task aimed at measuring biases in the interpretation of ambiguous information. In chapter 2, I initially ran replication studies (N=15 male rats) of past findings involving pharmacological manipulations of clinical importance, such as ketamine and amphetamine. After failed attempts to replicate published ketamine results, I designed a novel JBT variant to address shortcomings of the original task around the ambiguity of the test stimulus and the frequency of negative feedback and trial presentations. Pilot studies on this variant (N=16 male rats) revealed a different type of perceptual bias in the animals' responses. I discuss how this bias confounded the interpretation of results and how it relates to the original task design.

By collating data from past JBT studies, I conducted a large-scale analysis, which revealed that factors relating to past trials were important in determining animals' actions. Therefore, I designed statistical models that were able to account for these factors and any other biases in the animals' behaviour. Inference by model parameters, instead of summary statistics of actions, grants a more detailed view into the animal's decision-making process and reveals differences between the effects of ketamine and amphetamine.

Subsequently, I designed a novel foraging task, where animals were free to acquire reward or flee to avoid an imminent threat. Different versions of the task were tested in a pilot study (N=16 male rats). A statistical model reveals individual differences that become apparent when the threat was least predictable, as signaled by the constant presence of an odor in the operant chamber.

Finally, I present a theoretical model, based on reinforcement learning (RL) theory, which incorporates biases due to affective state. Simulated environments with naturalistic elements were also proposed. The model was compared to classical RL models within these environments to assess the benefits of affective biases. Overall, this thesis offers approaches to improve on how inference of affective state is performed, in addition to a hypothesis about why affective state is important for the survival of an animal.

Dedication and Acknowledgements

The first thing that I would like to acknowledge is that I was fortunate enough to be supported by more people than this page would allow me to name.

My supervisors, Claire Hales, Conor Houghton and Emma Robinson gave me the freedom to pursue my own ideas along the multiple forking paths of a PhD, while being there for every step of the way. Our conversations, whether they were about science or anything else, were always stimulating and I remember feeling more motivated after meeting with them.

I would also like to thank all members of Emma Robinson's and Richard App's groups in Bristol for being so inclusive and eager to help. I do hope that all of my future work environments will be as friendly and supportive as these two groups have achieved.

I am grateful to all my friends and my brother back home, as much as the new friends that I made in Bristol, for being there through good and bad times, greatly contributing to my best non-research related memories of the last four years. A special mention needs to go to the person who contributed the most to these memories, Christina, my wonderful partner. Doing a PhD during a global pandemic would not have been the same without her.

Most importantly, I would like to thank my parents. They gave me all the love and support I could ever ask for, and then they gave me more.

Finally, I would like to acknowledge the 66 rats that took part in my experiments. In the following pages, I have done my best to emphasise their contribution to science.

Author's Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..... DATE:.....

Contents

List of Figures

List of Tables

List of abbreviations

SSRI	Serotonin Selective Reuptake Inhibitor
NRI	Noradrenaline Reuptake Inhibitor
JBT	Judgement Bias Task
ACI	Ambiguous Cue Interpretation
ABT	Affective Bias Test
SNRI	Serotonin & Noradrenaline Reuptake Inhibitor
PRLT	Probabilistic Reversal Learning Task
GLM	Generalised Linear Model
RPE	Reward Prediction Error
MCMC	Markov Chain Monte Carlo
NUTS	No U-Turn Sampler
HMC	Hamiltonian Monte Carlo
DDM	Drift-Diffusion Model
RL	Reinforcement Learning
MEG	Magnetoencephalography
AC	Ambiguous Cue
LL	Lever Lights
HC	High-reward cue
LC	Low-reward cue
HA	High-reward action
LA	Low-reward action
HR	High reward
LR	Low reward
ITI	Inter-trial Interval
CRF	Continuous Reinforcement
CBI	Cognitive Bias Index
CST	Conditioned Suppression Task
CS	Conditioned Stimulus

PTSD	Post-traumatic Stress Disorder
FR	Fixed Ratio
US	Unconditioned Stimulus
ECDF	Empirical Cumulative Density Function
ESS	Effective Sample Size
BFMI	Bayesian Factor of Missing Information
FF	Free Food
LOOIC	Leave-One-Out Information Criterion
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
PSIS	Pareto-Smoothed Importance Sampling
DB	Deferred Bandits
TB	Tardy Bandits
MC	Monte Carlo

Chapter 1

General introduction

Emotions are an integral part of any human experience. Cognition and emotions are often thought of as two components of a reciprocal relationship (Damasio 2006, chapter 7), even though there are unresolved arguments around primacy; could emotions arise without a prior cognitive process (Zajonc 1984) or are they always the outcome of some cognitive appraisal (Lazarus 1982)? Affective state is a more general term for an internal state that includes some emotional component.

I have assumed in the present thesis that there exist emotional primitives that can be modelled across species (Anderson et al. 2014). For instance, the valence of an affective state, either positive or negative, and the level of arousal that characterises it are two common dimensions, along which human emotions, such as anger and happiness, could be mapped (Russell 2003). The current assumption then is that non-human animals can find themselves in the same states of valence and arousal, which would alter their behaviour accordingly, without necessarily labelling these states with the respective human emotions. These emotional primitives that underlie the affective state of both human and non-human animals could also be thought of as evolutionary conserved states that maximise the chances of survival (Bach et al. 2017). In this sense, affective state could drive pre-programmed behavioural motifs in situations when it is critical to act quickly and effectively without gathering more sensory evidence or considering every possible course of action.

Adding to its already broad definition, there are multiple somatic markers that are indicative of distinct affective states, for example changes in heart rate, respiratory patterns and skin conductance (Bechara et al. 2005). These changes in bodily states are thought to be modulated by the brain, in order to prepare the body to act in

accordance with its current affective state (Barrett 2017). Affective state can bias the process of deciding on an action at different points during the interaction of an individual with their environment; this could be as part of the background internal state influencing how incoming sensory stimuli are interpreted, prioritising which ones are attended to or which memories are recalled, or as specific emotions relating to the execution of an action, or as predicted emotions of the outcomes of an action (Dunning et al. 2017).

Biases due to affective state do not hold an explicitly negative connotation. The value of affective state as an adaptive modulator of actions has been corroborated in theoretical and empirical work (Hirsch et al. 2000; Haselton et al. 2006; Trimmer et al. 2013). However, biases exhibited through various forms of actions, when responding to affectively valenced stimuli, have been linked with vulnerability to mood disorders, such as anxiety and depression (Mathews et al. 2005). Such actions do not necessarily preclude the conscious awareness of being in an specific affective state (Winkielman et al. 2004; Winkielman et al. 2005). This statement has two important implications; it supports the argument that affective state could be studied in non-human animals, where subjective feelings are inaccessible (Mendl et al. 2020), and it highlights the value of using actions to infer affective state in humans, as opposed to self-reported questionnaires, which might not faithfully represent the individual's state. Thus, inferring affective state with high precision could aid in numerous steps towards the treatment of mood disorders; from diagnosis and prognosis to identification of novel treatment targets (see Slaney et al. 2018a for a discussion of the benefits of action biases as biomarkers for depression).

The present chapter introduces the various ways in which affective state can lead to biased actions in both humans and non-human animals. It follows a methodological categorisation, where each section presents a different way of inferring affective state given observations of actions. Firstly, there is inference by task design, where summary statistical measures of the observed actions were directly related to affective state by a verbal theory. Secondly, inference by statistical models will be addressed, which could account for multiple latent factors and sources of variability that might constitute an affective state given the observed actions, thus offering a more detailed look into the internal process that generates actions. The final section is concerned with inference by computational models, motivated by theories about the role of affective state in learning and decision-making processes.

This categorisation gradually presents more capable inference tools, which offer more

detailed ways of investigating how affective state could drive behaviour. The following chapters of this thesis will employ approaches based on these three categories in order to infer the affective state of rats in different tasks solely by observing their actions.

1.1 Inference by task design

This category constitutes the basis of any attempt to infer affective state. It involves the design and execution of behavioural tasks, according to a verbal theory, as opposed to one utilising mathematical tools (but see van Rooij et al. 2020 for an intergrated view of both), about the influence of affective state on actions. Affective state has been shown to bias multiple functions across the continuum between receiving sensory input and executing an action. The tasks that measure such biases have further motivated non-human animal experiments, in an effort to produce more realistic animal models of mood disorders (Slaney et al. 2018b) and to offer reliable means of measuring affective state for animal welfare research (Mendl et al. 2009).

1.1.1 Interpretation bias

The interpretation of ambiguous information has been shown to be biased by affective state. Healthy humans assigned probabilities to the occurrence of uncertain future events that were influenced by their underlying affective state (Wright et al. 1992). Either a positive or a negative mood was induced to each participant, by a suggestion to focus on happy or sad personal experiences under hypnosis. Consequently, participants made judgements about the likelihood of occurrence of either personal or nonpersonal future events. They assigned lower probabilities to positive events, irrespective of the context, after a negative mood manipulation, compared to control subjects, and conversely after a positive mood was induced. In another study, healthy participants were more likely to categorise a face of ambiguous emotional expression as happy rather than sad, after the administration of a selective serotonin reuptake inhibitor (SSRI) drug, a conventional antidepressant (Harmer et al. 2003).

Following the same experimental paradigm in a later study, depressed participants more readily interpreted ambiguous faces as expressing negative emotions, while this

bias was mitigated by an acute treatment with a noradrenaline reuptake inhibitor (NRI), another conventional antidepressant drug (Harmer et al. 2009). Importantly, in this study, the change in judgement bias preceded a change in self-reported mood by several days, thus indicating the prognostic value of measuring biases. Additionally, judgement biases in interpreting ambiguous faces were correlated with the risk of relapse in depressed individuals (Bouhuys et al. 1999). In a categorisation task, patients with moderate to severe depression were less accurate and slower in recognising a face as neutral, while frequently misclassifying it as sad (Leppänen et al. 2004). This recognition bias was apparent even after symptom remission, as evaluated by a clinical psychiatrist, however no longitudinal monitoring was performed to assess the probability of relapse.

Inspired by the observed interpretation biases in humans that were discussed above, a rat model was designed, using auditory stimuli, rewards and punishments, instead of emotional faces (Harding et al. 2004). Animals were trained to press a lever to gain reward when one reference tone was presented and withhold this action during presentations of a second reference tone in order to avoid a white noise punishment. Consequently, three tones of intermediate frequencies were presented during test sessions, corresponding to the ambiguous stimuli. Rats were subjected to an unpredictable housing condition, that involved various changes to their home cages over nine days. This manipulation induced a negative affective state, which led to the animals interpreting the ambiguous cues as ones predicting the punishing outcome and thus withholding their responses more often than pressing the lever that was associated with reward. This effect was more pronounced for ambiguous tones of frequencies closer to the frequency of the reference tone predicting the loud noise.

Multiple variants of this task followed in a large range of species, either as animal models of the affectively driven interpretation bias in mood disorders or for animal welfare research (see Roelofs et al. 2016 for a review of both types of studies). The central premise of the task was based on generalisation of the cue-action-outcome contingencies of reference cues, learned during training, to actions as a response to ambiguous cues, which were newly presented during testing. There were always two possible actions, each one reflecting an interpretation of an ambiguous cue, as one predicting a positive or a negative outcome. Actions with a learned positive outcome were inferred to be optimistic, whereas the alternative action was deemed pessimistic. The term Judgement Bias Task (JBT) will be used in the present thesis to refer to the family of tasks that implement variants of this design, although the

Ambiguous Cue Interpretation (ACI) task is another common name.

A version of the task, where both reference tones required a correct action, either to acquire a reward or avoid a punishment, revealed a negative bias in helpless rats, a rodent model of depression. These rats acted in order to avoid the shock more frequently during the ambiguous trials (Enkel et al. 2010). Utilising the same task design, a later study found that chronic social defeat also biased the interpretation of an ambiguous tone towards acting to avoid the punishment (Papciak et al. 2013). Moreoever, rats exhibited a negative shift in their bias after an acute pharmacological treatment with reboxetine, an NRI with antidepressant properties (Anderson et al. 2013). However, daily treatments with fluoxetine, an SSRI antidepressant drug, across one week, led to a small positive shift of the bias. More recently, a new version of the reward-punishment JBT was designed, involving more naturalistic actions and less aversive outcomes (Jones et al. 2018). According to this design, rats had to keep their heads inside a food trough in order to get rewarded or move away from it to avoid an air puff.

A version of JBT with only reward outcomes was also developed to minimise the exposure of animals to aversive experiences (Hales et al. 2016). In this task, the two alternative outcomes were a large and a small reward amount. Responding on the lever that was known to result in a large reward was considered the optimistic choice, while pressing the lever that led to the small reward constituted a pessimistic action. The acute induction of a negative affective state, via a treatment with an anxiogenic drug, resulted in a negative shift in the action bias. A second study using the same variant showed that positive manipulations of affective state by amphetamine, a psychostimulant drug that both releases and inhibits the reuptake of dopamine and noredrenaline, and ketamine, an NMDA receptor antagonist, were capable of a positive shift in bias (Hales et al. 2017).

A meta-analysis of pharmacological studies on JBT has revealed an agreement between the hypothesised change in affective state under the drug treatment and the shift in the animals' action bias as they interpret the ambiguous cues (Neville et al. 2020a). This work further supports the claim that affective state can be inferred directly from the animals' actions.

There are variants of the JBT that were not mentioned here, including other sensory modalities such as visual or tactile cues, compound stimuli of multiple modalities or spatial locations as the reference and ambiguous cues (see Nguyen et al. 2020 for a systematic review). The present thesis will involve the auditory variant of the task

that was presented above.

1.1.2 Memory bias

The modulation of memory recall processes by affective state has been observed in both healthy people and depression patients. Even though memory recall is not an explicit action, through which an individual interacts with their environment, it is an integral part of planning how to act (Buzsáki et al. 2014). In this regard affective state could set the context upon which congruent memories would be recalled, which in turn would lead to biases in action planning and execution. After an emotional face recognition test akin to the one presented above, depression patients managed to recall fewer happy faces than controls, yet they were more accurate in recognising previously seen sad faces (Ridout et al. 2003). Word lists are another common stimulus with multiple studies having measured a propensity to recall affectively valenced words that were congruent with an individual's affective state at the time of recall, in both patients and healthy controls (Blaney 1986 for a review). The recall of personal experiences from both the short-term and distant past was similarly influenced by affective state, where the valence of an affective manipulation prior to recall tended to be the same as the valence of the recalled memories (Bower 1981). Both the autobiographical and stimulus recall studies involved an explicit recall stage. However memory biases in depressed people have been observed implicitly. Patients that were exposed to words with an affective connotation via simplified scenarios of social interactions, depicted in drawings, were less likely than healthy controls to fill in a studied positive word in a later semantic task, where a definition was given for the missing word (Watkins 2002).

Results of memory bias studies on patients with general anxiety disorder have not been as consistent as those on depression patients (MacLeod et al. 2004 for a review). It was suggested that anxious people encode affective information more reliably, particularly in unpredictable environmental conditions (Pury et al. 2001). Such conditions could be emulated in the lab by the order that the test stimuli are being presented. In any case, it is evident that a direct association of the valence of recalled words or faces with affective state does not hold for anxiety. Thus, more nuanced explanations are required, potentially including a view of the disorder as a multidimensional construct with valence of the affective state being only a part of it. This view will be explored further in the section with inference by computational models.

Besides its effect on memory recall of congruent stimuli, affective state has also been shown to modulate the remembered association between learned action-outcomes contingencies. While learning the potential outcome of actions, animals' affective state has been shown to augment the learned outcome. The Affective Bias Test (ABT) was introduced as a behavioural assay in rats for testing whether they have differential preference between two actions, when both actions result in the same learned reward, but one of them had been previously paired with a manipulation on affective state (Stuart et al. 2013). During days 1 and 3 of training, rats were presented with two digging substrates A and C, with A hiding a single reward pellet and C containing no reward. On the same days rats were treated with a drug or a physical manipulation, such as social isolation, that was purported to change their affective state. During days 2 and 4 another substrate B that contains the same reward magnitude as A was presented along with the same substrate C. After training for four days on the associations between digging in a substrate and acquiring reward, rats were tested on day 5 by having to choose between A and B, the two previously rewarded substrates. Animals were shown to prefer substrate B, when A was paired with an affective manipulation of negative valence and conversely for positive-valenced ones. Manipulations of either valence were equally effective in causing a bias when they were administered before or after the learning sessions during days 1 and 3.

A recent ABT study compared the effect that social play had on the rats' bias, when they had been previously treated with an anxiogenic compound (Hinchcliffe et al. 2022). Social play between cage mate rats was associated with a positive affective state, as measured through its correlation with elevated 50 kHZ ultrasonic vocalisations. The positive affective state, induced by social play, managed to mitigate the negative effect of an anxiogenic drug injection prior to a training session on days 1 and 3, even when social play occurred several hours after the session.

The ABT has been validated as a preclinical model of affective state manipulations as it has exhibited predictive efficacy in capturing the effect of conventional antidepressants and making predictions for novel, rapid-acting ones such as ketamine (Stuart et al. 2015b). The case of ketamine is particularly compelling, as it was shown to mitigate a negative bias that was acquired previously, without any reexposuse to the task context prior to testing. Additionally, venlafaxine, a serotonin and noradrenaline reuptake inhibitor (SNRI), but not ketamine, when administered during training, caused a preference towards the rewarded substrate that was experienced on the same days.

The ABT utilises a type of retrospective inference across days. Observations of action preferences during the test day are being used to infer the animals' affective state during a past learning session, when they first encountered the substrate that was paired with the affective manipulation. Thus, the memory of the learned action-outcome associations is always implicated. One potential target for the affective state's influence is the learning process itself, in cases of the manipulation preceding a learning session. An example of such an effect was found in a modified version of the ABT that was specifically designed to measure learning deficits by raising the amount of reward for one of the substrates (Stuart et al. 2019). Rats that were subjected to maternal separation as a model for early life adversity did not show a preference towards the more richly rewarded substrate, as opposed to untreated healthy rats. However, another possible and complementary explanation is that animals conflate the experienced rewards with their affective state during the same training day, which in turn leads to action biases during the subsequent test day.

1.1.3 Attention bias

Affective state has been shown to affect the way people direct their attention. Attention biases can be manifested in a valence-congruent manner, akin to memory biases examined above, with evidence suggesting an interaction between biases in working memory and attention in depression (Raedt et al. 2010). For example, being in a negative affective state would prioritise attention towards negatively valenced stimuli and the retention of similar information in working memory. A similar effect of congruence was observed in a study with faces comprised of two characteristics, sex and expressed emotions (Gilboa-Schechtman et al. 2004). Participants belonging to a high depression-index group, as judged by Beck's Depression Inventory, were influenced by the emotional component of the faces, even when explicitly asked to ignore it and make judgements about the sex of the depicted person.

Acute induction of a negative affective state, by instructing participants to recall sad personal memories while listening to "sad" music, also increased the participants' vigilance towards depression-related, over anxiety-related or neutral words (Bradley et al. 1997). This effect was measured by presenting a pair of words, one emotional and one neutral, for a short duration and after a short delay presenting a probe word in the same location as one of the words in the pair. The participants then responded to the location of the probe word and were faster when they were in a negative mood and the probe word appeared in the same position as the depression-related

word of the pair. In a second experiment, highly anxious individuals exhibited greater sensitivity to threatening words and for smaller exposure times to the word pair compared to the participants with acute negative mood changes in the first experiment. These results were in agreement with earlier experiments showing that depressed patients had an attention bias only when the words were clearly presented without masking and the presentation duration was long (Mathews et al. 1996).

A meta-analysis of attention bias studies in highly anxious individuals, both clinically diagnosed and self-reported, indicated consistent results, across age groups and stimuli used, which typically consist of emotional faces or words with a threatening or neutral meaning (Bar-Haim et al. 2007). Although mood-induced biases potentially interact with the baseline capacity of a person to control their attention, as shown by a study involving a self-reported index of attention control prior to the experiment (Derryberry et al. 2002). Most of the studies around attention biases in anxiety included people with high levels of trait anxiety. The results of acute changes in affective state of healthy or anxious individuals were conflicting, after either experimental manipulations (in the form of newspaper photographs, Richards et al. 1992) or naturally occurring stressful events (testing several days prior to the start of an exam period, MacLeod et al. 1992). In healthy individuals though, biases in attention have been associated with resilience to stress (Thoern et al. 2016). Among participants that showed an attentional bias to emotional over neutral faces, those with a higher bias towards happy over angry faces scored higher on a self-reported stress resilience index.

The studies on attention biases discussed thus far describe an elevated attention towards and difficulty to disengage from threatening stimuli, particularly under an anxious affective state. However, biases on the opposite direction have been measured, when humans shift their attention away from threatening stimuli (Cisler et al. 2010 for a review of findings of all types of attention bias in anxiety). The choice of the location to attend to has been shown to vary with the timing of the presented threat in various studies discussed within the review. For instance, self-reported arachnophobes were fixating on a spider stimulus, while it was first presented, while shifting their gaze to a neutral location briefly afterwards (Pflugshaupt et al. 2005). This behaviour is thought to be a result of an individual's effort to regulate their affective state; initially fixating on a potential threat and then suppressing the stressful state by avoiding the same stimulus (Eippert et al. 2007).

1.1.4 Bias in effort exertion & vigor of movement

Moving away from the cognitive processes of interpretation, recall and attention, that might precede an action, there exists evidence for the effect of affective state on the execution of actions. The vigor of executed actions is particularly intertwined with affective state; "why do we run towards people we love and only walk towards others?" (Shadmehr et al. 2020). Shadmehr and Ahmed offer examples about this relationship, along with a theory about how the subjective value of potential actions affects the intensity at which they are executed and how a system that controls both motivation and valuation might have evolved.

Similar ideas gave rise to behavioural tasks, which aimed at measuring the relationship between anhedonia and the willingness to exert effort for reward. Anhedonia is defined as the reduced motivation in acting to acquire reward, or reward *wanting* (not to be confused with the enjoyment of already available reward, or reward *liking*, see Berridge et al. 1998; Gygax 2017). It has been shown to be a distinctive symptom of depression, thus closely associated with a negative affective state (Brown et al. 1998). There are multiple studies relating anhedonia with a negative affective state in depression and schizophrenia, although the importance of factors that are irrelevant to affective state, such as behavioural activation and the ability to predict future events in order to act with appropriate vigor, can not be understated (Salamone et al. 2016a).

An example of a task that measured anhedonia through action vigor was the effort expenditure for reward task (Treadway et al. 2009). Participants were required to press a button multiple times in order to gradually fill a bar on a computer screen. The number of required button presses varied between a hard, effortful condition and an easy one. Self-reported measures of high state and trait anhedonia through questionnaires were reversely proportional to the effort that participants were willing to exert. Moreover, patients diagnosed with major depressive disorder applied the least amount of effort in the same task, followed by people with subsyndromal depression, patients with remitted symptoms and healthy controls (Yang et al. 2014). These results were in agreement with an earlier task, where participants exerted effort by flexing their wrist to rotate a joystick, that in turn controlled the distance of a cursor to a goal location (Caligiuri et al. 2000 Jan-Mar). A group of depressed patients were shown to apply significantly less effort than the control group. An acute positive manipulation of affective state by treating healthy people with amphetamine yielded the opposite result (Wardle et al. 2011). People under

the drug treatment exerted more effort than controls, particularly in trials where the indicated probability of reward was at a minimum.

The amphetamine study was motivated by a previous study in the effort expenditure task in rats (Bardgett et al. 2009). In this study animals had to choose between a low-effort and low-reward action and a high-effort and high-reward alternative. Amphetamine made the animals more likely to choose to exert the required effort for the high reward, while there was no manipulation to the probability of reward. In another study, induction of a negative affective state by restraining the animals for an hour before the task led to reduced choices for the high-reward option and longer action latencies overall, which reflected the results of the human studies (Shafiei et al. 2012). Complementary tasks, within the same study, revealed that the effect of restraint stress was more likely to be specific to the animals' decision on the exerted effort rather than to their valuation of the different reward amounts or to the temporal discounting of delayed rewards. Dopamine has been thought to play a central role in the effort-related motivational deficits of depression. The animal version of the task has enabled more invasive and targeted research specifically on the role of dopamine (Salamone et al. 2016b for a review of studies that combine the effort expenditure task with physiological measurements and interventions).

Measuring vigor has diagnostic value when inferring an individual's state of depression. A study monitored the movements of participants, who wore a watch-like activity monitoring device for one month (Todder et al. 2009) . Elevated vigor of movements was associated with remission in depression symptoms and a more positive affective state during daytime activities, as indicated by the chosen questionnaires. In particular, vigor of movement was a more reliable predictor than quantity of the executed movements.

1.1.5 Learning bias

The pathologically negative affective state of depressed people has additionally been shown to affect their responsiveness to reward. A probabilistic learning task measured such a bias by presenting a simplistic drawing of a human face with varying lengths of the single mouth line (Pizzagalli et al. 2005). Participants then had to judge whether the length of the mouth was long or short, while one length category lead to a monetary reward three times as often as the other one. People that scored high on a depression index questionnaire chose the more frequently rewarded action

significantly less often than healthy controls.

A subsequent study on clinically diagnosed patients with depression revealed the same lack of a response bias, supporting the claim about a limited capacity to adapt the action selection process, given a history of relevant rewards (Pizzagalli et al. 2008). This finding was replicated on a later study, which further measured a smaller bias towards the richly rewarded option in depressed patients with elevated anhedonia, as opposed to ones with a low anhedonia index (Vrieze et al. 2013). This result links the reward learning deficits with the diminished vigor of movement that was discussed previously, with the common element being the presence of anhedonia. A reduction in the value that individuals assign to potential actions could be the outcome of the learning deficit and a causal factor behind the vigor deficit.

Inducing a negative affective state via acute manipulations has resulted in deficits akin to those measured in patient groups. Performing the reward learning task presented above under a threat of electric shock reduced the response bias towards the more rewarded action, relative to a condition without a threat (Bogdan et al. 2006). The stressful condition induced by the threat of shock was shown to cause a negative affective state through multiple questionnaires prior to and following the task. Another study involving choices between two cues, whose outcomes had to be learned, found that threat of shock resulted in diminished response bias towards the optimal action only when the action led to reward and not for an action that led to the avoidance of punishment (Berghorst et al. 2013). The effect was pronounced in the subjects, for whom the threat of shock was most successful in inducing a negative affective state, thus highlighting the influence of affective state on adaptations to rewarded actions.

A version of the probabilistic learning task was designed for rats as well (Der-Avakian et al. 2013). In it, animals were initially trained to discriminate between two reference tones of distinct durations and the same frequency by pressing the matching lever. During a subsequent test session, two tones of intermediate duration were presented as ambiguous cues, with one of them being three times more likely as the other one to lead to a reward after a correct action. Rats exhibited a response bias in choosing the richly rewarded action more often than its alternative during either ambiguous cue and this bias was potentiated under an amphetamine treatment. Conversely, a negative manipulation to the animals' affective state, exposure to social defeat across three days, diminished the response bias, suggesting that animals were less able to adapt to the reward feedback under this state (Der-Avakian et al.

2017).

A behavioural task that is related to the probabilistic learning one is the probabilistic reversal learning task. In this case, subjects learn cue-action-probability for reward contingencies across trials, for two simultaneously presented cues, and after a number of trials the reward probabilities swap. Depressed patients were not able to follow the optimal strategy either before of after the reversal of reward probabilities, as they switched their choices more frequently after a negative feedback (Murphy et al. 2003). Similar results were found in a subsequent study, thus corroborating the increased sensitivity to the omission of reward (Taylor Tavares et al. 2008). However, acute positive, negative or neutral affective manipulations, in the form of videos, did not cause any changes in the strategies that healthy people employed during this task (Nusbaum et al. 2018).

Rats performing the probabilistic reversal learning task after an acute treatment with citalopram, an antidepressant SSRI drug, were less sensitive to switching their preferred choice after negative feedback compared to controls (Bari et al. 2010). Moreover, daily treatment, across a week, with the same compound led to an elevated sensitivity to positive feedback, while both treatments caused a better adaptation to rule changes, when reward probabilities swapped. An acute ketamine treatment in another study led to a selective decrease in negative feedback sensitivity, without affecting the rats' actions after positive feedback (Rychlik et al. 2017).

1.2 Inference by statistical models

In the previous section, inference on affective state was conducted by analysing summary statistics of subjects' actions. This was a type of direct inference, immediately linking the summary statistics with affective state. The present section provides an overview of statistical models that can extend this inference, by accounting for individual data points, such as actions in a sequence of trials, and by allowing for multiple factors to influence these actions. This way, the effect of affective state on actions can be dissected into individual components and the relative importance of each component can be quantified using standard statistical models. Avoiding data summaries, particularly across time, can enable the correlation of the latent factors, that are hypothesised to constitute affective state, with other time-sensitive measures, such as electrophysiological recordings, fMRI and questionnaires.

Commonly used statistical models fall into the category of generalised linear models (GLM) (McCullagh et al. 2019). The structure of a GLM could be broken down into three primary components; a distribution of the data, for example a Binomial distribution for actions in a two-alternative choice task, a linear function, including slope parameters that quantify the effect of the factors under consideration and intercepts for factor-independent effects, and a nonlinear *link* function connecting the linear function to the parameters of the data distribution.

One study used a GLM on an effort expenditure for reward task, as it was presented above, to dissect the effects of several factors, which could potentially drive the actions of healthy and depressed participants (Treadway et al. 2012). The probability of being rewarded, the reward amount and the required effort were shown to participants in the beginning of each trial. By incorporating these three factors as potential drivers of the participants' actions, the authors found that depressed people were less willing to exert effort and this behaviour was more insensitive to both reward probabilities and amounts, compared to healthy controls.

Linear models have been used to assess whether reward magnitude and its surprisingness could modulate affective state. An initial study implemented a two-choice task between an action with a certain reward and a gamble with known probabilities and possible outcomes, along with intermediate questions regarding the participant's happiness (Rutledge et al. 2014). The fitted model parameters indicated that the unexpectedness of reward, as measured by a reward prediction error (RPE), was a better predictor of the participants' happiness scores, as opposed to the magnitude of the earnings. In a follow-up experiment, patients with major depressive disorder exhibited the same influence of RPEs to their affective state (Rutledge et al. 2017). Moreover, the same RPEs correlated with blood oxygen level activity within the ventral stratium, a region that has been thought to represent RPEs, almost equally well in depressed patients and controls. Similar results around RPEs were found in a more ecological setting (Villano et al. 2020). The unexpectedness of an exam result was a better predictor of both a positive or negative affective state compared to the exam grade. Additionally, a negative unexpected exam outcome could predict a negative affective state for longer, as opposed to a pleasant surprise.

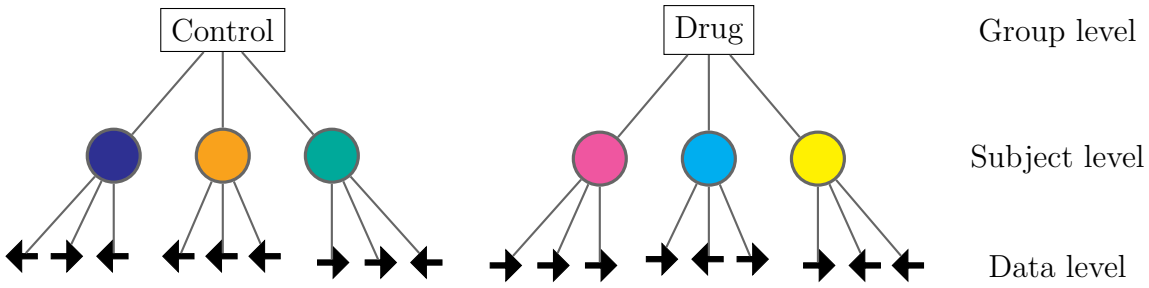


Figure 1.1: **An example of a hierarchical model**: There are two experimental groups, one under a drug that is hypothesised to alter the subjects' affective state and a control one. Each group includes the subjects under the matching condition and each subject generates actions during the task. The actions are denoted as left or right arrows.

1.2.1 Hierarchical statistical models

Statistical models, such as the GLMs of the previous section, could account for multiple factors that affect subjects' actions. On their own though, they are unable to incorporate information about the structure of the observed data. Namely, multiple actions are generated by multiple subjects that either belong to different population groups, for example diagnosed patients and healthy controls, or perform the task under different experimental conditions, such as a vehicle and a drug treatment, purported to change their affective state (Figure 1.1). Thus, the data has a hierarchical structure with clusters of actions, subjects and population or treatment groups. Another common term for this structure is nested data.

Important heterogeneity can exist within each level of clustering (Bell et al. 2015). A common source of heterogeneity is individual subject differences; even though an experimental group includes people diagnosed with depression, it should not be expected that all of them would behave in an identical manner in the experiment. Equally, the effect of a drug could vary across the subjects that were treated by it. Generalised linear models could account for such sources of heterogeneity within each level of the observed data by being embedded in a hierarchical model (Leeuw et al. 2008). In a hierarchical model, parameters, such as intercept and slope terms, for each subject are assumed to be samples out of a distribution that is common for all subjects within a group, that is a population or an experimental condition. The result of an inference process in this case would be both subject-specific and group-specific parameters, with the latter representing the parameters of the group-level distributions.

A hierarchical model with a GLM incorporated within it for each subject was fitted

on observational data in an urban environment to test whether the unexpectedness of rewards also influences affective state outside the lab (Otto et al. 2016). People residing within distinct zip codes in New York City were organised in groups according to an index of the socioeconomic status of each neighbourhood. It was then shown that people were more likely to gamble after recent unexpected positive events that were germane to them, such as a local sports team winning or local weather conditions being better than expected. The effects of both sports- and weather-related unexpected outcomes (slope parameters) were similar for the high and low socioeconomic group, while the latter group exhibited higher gambling activity overall (intercept parameter). This observation links to an optimism bias that was hypothesised to be the result of a more positive affective state.

Using GLMs for individual subjects without introducing a group level, to cluster subjects under, is feasible and allows for individual differences. However, in a simulation study using synthetic data, it was shown that the use of an explicit hierarchical structure in conjunction with GLMs when analysing nested data could reduce the false positive rate for effects and increase statistical power to detect an experimental effect (Aarts et al. 2015). Using data from published studies to inform the number of observations within each group, or cluster, another simulation study corroborated the same claim about the inflation of false positive rates and used visualisations of the data within each group to discuss this observation (Aarts et al. 2014). Moreover, hierarchical models can be used without GLMs, as a way to calculate summary statistics while accounting for the hierarchical structure of the data via hierarchical bootstrapping, which partially accounts for the variability within each group (Saravanan et al. 2020).

There are technical benefits to hierarchical models that underlie the inferential benefits around false positive rates and statistical power. In particular, parameters of group members are being drawn towards a group average during inference, which is a way to regularise inference and deal with outlier data (McElreath 2016, chapter 13). This process is known as partial pooling, which lies in between assuming identical parameters for all subjects (complete pooling) and assuming that each subjects' parameters are independent (no pooling), as the examples of the previous section did. The benefits of partial pooling will be more pronounced in the following section, where Bayesian inference will be introduced as a way of calculating the model parameters that best describe the data.

1.2.2 Bayesian inference

The inference process discussed previously, whether GLMs were part of a hierarchical structure or not, resulted in point estimates. That is, the inferred model parameters were scalar numbers, representing the parameter values that best describe the data. This type of inference is commonly called frequentist. This section will present an alternative approach, known as Bayesian inference (Gelman et al. 2015; McElreath 2016, but see Quintana et al. 2018 for a discussion on the logic behind each approach and equivalent statistical tests between them). Bayesian inference is based upon the Bayes rule from probability theory

$$P_m(\theta|D) = \frac{P_m(D|\theta)P_m(\theta)}{P_m(D)} \tag{1.1}$$

where θ are the model parameters, such as the intercept and slope terms of a GLM for individual subjects and parameters of the population-level distributions, D is the observed data, representing a subject's actions, P represents a probability and the $|$ symbol denotes a conditional probability. Then $P_m(D|\theta)$ is the likelihood of observing the data given, or conditioned on, the model parameters θ, $P_m(\theta)$ is the prior distribution of model parameters and $P_m(D)$ is the probability of the data, irrespective of model parameters, which is also known as the marginal likelihood or model evidence. Combining these three probability terms in Bayes rule yields the probability of the model parameters after observing the data, $P_m(\theta|D)$, that is the posterior distribution of model parameters. The index m on each term indicates that the inference process is dependent upon a chosen model m, for instance a hierarchical model with GLMs for each subject.

The result of Bayesian inference is a joint distribution of possible values for all model parameters, that is the posterior distribution. In order to calculate it, the likelihood of observing the data, given a model and its parameters, are combined with a belief about possible model parameters before observing the data. Thus, the prior distribution is transformed into the posterior via the likelihood distribution.

Instead of the point estimates of frequentist inference, Bayesian inference yields a joint distribution on the model parameters, thus quantifying the uncertainty around the values of each parameter. This uncertainty is valuable for evaluating the extent at which an effect was present and for performing model comparisons between multiple possible models, each one representing a separate hypothesis about how data was generated (McElreath 2016, chapter 7).

Related to the benefit of having an estimate of uncertainty is the pronounced advantage of the partial pooling property of hierarchical models in Bayesian inference. There is a bidirectional influence between lower- and upper-level parameters within the hierarchy that constitutes partial pooling (McElreath 2016, chapter 13). For instance, subject-level parameters will be chosen to match individual subject data and their values will inform the selection of the parameters of the group-level distributions, that should include the individual subjects. In turn, the group-level parameters constrain the values that individual subject parameters can have, as the subjects are samples out of the groups. Thus, uncertainty about individual subjects will propagate upwards to uncertainty about the corresponding group, which can consequently inform parameter comparisons. Furthermore, subject-level parameters are pooled closer to the expected value of their group-level distribution, with the extent of this pooling being weighted by the uncertainty around the likelihood of observing a subject's data given their parameters.

In the current example of a hierarchical model, the subject-level parameters essentially have an adaptive prior distribution, which is shaped by the group-level parameters as they adapt to the data during inference (McElreath 2016, chapter 13). However caution must be taken when defining priors for the group-level parameters themselves. Different choices of prior distributions for the highest level of the hierarchy could both hide or yield an effect with high certainty, as it was shown in an analysis of simulated data of healthy and depressed individuals making choices in a probabilistic learning task (Valton et al. 2020).

The product of the likelihood and prior distributions, $P_m(D|\theta)P_m(\theta)$, constitutes a generative model of how the data and the model parameters are jointly distributed, $P(D, \theta)$. This means that the product could be used to simulate fictive data, first by drawing samples out of the prior and then by using these parameter values to sample fictive data out of the likelihood. This step is important both before inference, in order to inspect the data that the model expects to observe, and after inference, by using the posterior in place of the prior (Gelman et al. 2020 and Wilson et al. 2019 for a discussion of the benefits of simulating data specifically for behavioural tasks). Simulating fictive data using the prior could help to diagnose unreasonable prior distributions, particularly on the highest group level of the hierarchy. The case of producing fictive data after inference could be a useful process for assessing the quality of inference by comparing the generated with the observed data and for using the generated data as a prediction of future observations, that could for example inform future experiment designs (van de Schoot et al. 2021).

The frequentist inference framework comes with multiple statistical tests. Reliable inference in this case depends upon the type of data, strict cut-off thresholds for significance and correction after the tests (McElreath 2016 chapter 1 for an overview). This fact could make correct inference harder to accomplish and in many cases to replicate as well. On the other hand, the Bayesian framework offers a simpler process to achieve good quality in the inference results. This process relies solely on Bayes rule and mathematical operations on the posterior distribution results, for example taking the difference between the posterior of a parameter from a drug treatment group and the same parameter from the vehicle control group to calculate what is the probability of a nonzero effect.

Nevertheless, Bayesian inference often is not as straightforward as an application of the Bayes rule. The marginal likelihood term is calculated as

$$P_m(D) = \int P_m(D|\theta)P_m(\theta)d\theta$$

the integral of the product of the likelihood and the prior over all model parameters θ is often too computationally expensive or intractable to calculate. Two families of methods offer a solution to this issue by approximating the posterior distribution without calculating the marginal likelihood: variational inference and sampling-based inference. Variational inference approximates the posterior distribution with a known distribution, which is called the variational distribution (MacKay 2003, chapter 4). The parameters of the variational distribution are then determined by minimising a distance metric between the variational distribution and the actual posterior. Given Bayes rule, the actual posterior within such a distance metric is decomposed into a likelihood and a prior, thus circumventing the issue of having to calculate the marginal likelihood. This approximation is particularly accurate for specific pairs of likelihood and prior distributions that fulfill the conjugate prior property (Fink 1997).

The second family of methods approximates the actual posterior by drawing samples from it. It does not calculate the marginal likelihood, however it does perform a normalisation to the samples so that the result is a probability density function. Algorithms belonging to this family will be employed in the current thesis and they are presented in the following section.

Sampling-based approximate Bayesian inference

Approximating the posterior distribution by drawing samples from it can be achieved by multiple algorithms (McElreath 2016, chapter 9). A common element across these methods is that they build a sequence, or chain, of samples, that follows the Markovian property. That is, each new sample in the chain will only depend on the previous one in the sequence. This family of methods is called Markov chain Monte Carlo (MCMC). Methods of this family use different processes for taking a new sample and deciding on whether to attach it to the sequence or redraw.

A commonly used MCMC method that can draw samples efficiently and robustly is the No-U turns sampler (NUTS), an extension of the Hamiltonian Monte Carlo (HMC) method (Hoffman et al. 2014). The goal of HMC is to move around the typical set of the posterior distribution, that is the area in parameter space where most of the probability mass of the posterior lies, and draw independent samples from it (Duane et al. 1987; Betancourt 2018). In order to achieve this without diverging away from the target region, it utilises the notion of Hamiltonian mechanics, that is a way to describe the trajectory of dynamical systems. The posterior probability density function is mapped to a potential energy term, while a kinetic energy term is calculated through momentum values for each parameter. The momentum values are sampled according to the respective parameter values using various distributions, with the Gaussian distribution being a typical choice.

The sum of the potential and kinetic energy terms constitutes the Hamiltonian, which is then used to draw a trajectory in the extended space that includes both parameters and their momentum vectors. This way of drawing a trajectory allows for the exploration of the parameter space in closed contours, where the total Hamiltonian is approximately preserved and there is only an exchange of energy between potential and kinetic (Betancourt 2018). Consequently, the trajectories do not diverge away from the region of importance (Figure 1.2). As the goal of the method is to draw independent samples, the NUTS sampler allows the trajectory to move in parameter space until it begins to move towards its starting point, making a U turn. At that point the trajectory is stopped and the sampler decides on whether to keep this final point as the next sample or reject it and start again. The decision is based upon the ratio of probability densities at the initial and the final point of the trajectory.

In addition to the simulation-based checks that were discussed in the previous sec-

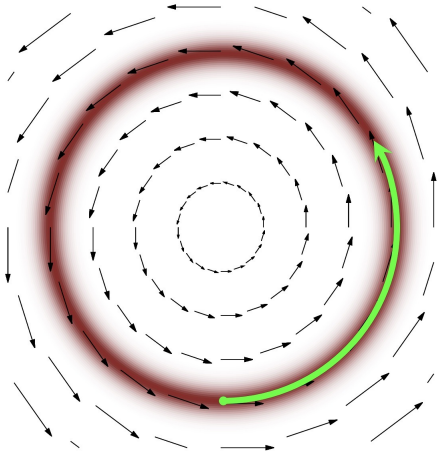


Figure 1.2: **A trajectory during inference by Hamiltonian Monte Carlo methods:** The red band represents the typical set, or the target region, of the posterior distribution to be explored. The ideal case for the trajectory in green is to move within this band to maximise the number of representative samples. This is achieved in HMC methods by using the Hamiltonian, the sum of a potential and a kinetic energy term, to define the contours, which constrain the directions that trajectories can move through. Each contour is represented by arrows that draw a closed curve. In the NUTS variant of HMC, the current trajectory will stop before it starts to come around the left side of the circle towards its initial point and a decision will be made on whether to keep the sample or not. Subsequently, a new initial point will be chosen close to the red band and a new trajectory will be drawn. The HMC methods negotiate a balance between kinetic and potential energy to sample points and draw trajectories as close to the red band as possible. Taken from Betancourt 2018.

tion, there are multiple diagnostic checks specific to sampling-based inference methods (Baribault et al. 2021; Vehtari et al. 2021b). Thus, the extent at which the posterior distribution was explored sufficiently is verifiable. These diagnostics will be presented more thoroughly in the following chapters when they are applied to the posterior samples.

1.3 Inference by theoretical models

The final approach to inferring affective state through observations of a subject's actions concerns theoretical models of behaviour. Such models are mathematical descriptions of mental processes that precede the selection of an action. They represent distinct hypotheses about how actions are being generated, given sensory input and any internal process, such as memory replay. The two broad categories where the influence of affective state has been described mathematically are learning and

decision-making tasks.

A theoretical model is defined by a set of parameters that represent different factors and a set of mathematical equations that describe how these factors give rise to actions. The difference between theoretical models and the statistical models of the previous section is that theoretical models are domain-specific and based upon formal theories around learning and decision-making. In order to quantify the effect of affective state on the model parameters, tools from the previous section will be utilised. Models are fit to action data, via either frequentist or Bayesian inference, and the resultant parameters are analysed in order to assess how affective state modifies each one of them.

The application of theoretical models particularly to the behaviour of patients with psychiatric disorders, including anxiety and depression, falls under the recent field of computational psychiatry (Friston et al. 2014; Huys et al. 2016). The combination of behavioural tasks with theoretical models of learning and decision-making, informed by observations of actions, has contributed to a better understanding of mood disorders (Huys et al. 2020). Examples of such applications, along with studies with acute affective state manipulations on healthy subjects, will be presented in the following sections.

1.3.1 Decision-making models

A classical approach to modelling decision processes is the drift-diffusion model (DDM) (Ratcliff et al. 2008 for an overview of applications). The DDM (Figure 1.3) models the real-time integration of sensory evidence towards one of two alternative interpretations as a stochastic process. The drift of the process corresponds to the ratio of the likelihood of the evidence supporting one interpretation over its alternative, while the added diffusion on the drift represents the quality of the accumulated evidence. The result is a noisy integration of the sensory input towards one of two decision bounds, with each bound corresponding to an action. Moreover, the model is able to account for biases towards any action via the position of the starting point of the drift-diffusion process, relative to the decision bounds.

The DDM has been used to explain the behaviour of rats in the Judgement Bias Task, where the interpretation of an ambiguous cue could lead to either a high or a low reward (Hales et al. 2016). It was shown that the negative affective state after an anxiogenic drug treatment or after repeated restraint stress and social isolation led to

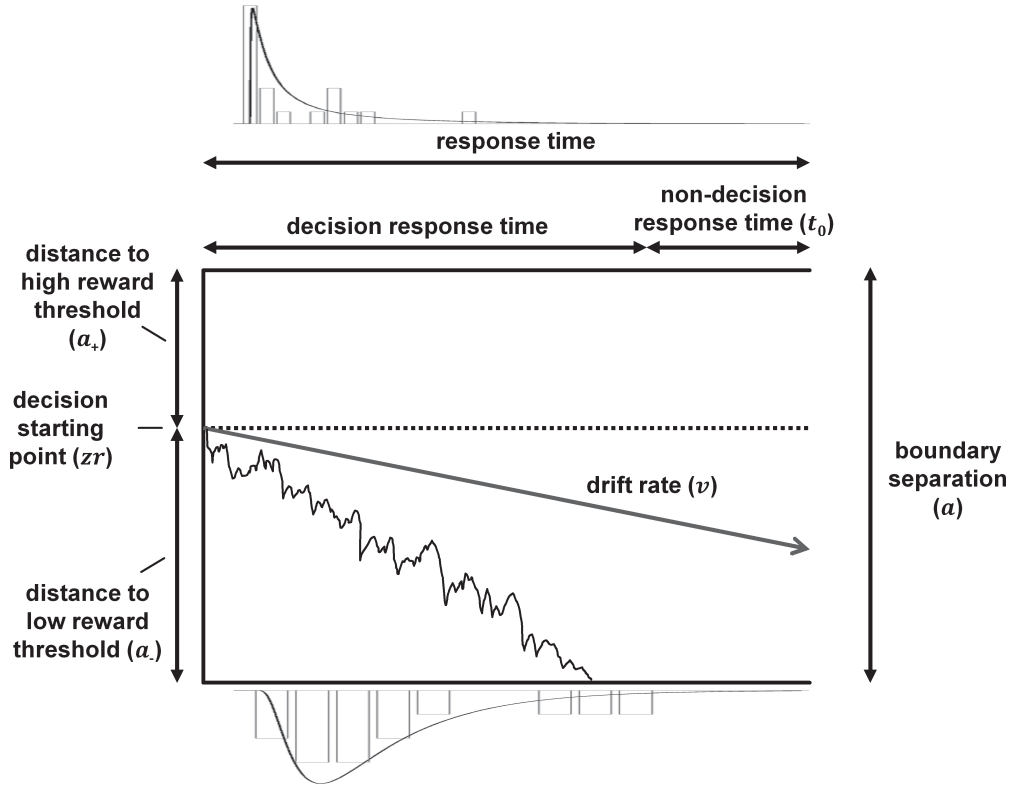


Figure 1.3: **The drift-diffusion model of decision-making:** The response time of a decision is modelled as a noisy evidence integration process. The evidence accumulation initiates at the decision starting point and proceeds according to a drift and a diffusion term towards either one of the decision bounds. When a bound is reached, the corresponding action is taken. The model is able to account for multiple factors that are thought to be important in a decision process; a bias towards an action via the distances between the starting point and the bounds, a non-decision response time potentially due to the execution of the action, noise in the sensory evidence to be accumulated via the diffusion term and interpretation of the sensory input via the drift. Taken from Hales et al. 2016.

an increased interpretation of the ambiguous cue as the reference cue associated with the low reward. A reduced anticipation of high reward and an elevated anticipation of low reward were measured under the negative state, indicated by the position of the decision starting point and the distance between the bounds. A follow-up study using positive affective manipulations found comparable changes to the summary statistics of the animals' actions under chronic treatment with fluoxetine, an SSRI antidepressant, and acute treatment with ketamine, an NMDA receptor antagonist and novel antidepressant (Hales et al. 2017). However the model revealed differences in how the drugs affected the decision-making process. Ketamine resulted in a more optimistic interpretation of the ambiguous cue towards the high-reward cue, whereas fluoxetine increased the decision starting point towards the high-reward action bound. The same version of JBT was translated into a human task (Aylward

et al. 2020). In this case, symptomatic people with anxiety disorder exhibited a more negative drift rate, an observation that suggests an interpretation of the ambiguous cue as the low-reward one.

Other studies have used Bayesian decision models to investigate the effect of affective state. These models have been shown to be equivalent to the DDM model under certain conditions (Bitzer et al. 2014; Fard et al. 2017). The Bayesian models assume that a subject is using Bayes rule, as it was presented in the previous section, to reason about incoming sensory input and potential outcomes that alternative interpretations could lead to (Whiteley et al. 2008). Thus, unlike DDM these models make use of information about the magnitude of potential outcomes.

A JBT variant in humans used visual stimuli and a choice to either take a gamble, for a monetary win or loss, or opt out (Iigaya et al. 2016). The actions of participants were fitted to a Bayesian decision model using variational inference and a hierarchical model structure of trials, subjects and two groups. People that were given a bag of sweets prior to the task and placed into a pleasantly decorated room for the experiment were shown to value losses more than wins and were less risk-averse, compared to a second group of participants, who were placed in an undecorated room and received no food reward beforehand.

In a similar JBT variant with auditory cues, rewards and punishments, it was shown that when rats experienced more rewards shortly before a test session, they were subsequently more risk-averse when interpreting the ambiguous cues (Neville et al. 2020b). Conversely, rats that were previously exposed to multiple air puffs exhibited a more optimistic bias, since they interpreted the ambiguous cue as one leading to reward and not punishment. A model fitting process similar to the human version was employed, using a hierarchical model and variational inference. Model parameters for the group of rats in the reward condition revealed an increased sensitivity to punishments and a bias towards the risk-averse action, while conversely the air-puff group had an increased sensitivity to reward.

A more recent JBT variant in humans used groups of dots on a screen moving randomly towards the left or right direction, with various degrees of coherence between the direction of individual dots (Neville et al. 2021). The average movement direction of the group determined whether the outcome would be a win or a loss of known magnitude. Participants had to keep pressing a button to achieve the outcome or stop pressing it to opt out of a trial. The authors extended the Bayesian decision model by adding a GLM in order to account for the influence of the recent

history of reinforcement on the subjects' actions. Periods of high past earnings and positive prediction errors were correlated with positive affect, as measured by intermittent questionnaires, and were biasing the subjects away from the risky action. Furthermore, participants in a more negative affective state were less able to correctly estimate the amount of time between initiating the button press and receiving the outcome.

The DDM model has also been applied to the effort expenditure for reward task, where depression patients and healthy controls had to press a button multiple times for a high reward or choose a less effortful low-reward alternative (Berwian et al. 2020). The drift rate of the model was the difference between a weighted sum of the effort and the reward of the high-effort action and the equivalent sum for the low-effort one. Patients were more sensitive to the required effort, as their drift rate was less positive, resulting in more low-effort actions. Moreover, the amount of evidence required to make a decision for the low-effort option, as measured by the low-effort decision bound, could predict the relapse of patients within 6 months of stopping their medication.

1.3.2 Learning models

The role of affective state is important in cases when the potential outcomes of actions are not known a priori and need to be learned via interaction. Reinforcement learning is the primary framework that describes such interactions between an agent and its environment (Sutton et al. 2020). A reinforcement learning agent acts in order to change the state of its environment and receives rewards or punishment from it as an outcome. In doing so, the agent learns about the optimal actions to take in a given state of the environment in order to maximise its accumulated reward in the future.

States of an environment could be thought of as sensory input to the agent. The process of learning constitutes updating an estimate of how beneficial an action is in a certain state. The estimate of an action in a state reflects either a preference for the action given a behavioural policy or the expected outcome that this action would lead to (see Bennett et al. 2021 for a discussion on the distinction between preferences and expected outcomes). In both cases, the running estimate of an action is compared to the experienced outcome during learning. This difference is the reward prediction error (RPE), which guides the update of the action estimate

via a learning rate parameter that determines how much distant experiences should be weighted relative to recent ones. The final component of a reinforcement learning agent is an action selection rule that weighs the expectations of available actions and decides on which one to take.

Reinforcement learning models have been widely used across learning experiments, particularly in conjunction with the tasks that were discussed in section 1.1.5 (Huys et al. 2013). They have aided in the characterisation of differences between patient and control groups in sensitivity to positive and negative feedback, when learning the expected outcome of actions (Pike et al. 2022).

Probabilistic learning tasks have been used in conjunction with reinforcement learning models to elucidate how affective state modulates the adaptation to feedback. One study included actions that could lead to both a win and a loss with independent probabilities (Pulcu et al. 2017). The authors manipulated the probabilities of wins and losses separately to distinguish the effect of positive and negative feedback on learning. Each probability was either in a stable or a volatile condition, where it varied frequently across trials. A reinforcement learning model with distinct learning rates for wins and losses was fitted to the choices of healthy humans performing the task. The model parameters that were retrieved indicated that participants had an elevated learning rate for the probability that was volatile within each block of trials, suggesting that they were following the most informative feedback.

Using the same probabilistic learning protocol with stable and volatile conditions, another study found that highly anxious individuals were less able to adjust their learning rate when the condition changed (Browning et al. 2015). Individuals with a high anxiety index were using a higher baseline learning rate in a probabilistic reversal learning study, which involved choices between three options, whose the probability of reward varied across three values (Huang et al. 2017). The higher learning rates led to an elevated lose-shift behaviour due to the higher sensitivity to recent negative feedback. A subsequent article discussed these and similar findings and suggested that people with anxiety and depression possibly misestimate the amount of uncertainty in their environment and find negative events more informative to focus on (Pulcu et al. 2019).

A disproportionate focus on negative feedback has also been observed in healthy humans, when performing a multi-step probabilistic learning task (Eldar et al. 2020). Participants' actions given a visual cue could either result in an immediate outcome or in another cue, where a second action would lead to an outcome. By

manipulating the cue-cue and cue-outcome contingencies, while recording magnetoencephalographic (MEG) data, the authors observed that the best performing participants had a propensity to replay trials that resulted in a negative feedback. This occurred when participants were not engaged with the task and it was detected by matching the MEG activity during the task with that during periods of rest, in between blocks of trials. Subsequent work using an extended reinforcement learning model suggested that replaying negative feedback is beneficial in cases when learned contingencies are being forgotten quickly or the learner is uncertain about them (Antonov et al. 2021).

The parameters of reinforcement learning models could interact in ways that are not immediately obvious. For instance, in a similar probabilistic learning task with fixed probabilities of wins and losses, participants were split into a high and a low depression index group, as judged by a questionnaire (Kunisato et al. 2012). The high depression index group was found to have a higher average learning rate for losses. However, a subsequent meta-analysis of multiple studies involving people with depression indicated that this adaptation deficit was more likely to be due to a diminished reward sensitivity, rather than an increased learning rate (Huys et al. 2013). In this work, reward sensitivity was modelled as a multiplier to the acquired reward, which is then incorporated into the value of the relevant action.

The learning models discussed so far have described ways in which affective state influences learning. However affective state could also be modulated by ongoing experiences. For example, the effect of positive and negative surprises, corresponding to positive and negative RPEs, has already been discussed in section 1.2. A study investigated the reciprocal relationship between affect and learning by extending a classical reinforcement learning model with a mood parameter, which was fit on the actions of humans during a probabilistic learning task (Eldar et al. 2015). Mood was calculated as a running average of RPEs and it modulated the sensitivity to wins and losses. Intermittent trials of a wheel of fortune producing an unexpected outcome were implemented as a means to cause sudden changes in mood. The mood parameter correlated closely with scores from a self-reported mood questionnaire, suggesting that the model could reliably infer the participants mood. Additionally, a mood instability score was correlated with a positive modulation of outcome sensitivity by the mood parameter. This positive modulation is equivalent of a magnification of wins and losses and it could lead to mood instability, as model simulations indicated.

Another study that investigated how affective state changes during ongoing learning experience found that this effect was specific to learning-relevant surprises (Blain et al. 2020). This study implemented a probabilistic reward learning task in a stable or a volatile condition, where probabilities of reward reversed after a set amount of trials. Participants had to learn the probability of reward for each action, while the reward amount was irrelevant. The authors used a reinforcement learning model to update expectations and choose actions, along with a GLM to correlate prediction errors around probabilities and rewards with self-reported mood. Happiness was more sensitive to errors in the estimation of the probability rather than reward. Elevated depressive symptoms correlated even more with the modulation of affective state by probability prediction errors, particularly in the volatile condition.

1.4 Aims and objectives

The overarching aim of the present thesis was to develop reliable approaches for inferring the affective state of animals through their actions in decision-making and learning tasks.

Chapter 2 involved experimental work on the Judgement Bias Task. The objectives were initial replications of previous findings with pharmacological manipulations and the design of a new task variant in an effort to enhance the ambiguity of the testing cues.

Chapter 3 describes the design of a novel behavioural task with the goal of measuring affective-state related biases in animals' responses as they are conflicted between foraging for reward or fleeing from an imminent aversive event. The task was inspired by the conditioned suppression task that has been previously used, particularly as a model of fear with high-intensity inescapable punishments. The aim was to assess whether escapable aversive events of low intensity could be employed to elicit a suppression of the reward-seeking behaviour. This way our novel protocol could be an improvement in terms of animal welfare and it could be used for chronic studies. Complementary to the task design was a hierarchical Bayesian model that was designed to capture individual differences in the animals' behaviour.

Chapter 4 returns to the JBT task with the aim of looking into the animals actions in more detail, by considering the actual sequence of trials rather than averaging across an experimental session. A large scale analysis of previously acquired JBT data

suggested that factors such as the history of reinforcement and the feedback of the most recent trial were important determinants of the animals' actions. Hierarchical Bayesian models were designed and fitted to a large dataset of animals' actions during JBT. These models combined principles of the decision models, discussed in section 1.3.1, and the learning models of section 1.3.2, in order to capture both the interpretation of the ambiguous stimuli and the influences from the past. Each model represented a distinct hypothesis about which factors were most important. A model comparison was performed to evaluate the probability of each model being correct. Differences between animal populations originating in distinct breeding facilities and between the pharmacological manipulations of chapter 2 were investigated in light of the most accurate model.

Finally, the work of chapter 5 was inspired by the action biases, which were measured in the Affective Bias Test (ABT) and involved non-immediate contingencies between actions and outcomes. We designed a novel reinforcement learning (RL) agent model that could account for the ABT observations by incorporating affective state in its learning process. Moreover, novel simulated tasks were designed and implemented in code, as examples of naturalistic environments that deviate from the standard trial-based structure that only involves immediate action-outcome contingencies. The functional benefits of our model were evaluated in these environments and compared against alternative models, as a way of assessing the functional benefits of including affective state in an RL agent.

Chapter 2

Judgement bias task: interpretation bias under ambiguity

2.1 Introduction

There are multiple versions of the Judgement Bias Task (JBT), also known as the Ambiguous Cue Interpretation task, varying in terms of sensory modality, outcome valence and reinforcement schedule among other factors (Roelofs et al. 2016 for a review). This chapter includes experimental work on two versions of the task in rats using reward outcomes, yet different reinforcement schedules and sensory stimuli. Besides these differences, the premise of the task remained the same: decisions under an ambiguous stimulus are primarily driven by expectations of outcomes. The valence and the magnitude of the expected outcomes in turn reflect an animal's affective state.

The JBT has exhibited efficacy as an animal model of antidepressant manipulations. Past experiments have measured a positive shift in the interpretation bias of rats under an acute ketamine treatment, while conventional antidepressants only had an effect after a chronic treatment (Hales et al. 2017). These results have supported the claim that ketamine is a compound with novel rapid-acting antidepressant properties (Carboni et al. 2021 for a recent review). Thus, the positive ketamine finding was an important one for the translational validity of the JBT. The first study of cohort 1 was designed as a replication study of this result.

Results from the Affective Bias Test (ABT), a different animal model of affective state manipulations, have indicated that administering ketamine 24 hours before a test session had an effect in mitigating a negative bias (unpublished data). The ABT measures affective biases of previously learned stimulus-outcome contingencies, however the differential influence of the drug depending on the pretreatment time was deemed relevant for JBT as well. Even though ketamine previously caused a rapid change in interpretation bias, the duration of this effect is unknown. Thus, the second study of cohort 1 measured the delayed effects of a ketamine treatment, which was administered 24 hours before testing.

Given a failure to replicate the positive shift in interpretation bias under ketamine, an amphetamine study was subsequently conducted in cohort 1 as a positive control. An acute amphetamine treatment has previously resulted in the largest positive shift in the rats' interpretation bias, as compared with ketamine, cocaine and conventional SSRI antidepressants (Hales et al. 2017). The effect of amphetamine was replicated successfully. Given this result and the fact that animals of cohort 1 were supplied by a different breeding facility, compared to animals that took part in the original ketamine experiments (Hales et al. 2017), a ketamine dose-response study was performed. The aim of this experiment was to assess whether the failed replication was due to a difference in sensitivity to the drug between cohort 1 and the animals of past studies.

In past JBT studies in rats, different frequency tones were used as reference cues and the ambiguous cue (AC) was a tone, whose frequency was between the two references (Harding et al. 2004; Hales et al. 2016; Hales et al. 2017; Jones et al. 2018). However, there is no common agreement across studies about which tone frequency would make AC purely ambiguous. Thus, setting the frequency for the AC constitutes a design choice, which might promote biased actions in JBT, caused by perceptual interpretations of the ambiguous cue as one of the other two cues.

Animals could learn new stimulus-action-outcome contingencies for the newly presented AC, akin to their past training with reference cues. Several JBT studies have measured a drift in the animals' responses to the AC over trials, which was indicative of learning from past feedback (Roelofs et al. 2016 for a review). Introducing fewer AC trials decreased the amount of drift, making CBI more stable, thus more likely to be driven by an interpretation bias, based on prior reward expectations, instead of a learned expectation for the value of each action.

Both the perceptual biases and learning effects could potentially confound the in-

terpretation of the animals' actions according to the premise of the task. Namely, animals could respond to the AC based on a perceptual interpretation of the cue or the feedback during past trials and not according to their expectation of a large or small reward.

A novel variant of the JBT was designed with the aim of addressing these confounding factors. Cohort 2 took part in the pilot run of this task. The reference cues were simultaneous tone and light presentations, with a common light component. The ambiguous cue consisted of a sole presentation of the same light stimulus. Additionally, there were fewer ambiguous trials within a test session, compared to the JBT variant of cohort 1. The reference cues resulted in probabilistic feedback, instead of always leading to the corresponding reward on correct choices, as was the case of the original JBT of cohort 1. The aim of this design was to enhance ambiguity by omitting information, as there was the familiar light but no sound during ambiguous trials, instead of placing the ambiguous stimulus on the same scale as the reference cues. Moreover, the ambiguous trials were further apart in time and omissions of reward were partially expected, since the animals were trained on probabilistic feedback. It was hypothesised that these changes would hinder any potential learning of new contingencies about the ambiguous cue and promote responses driven by an interpretation bias.

Cohort 2 was initially assessed in a test to evaluate how ambiguous the chosen light stimulus was with respect to the multimodal reference cues. Reward magnitudes for correct actions were kept the same in this experiment, so that responses would be purely driven by a perceptual interpretation of the AC.

Previous work in humans has measured disruptions in a perceptual decision-making task without differences in reward magnitudes, where it was suggested that ketamine could make the accumulation of ambiguous sensory stimuli more uncertain (Salvador et al. 2022). A potential change in perceptual processing under ketamine could confound the effect of the drug on the interpretation bias. Thus, in the second experiment of cohort 2, I tested whether ketamine could shift the perceptual interpretation of the AC, when both reward magnitudes were equal.

Following these experiments, it became evident that the animals could have been using the tone loudness when responding to the AC. In order to test this hypothesis, the third experiment of cohort 2 changed the loudness of the tone of one of the reference cues, while keeping the rewards magnitudes equal.

2.2 Methods

2.2.1 Animals

The experiments in this chapter were conducted on two cohorts of male Lister Hooded rats (N=15 for cohort 1 and N=16 for cohort 2), supplied by Envigo UK. Rats weighed approximately 270-300 g at the start of training, and approximately 400 g at the start of the pharmacological manipulations. They were housed in groups of two in cages with environmental enrichment, which consisted of a red 3 mm Perspex house (30 x 10 x 17 cm), a cardboard tube and a cotton rope suspended across the cage lid. Temperature was controlled within the range of 19-23 ^{o}C and humidity at 45-65%. Rat cages were placed in rooms with a twelve-hour reverse light cycle (lights off at 0800 hours) to ensure behavioural testing was carried out during their active period (between 0800 and 1800 hours). Water was available ad libitum in the home cage. During training and testing, rats were maintained at no less than 90% of their free-feeding body weight with restricted access to laboratory chow (LabDiet, PMI Nutrition International) to 18 g per rat per day. All procedures were carried out under local institutional guidelines (approved by the University of Bristol Animal Welfare and Ethical Review Board) and in accordance with the UK Animals (Scientific Procedures) Act 1986. During experiments all efforts were made to minimise suffering, and at the end of experiments rats were killed by giving an overdose of sodium pentobarbitone.

2.2.2 Hardware

Standard rat operant chambers (Med Associates, Sandown Scientific, UK) were used to conduct all experimental studies. The dimensions of the chambers were 30.5 x 24.1 x 21.0 cm. Each chamber was positioned inside a light-resistant and sound-attenuating box that contained a ventilation fan. All chambers included two retractable response levers, which were located on each side of the food magazine. Two lever lights (LL) were positioned above each lever (1" White Lens, 28 V). A rubber tube connected the food magazine to the pellet dispenser, while a house light (28 V, 100 mA) was located above the magazine. The pellets used as reward in all sessions were 45 mg grain-based sweetened food pellets (5TUL Test Diet, Sandown Scientific, UK). An audio generator (ANL-926, Med Associates, Sandown Scientific, UK) produced tones that were delivered to each chamber via a speaker placed at

the top of the chamber, above the left lever.

2.2.3 Software

The K-Limbic software (Conclusive Solutions Ltd., UK) operated as an interface between the operant chamber and a computer. It managed all input parameters to the chamber, such as dispensing pellets, extending and retracting levers and switching lights on and off. The software further controlled the audio generator. It received chamber output through lever presses and magazine entries, which it turned into a Microsoft Excel file (.xlsx format) after the end of each session. Custom code was developed in Julia (v1.6) (Bezanson et al. 2015) to read through the output forms, map the data into a common table format and produce all figures (`https://github.com/harisorgn/JBT`).

2.2.4 Task Design

Each rat of both cohorts of the present studies was trained to associate one of two reference cues with one of two available actions, which in turn led to one of two different reward amounts. In previous studies, to optimize the design of this version of the JBT, the magnitude of the the reward outcomes was chosen as one and four reward pellets (Hales 2016). The cues corresponding to the high and low reward magnitudes are presently referred to as HC and LC respectively. The available actions throughout all experiments were single presses on either one of the two levers located on either side of the magazine inside the operant chamber. The levers remained extended throughout each session. These actions are referred to as high action (HA) and low action (LA), to represent the correct action that leads to a high reward (HR) and a low reward (LR) value respectively. Both cue-action and action-outcome pairs were counterbalanced across all animals of both cohorts.

Figure 2.1 includes a graphical representation of the task. Successful graduation from all training stages ensured that rats were able to respond to each of the HC and LC stimuli with the appropriate action to receive an HR or an LR outcome. During the subsequent test sessions, the AC was presented for the first time. Given no prior association between the newly presented ambiguous cue and the available actions, the task's premise is that a rat would use its prior expectation of a small or large reward and take the action that corresponds to the expected reward amount.

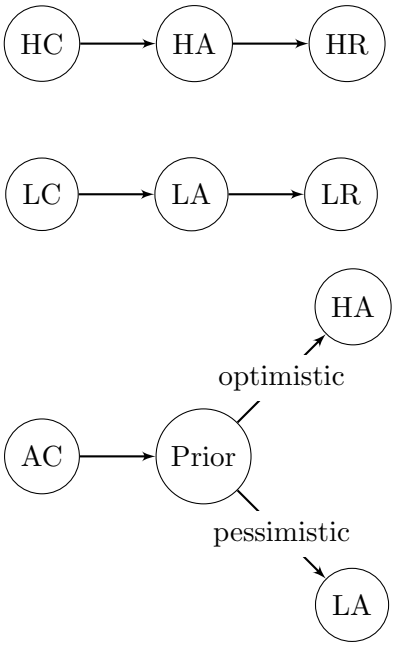


Figure 2.1: **Graphical representation of the basic premise of JBT**: The top two rows are sequences of cue-action (HC-HA, LC-LA) and action-reward (HA-HR, LA-LR) associations, being built through training. The ambiguous cue (AC) is presented during test sessions for the first time, thus it is not linked to a particular action. The premise of the task is that actions towards the previously high rewarded option, HA, reflect a more optimistic prior expectation and therefore a positive affective state, whereas choosing LA, which was previously led to a low reward value, would imply a pessimistic prior and a more negative state.

As depicted in the bottom sequence of Figure 2.1, an action taken towards the lever that had been previously associated with a high reward was deemed an optimistic action, reflecting a positive affective state. Conversely, an LA response, which had a trained association with the LR, would imply a more negative affective state.

2.2.5 Task parameters

Tables 2.1 and 2.2 summarise the task parameters, where there were differences between the task versions implemented for cohort 1 and cohort 2 respectively.

The magnitude values of all tones were based on a linear approximation of the audiogram of hooded rats (Heffner et al. 1994), while the audiogram curve in the range between 2 and 8 KHz was the closest to linear (Syka et al. 1996). The loudness adjustment was performed so that all tones were equally audible by the rats. Counterbalance between reference cues and reward amounts was not implemented, as it has been shown that rats were unable to be trained successfully when the 8 KHz

tone was paired with the high reward of four pellets (Hales 2016). The amplitude of all tones remained constant during cue presentation, after an initial rise period of 0.025 seconds.

In the novel JBT design for cohort 2, there was no need to carefully adjust the frequency of the AC to ensure that the AC was perceptually equidistant from the two reference tones. The AC was a visual stimulus (LL), which was a common component between the reference cues. The reference cues were compound stimuli, consisting of the lever lights switching on and a tone of constant frequency. The tone parameters were initially the same as for cohort 1. However the loudness of the LC changed from 64 to 76 dB for experiment 3. In order to make sure that rats were trained efficiently to discriminate between the reference cues, the tone started playing 0.5 seconds before the lever lights came on, during all stages of the task. Moreover, the number of ambiguous trials in a test session was reduced and the reinforcement schedule for correct responses on reference cues was probabilistic for this task.

	HC	AC	LC
Parameters	Tone (2 KHz, 75 dB)	Tone (5 KHz, 70 dB)	Tone (8 KHz, 64 dB)
N of trials in test session	40	40	40
Reinforcement schedule	100%	50%	100%

Table 2.1: **Task parameters for the JBT of cohort 1**

	HC	AC	LC
Parameters	Tone (2 KHz, 75 dB) +LL	LL	Tone (8 KHz, 64/76 dB) +LL
N of trials in test session	40	20	40
Reinforcement schedule	80%	50%	80%

Table 2.2: **Task parameters for the JBT of cohort 2**: LL: lever lights on

2.2.6 Trial structure

Figure 2.2 depicts a flowchart of the structure of a trial. This format was implemented exactly as shown for the Reward Magnitude Training stage and during all consecutive experimental studies. The earlier Discrimination Training stage was based on the same structure with the only difference being that the size of both rewarded outcomes, HR and LR, was set to one pellet. All trials were self-initiated

by the rat with a nose-poke into the magazine. An inter-trial interval (ITI) between the beginning of the trial and the presentation of a cue ensured that rats learned to hold their responses until a cue was presented to avoid a time-out due to a premature lever press. All cues were played for a maximum time of 20 seconds and were stopped if a response was made during this time. No response within the allotted time resulted in a time-out. Reference cues, HC and LC, were always rewarded after a correct response, while the AC was rewarded 50% of trials, with the reward amount being dependent on the action taken.

This flowchart was used for trials of both cohort 1 and 2. The only deviation from this trial structure is that for cohort 2 a correct action HA or LA, during a trial where HC or LC was presented respectively, led to a reward 80% of the time, with no reward the rest of the time. A graphical example depicting the rat in the operant chamber during an HC trial is shown in Figure 2.3.

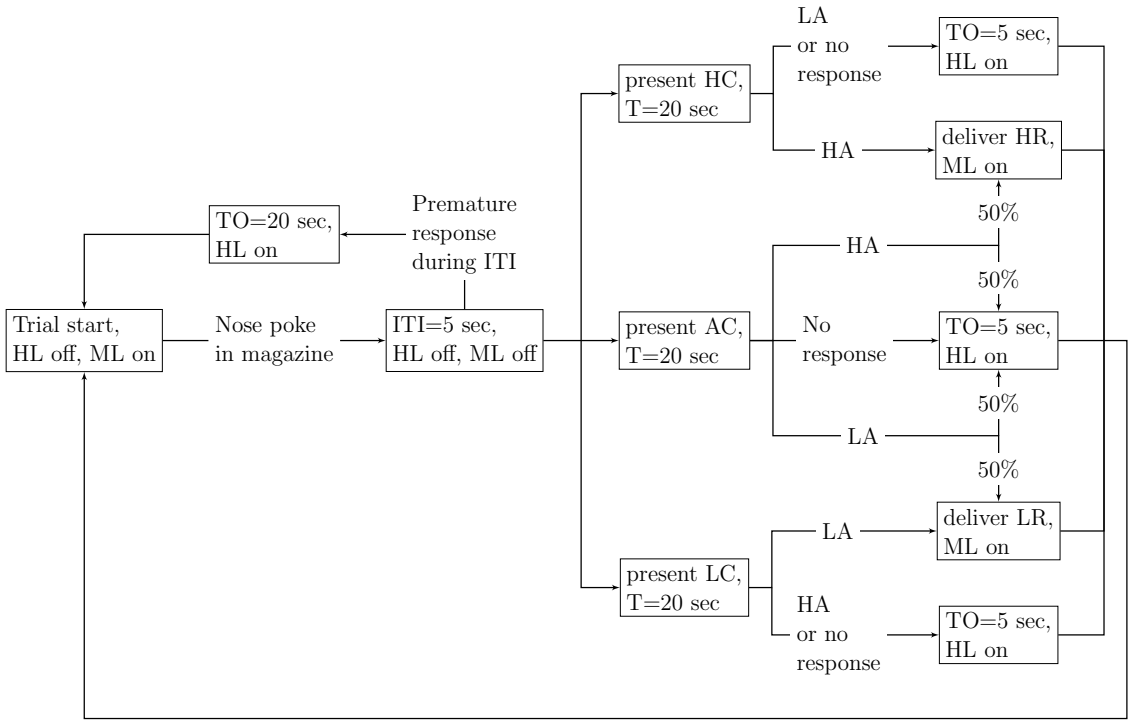


Figure 2.2: **Structure of a JBT trial**: Boxes represent states during a trial and arrows contain in unframed text the conditions for transitioning between consecutive states. Trials are self-initiated by a nose-poke into the magazine, followed by an ITI and then the presentation of a cue, chosen pseudo-randomly. Cues are presented for a maximum of 20 seconds and if no response if made within this time, the trial is registered as an omission. Responses during the AC lead to probabilistic outcomes 50%/50%, dependent on the type of response. HL : House Light, ML : Magazine Light, ITI : Inter-Trial Interval, TO : Time Out, T : Time of cue duration, HC : High Cue, AC : Ambiguous Cue, LC : Low Cue, HA : High Action, LA : Low Action, HR : High Reward, LR : Low Reward.

Figure 2.3: **Graphical depiction of a JBT trial**: A tone is presented at the beginning of a trial (left panel). The rat makes a choice by pressing one of the two levers (middle panel). The tone stops after an action is taken. The rat receives a reward, which is dependent on the presented tone and the pressed lever (right panel). In this example the tone was a high-reward cue (HC, 2 KHz) and the rat correctly chose the high-reward action (HA), resulting in the high reward (HR, 4 pellets).

2.2.7 Behavioural metrics

Several metrics were calculated from the raw output of each session. These quantities were extracted for each rat of both cohorts. The number of indices on each metric denotes the its number of dimensions, while the type of each index describes the granularity used for its calculation. First of all responses were measured as

$$\text{Response percentage}: R_{c,a} = \frac{N_{c,a}}{N_c} 100\% \tag{2.1}$$

where $N_{c,a}$ is the number of trials when cue C was presented and action A was taken, and N_c is the total number of trials where cue C was presented, regardless of the chosen action. In the following Results section, response percentages to all cues are reported only for the HA action, that is matched with the HC reference cue to deliver the HR reward. This way, accuracy for the HC can be directly monitored, along with inaccurate responses during LC, since HA was chosen incorrectly in these trials. Most of the LC trials when an HA response was not made, a correct LA one was, since the omissions were consistently low. Thus an estimate of LA accuracy could be gained indirectly.

When no response was made during a trial, that trial was registered as an omission in the output. The rate of omissions was measured for each cue as

$$\text{Omission percentage}: O_c = \frac{N_{c,\emptyset}}{N_c} 100\% \tag{2.2}$$

where $N_{c,\emptyset}$ represents trials of cue C where no action was taken within the cue presentation time.

Responses during the ITI, before cue presentation and after the nose-poke that initiates a trial, were registered as premature actions.

$$\text{Premature percentage}: \ P = \frac{N_p}{N} 100\% \tag{2.3}$$

with N_p being the total number of premature responses in a session and N being the total number of trials in a session, including the premature ones.

Response times measured the time in seconds between cue presentation and action taken in response as

$$\text{Response time}: \ \text{RT}_{a,c,i} = T_{a,c,i} - T_{c,i} \tag{2.4}$$

where $T_{a,c,i}$ is the time when action a was taken during trial i when cue c was played and $T_{c,i}$ denotes the start of cue c presentation in trial i. The response time of omission trials was set to the maximum cue presentation time of 20 seconds. In the current chapter, response times are reported as averages across all correct response times during reference cues, HC and LC, and for either response made during the AC for each animal in each session.

The four metrics above were control measures. The pharmacological manipulations used in our experiments were purported to only change the affective state and therefore the prior expectation for reward that drives actions during the AC (Figure 2.1). Thus, any changes to these four control measures would imply some off-target effect of the drug.

The Cognitive Bias Index (CBI) was the final metric. It measured the affective biases of rats during the AC trials of test sessions. This was the principal metric for JBT:the core hypothesis is that affective state influences actions under ambiguity and, therefore, change the lever preference under the AC. This is what the CBI measured and it was defined as

$$\text{CBI} = \frac{N_{AC,HA} - N_{AC,LA}}{N_{AC,HA} + N_{AC,LA}} \tag{2.5}$$

where $N_{AC,HA}$ and $N_{AC,LA}$ were the number of trials when AC was presented and the HA or LA action was made respectively. The CBI varied between -1 and 1, with more positive values representing a bias towards choosing the high reward action, HA, thus implying an optimistic prior expectation of reward and conversely for negative values.

2.2.8 Training stages: cohort 1

The original training stages of the JBT protocol were designed based on the phenomenon of Pavlovian-to-instrumental transfer (Cartoni et al. 2013). Pavlovian conditioning sessions of cue-outcome associations were followed by action-outcome training without cues. Cues, actions and outcomes were then combined in a third training stage. Here, an altered version of this protocol was tested in an effort to reduce the length of the training period. All training stages, the number of sessions spent on each one and the criteria for moving to the next stage are presented in Table 2.3. Training sessions took place 5 days per week, Monday to Friday.

Training Stage	Duration (Sessions)	Criteria
Magazine Training	2	• >20 pellets eaten
Tone Training	4	• >70% accuracy on two consecutive sessions
Discrimination Training	16	• >1:1 correct:premature responses ratio on two consecutive sessions
Reward Magnitude Training	10	• >60% accuracy on two consecutive sessions • 1:1 correct:premature responses ratio on two consecutive sessions

Table 2.3: **Training stages for cohort 1**: Training stages are shown along with the number of sessions it took for the cohort to graduate each stage and the corresponding graduation criteria. Experiments started after the end of Reward Magnitude Training.

Magazine training had the form of typical Pavlovian conditioning. It consisted of two sessions, during which one of the two reference cues was presented for 5 seconds, followed by a single pellet reward. Trials were not self-initiated at this stage. The ITI between cues was 30 seconds. Each session included presentations of only one of the two reference cues, while the same process was repeated for a second session with the other cue.

The next stage was Tone Training, where the goal was to teach the rats to respond to each reference cue with the correct action. Each session included repetitions of a single reference cue, while the session cues alternated for consecutive sessions. Only the correct lever for each cue was extended. Rats were rewarded with a single pellet if they pressed on the lever during the 20 seconds of cue presentation. This stage introduced self-initiation trials, an ITI between trial start and cue presentation, premature actions and time-out, as shown in the flowchart of Figure 2.2.

The Discrimination Training stage was the first one to include both cues within a session. Cue presentations followed a pseudo-randomised order and each cue was presented for 50 trials. Both levers were extended at this stage. Rats had to learn

to respond to each cue with the correct action and be more than 70% accurate on both cues in order to graduate from this stage.

After successful discrimination performance, the Reward Magnitude Training stage introduced the difference in reward values. The HC cue, followed by the correct HA action now led to four sugar pellets. All other parameters were identical to the Discrimination Training Stage. The accuracy criterion at this stage was reduced to 60% to allow for biases due to the differential reward outcomes.

Experiments with pharmacological manipulations were initiated after successful fulfillment of the Reward Magnitude Training criteria. The only experiment that was conducted before that point was a probe test session that introduced the ambiguous cue, immediately after the end of Discrimination Training. This session was carried out in order to asses perceptual biases when responding to the AC. Since the reward values were both set to one pellet, any deviation from choosing randomly between HA and LA during the AC would imply that this stimulus is not truly ambiguous and the rats interpret it as one of the reference cues before responding.

2.2.9 Experiment design: cohort 1

Experiment 1: 1vs1 perceptual test

This experiment included a single session, which introduced the AC for the first time. There were no differences in rewarded outcomes, with LA and HA resulting in a one-pellet reward 100% of the time during HC and LC trials and 50% of all AC trials.

Experiment 2: 4vs1 baseline test

In experiment 2, the baseline behaviour of cohort 1 was measured, when both the AC and the different reward values were present. This experiment took place over a week. On Monday and Thursday a session identical to Reward Magnitude Training was run. Tuesday and Friday included probe test sessions, the same as the Experiment 1 session, with the only exception being that there was now a difference in reward magnitude.

Overview of pharmacological experiments 3-6

Subsequent experimental studies involved pharmacological manipulation. The details of each experiment are included in Table 2.4. Each one of these experiment was a fully counterbalanced, double-blinded, within-subject study. All drugs were supplied by Sigma-Aldrich (UK). They were dissolved in 0.9% sterile saline vehicle. The same saline was used as a control treatment in every experiment. All substances, control and drugs, were administered through intraperitoneal (i.p.) injection, using non-restrained technique, shown to reduce reduce stress in animals (Stuart et al. 2015a). The week schedule during each experiment was identical to experiment 2, with pharmacological manipulations occurring on Tuesdays and Fridays, while Reward Magnitude Training sessions were run on Mondays and Thursdays.

	Experiment 3	Experiment 4	Experiment 5	Experiment 6
Drug	Ketamine	Ketamine	Amphetamine	Ketamine
Dose volume [mg/kg]	1.0	1.0	0.3	0.3, 1.0, 3.0
Pretreatment time [h]	1	24	0.25	1

Table 2.4: **Pharmacological manipulations for Experiments 3-6.**

2.2.10 Training stages: cohort 2

Table 2.5 summarises the training stages, their criteria and the time spent on each. one for the new JBT variant. The initial stage, Magazine Training, was the same as for cohort 1. Presentations of one reference cue for 5 seconds per trial were followed by a reward pellet, every 30 seconds.

Training Stage	Duration (Sessions)	Criteria
Magazine Training	2	• >20 pellets eaten
CRF Training	7	• >70% accuracy on two consecutive sessions
Discrimination Training	15	• >1:1 correct:premature responses ratio on two consecutive sessions
Experiments 1 & 2		
Discrimination Training (new contingencies)	18	• >70% accuracy on two consecutive sessions • >1:1 correct:premature responses ratio on two consecutive sessions
Experiment 3		

Table 2.5: **Training stages for cohort 2**: Training stages are shown along with the number of sessions it took for the cohort to graduate each stage and the corresponding graduation criteria. Experiments 1 and 2 took place after the end of the first round of Discrimination Training. Then, Discrimination Training was repeated with new cue-action contingencies before Experiment 3.

The following stage was CRF (Continuous Reinforcement) Training, where levers were extended and rats were rewarded after either lever press during the presentation of either one of the reference cues. This stage was implemented in order to train the rats to wait for the light component of the compound reference cues, before making a response. The tone component of each cue was initiated 0.5 seconds before the lever lights were switched on during all trials. Responses before the lever lights were on counted as premature.

The design of the third stage, Discrimination Training, was similar to that of cohort 1. Presentations of both reference cues were interleaved in a pseudo-randomised order. Only responses to the lever matched to the cue presented were considered correct. For the first seven sessions, correct responses were always rewarded, while after that point the probability of reward changed to 80% and remained constant throughout all future stages and experiments.

After achieving the criteria for successful discrimination, two experiments took place. The first one was a baseline 1vs1 experiment, in order to assess how the rats responded to the novel ambiguous cue. Consequently, the effect of an acute treatment with ketamine on the perception of this ambiguous cue was measured in a second experiment.

The results from the above experiments indicated potential sources of bias in the rats' responses, not caused by their affective state. Therefore, training was restarted, with alterations on the auditory component of the reference cues, to further investigate our observations. In order to restart the training process, cue-action contingencies were swapped for reference cues. Thus, rats had to learn anew to press a previously incorrect lever for both cues. After fulfilling the discrimination training criteria for the new contingencies, experiment 3 commenced.

2.2.11 Experiment design: cohort 2

Experiment 1: 1vs1 perceptual test

The first experiment of the compound JBT was to test the ambiguity of the chosen AC. Correct responses to the reference cues were rewarded at 80% of their trials, while the reward rate for ambiguous cues was set at 50%. All rewards had the same magnitude. Two test sessions were conducted as part of this experiment.

Experiment 2: Effect of an acute ketamine treatment on the perception of the ambiguous cue

This was a fully-counterbalanced, double-blinded, within-subject study with a single dose volume of ketamine at 1.0 mg/kg and a vehicle drug as control. All rewards had the same magnitude.

Experiment 3: 1vs1 perceptual test with altered reference cues

Experiment 3 was a repetition of experiment 1 after an adjustment to the loudness of one of the reference cues was made. After retraining the animals to discriminate the new cues, two test sessions took place.

2.2.12 Statistical Analysis

Cohort 1 studies were a direct replication of past experiments that had shown an effect of a drug treatment on CBI by performing one-sample t-tests on the difference in CBI between drug and vehicle treatments. Thus, the same statistical test was used to assess the replicability of the past published results (Hales et al. 2017) and to test for effect of novel drug treatments. The cohort 2 experiments were designed to evaluate whether a novel AC could induce greater ambiguity. One-sample t-tests were then used to calculate deviations of CBI values from the value of zero, which corresponded to total ambiguity under our hypothesis.

2.3 Results

2.3.1 Cohort 1

Experiment 1: 1vs1 perceptual test

The first experiment conducted in cohort 1 was a test of the ambiguous cue under equal reward values. The first test session of this experiment had to be excluded, since no animal met the criterion of having the ratio of correct responses over premature being more than 1. Moreover on the second session two animals were excluded

for the same reason.

The results for the included animals and sessions are shown in Figure 2.4. The CBI was negative for most animals, with the standard error of the cohort mean not overlapping 0. A t-test on the Cognitive Bias Index (Figure 2.4A) revealed an effect that was not statistically significant (one-sample t-test: $t(12) = -2.14, p = .054$). The HA response here was paired with the 2 KHz tone, thus negative values of CBI imply a bias towards the 8 KHz-paired lever (LA) when responding during the ambiguous 5 KHz tone. Accuracy values for both reference cues were around 80%, as judged by looking at the HA responses (Figure 2.4B) and the omission rates (Figure 2.4D) that were very low across cues. The response times (Figure 2.4C) were comparable across cues, although being more variable for the 5 and the 8 KHz tones. The premature responses (Figure 2.4E) were elevated relative to future test sessions, however they were within the criteria for 13 out of 15 animals in the second session.

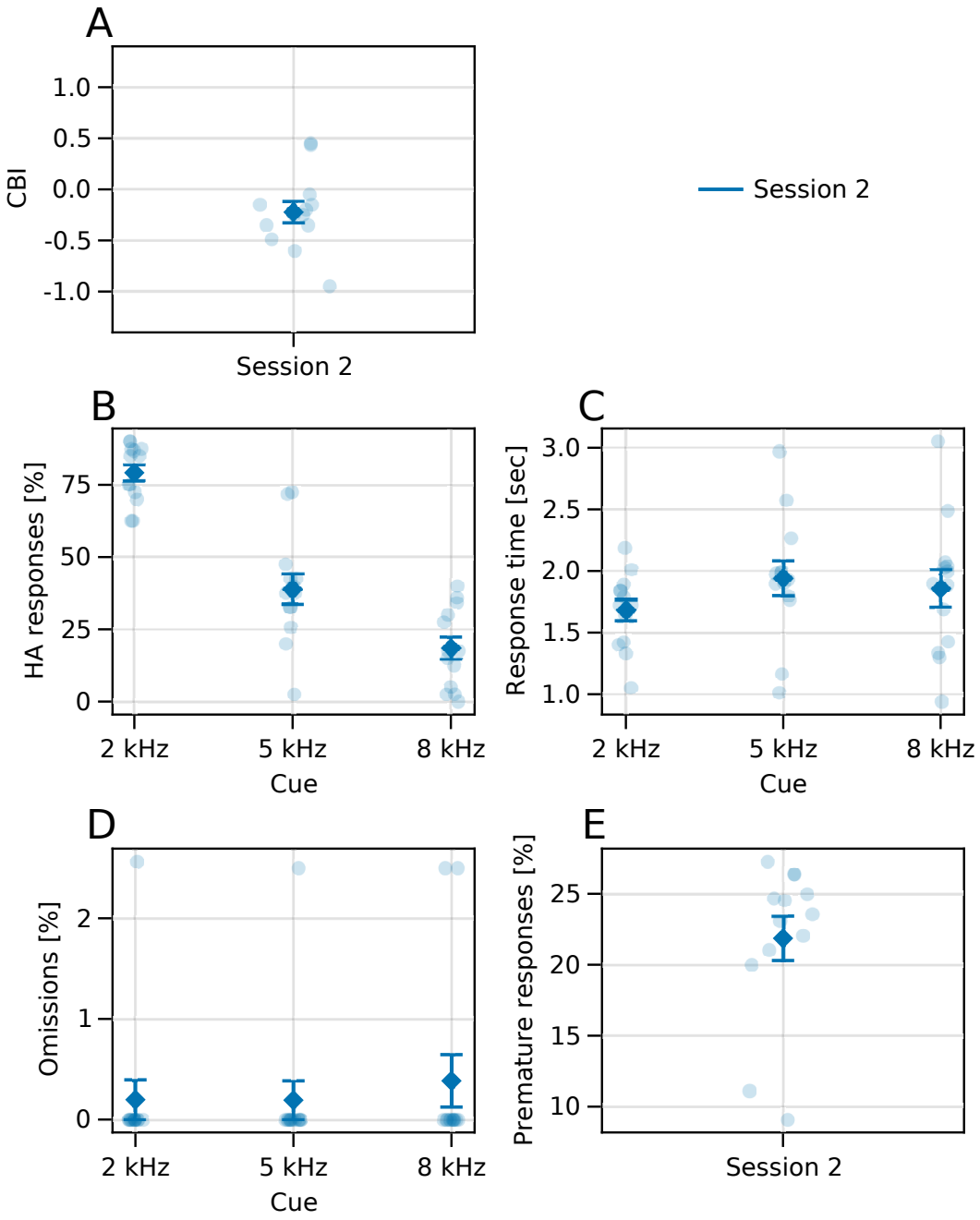


Figure 2.4: **Behavioural metrics for the 1vs1 perceptual test**: **A**: Cognitive Bias Index, **B**: percentage of trials when HA was chosen for each cue, **C**: response times for correct trials during 2 and 8 KHz and either responses during 5 KHz trials, **D**: omission percentages for each cue, **E**: premature response percentages during the session. Session 1 of the 1vs1 was excluded as all animals failed to meet the required criteria. Diamonds are cohort means, whiskers are SEM, circles are individual animals, N=15. HA : High Action, corresponds to the correct action for the 2 KHz cue

Experiment 2: 4vs1 baseline test

In experiment 2, I measured the baseline behaviour of cohort 1 under differential reward values. These were four pellets for HA responses at HC and one pellet for responding LA during LC trials. The CBI values of both sessions (Figure 2.5A) were centered around 0, though individual animals covered most part of the range of possible CBI values (first session one-sample t-test: $t(14) = .07, p = .941$, second session one-sample t-test: $t(14) = -0.63, p = .537$). The CBI values were stable across sessions (paired sample t-test: $t(14) = 1.09, p = .293$). The response times (Figure 2.5C) during correct responses on the HC were much faster than those during the LC, with minimal overlap between the two conditions. Response times during the AC, irrespective of the chosen action, were closer to those during the LC than responses during HC. Additionally, omission rates (Figure 2.5D) were higher for LC trials compared to HC ones. Omissions during AC trials shifted upwards from the first to the second session, with the sample statistics of the cohort overlapping those during the LC.

Experiment 3: Effect of an acute ketamine treament

The first drug study for cohort 1 was a single-dose, acute treatment with ketamine. Figure 2.6A depicts the difference in CBI between the ketamine session and the vehicle one. Most animals exhibited a positive shift in CBI under ketamine, however the drug's effect on the population mean was not significant (one-sample t-test: $t(14) = 1.9, p = .077$, Figure 2.6A). Ketamine did not disrupt the accuracy during the HC or the LC (Figure 2.6B). Inaccurate responses during the LC occurred at the same rate between conditions, even though the omission rate in these trials was decreased under ketamine (Figure 2.6D). Response times for correct responses during the HC were consistently the fastest, however correct responses during LC trials and particularly AC trial responses exhibited a downwards shift under ketamine (Figure 2.6C). Finally, the ketamine treatment did not alter premature response rates (Figure 2.6E).

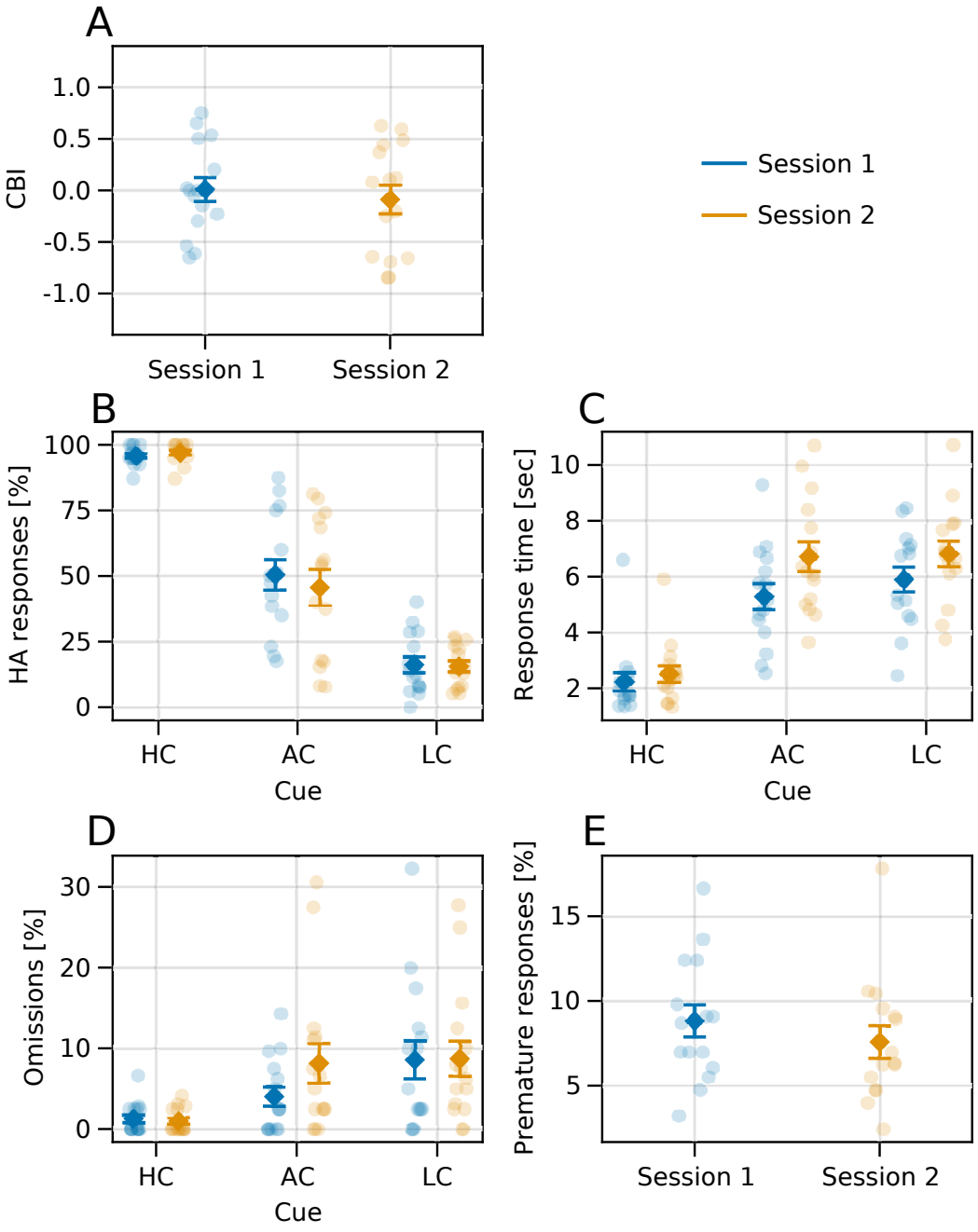


Figure 2.5: **Behavioural metrics for the 4vs1 test sessions**: **A**: Cognitive Bias Index, **B**: percentage of trials when HA was chosen for each cue, **C**: response times for correct trials during 2 and 8 KHz and either responses during 5 KHz trials, **D**: omission percentages for each cue, **E**: premature response percentages during the session. Diamonds are cohort means, whiskers are SEM, circles are individual animals, N=15. HA : High Action, HC : High Cue, AC: Ambiguous Cue, LC : Low Cue

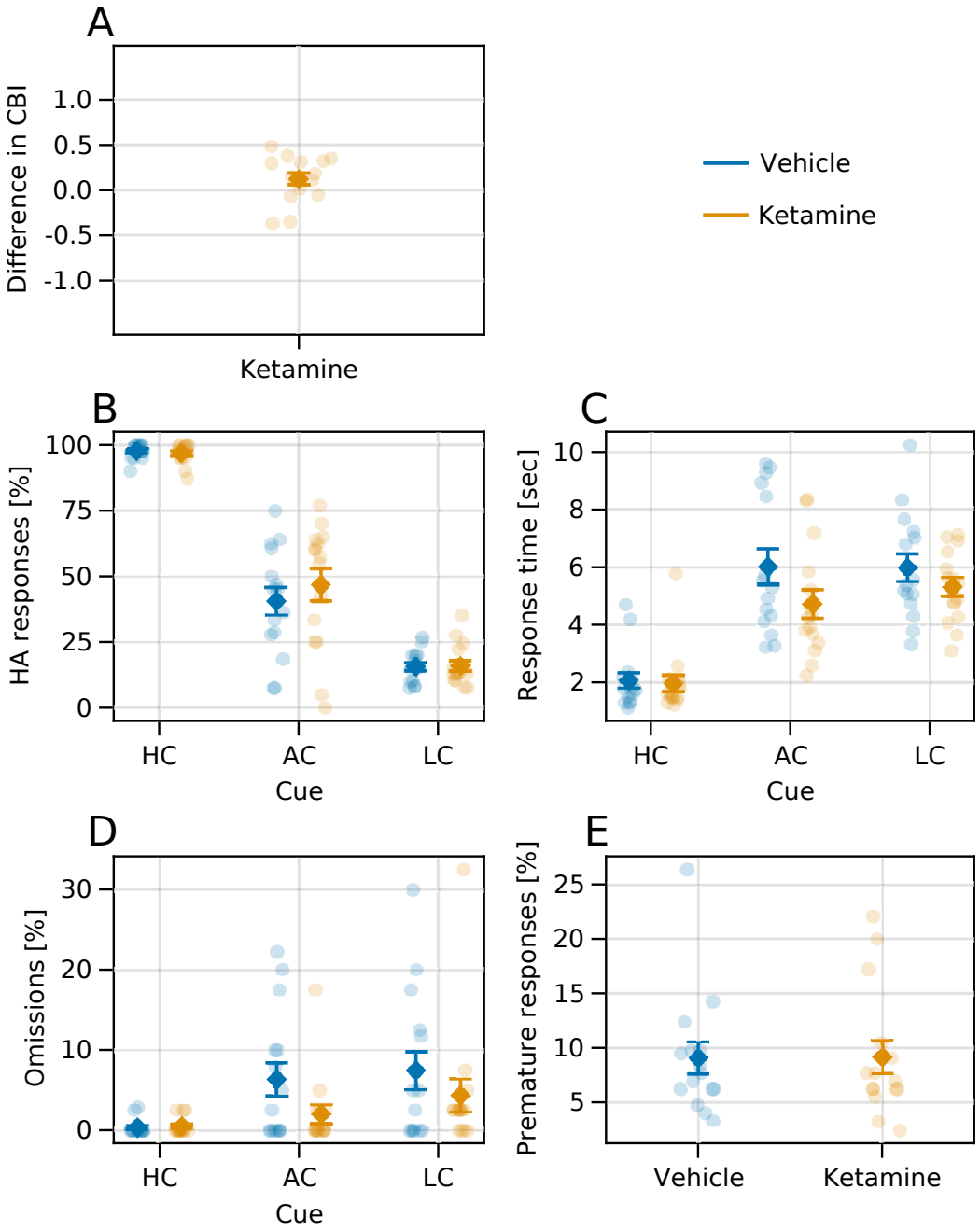


Figure 2.6: **Behavioural metrics for the acute ketamine treatment sessions**: **A**: Difference in Cognitive Bias Index between ketamine and a vehicle drug, **B**: percentage of trials when HA was chosen for each cue, **C**: response times for correct trials during 2 and 8 KHz and either responses during 5 KHz trials, **D**: omission percentages for each cue, **E**: premature response percentages during the session. Ketamine dose was 1.0 mg/kg, delivered via i.p. injection, 1 hour before the test sessions. Diamonds are cohort means, whiskers are SEM, circles are individual animals, N=15. HA : High Action, HC : High Cue, AC: Ambiguous Cue, LC : Low Cue.

Experiment 4: Delayed effect of an acute ketamine treament

Experiment 4 involved the same drug treatment as experiment 3, a single dose of ketamine. In experiment 4 though, the pre-treatment time was 24 hours before each test session to determine if there were any sustained effects of ketamine. There was no significant effect of the ketamine treatment on CBI (one-sample t-test: $t(14) = 1.67, p = .115$, Figure 2.7A). No changes were observed on the rest of the behavioural metrics, indicating no off-target effect of the drug.

Experiment 5: Effect of an acute amphetamine treatment

Experiment 5 aimed at measuring the effect of an acute treatment with a single dose of amphetamine on actions during the AC. There was a significant positive difference in CBI between amphetamine and vehicle (one-sample t-test: $t(14) = 2.82, p = .013$, Figure 2.8A). Amphetamine additionally caused an increase in incorrect responses during LC (Figure 2.8B) and an elevated premature response rate (Figure 2.8E). There were no differences in the rest of the measures, except for a small decrease in response times during LC trials under amphetamine (Figure 2.8C).

Experiment 6: Dose-response ketamine study

The final experiment for cohort 1 was a dose-response study with three dose volumes of ketamine. No dose caused a significant shift in CBI (Figure 2.9), relative to the vehicle treatment (one-sample t-tests: 0.3 mg/kg $t(14) = 0.18, p = .857$, 1.0 mg/kg $t(14) = 0.63, p = .541$ and 3.0 mg/kg $t(14) = 1.13, p = .276$). Similarly, no ketamine dose caused changes to the rest of the behavioural metrics.

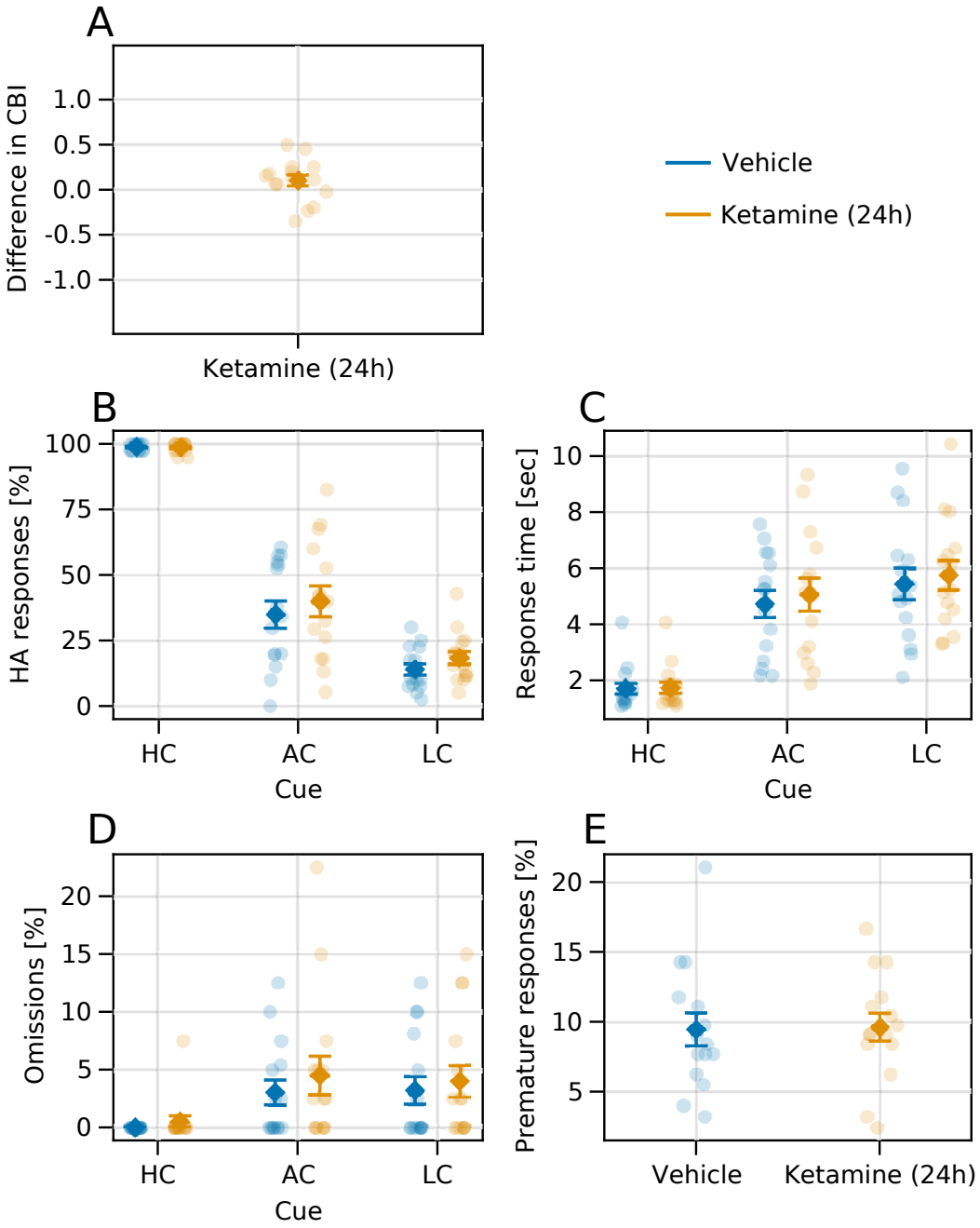


Figure 2.7: **Behavioural metrics for the acute ketamine treatment, 24 hours before test sessions**: **A**: Difference in CBI between ketamine and a vehicle drug, **B**: percentage of trials when HA was chosen for each cue, **C**: response times for correct trials during 2 and 8 KHz and either responses during 5 KHz trials, **D**: omission percentages for each cue, **E**: premature response percentages during the session. Ketamine dose was 1.0 mg/kg, delivered via i.p. injection, 24 hours before the test sessions. Diamonds are cohort means, whiskers are SEM, circles are individual animals, N=15.HA : High Action, HC : High Cue, AC: Ambiguous Cue, LC : Low Cue.

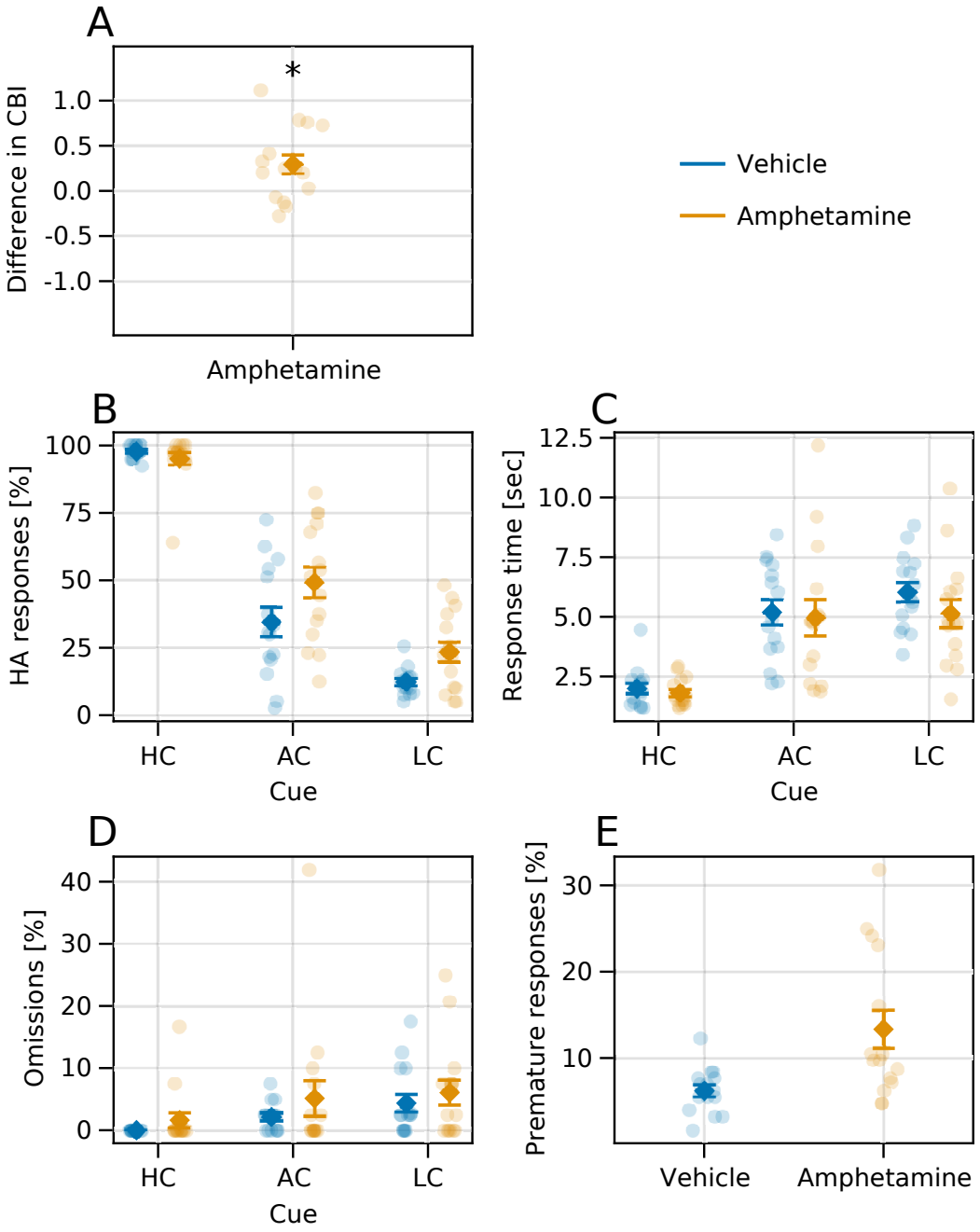


Figure 2.8: **Behavioural metrics for the acute amphetamine treatment**: **A**: Difference in Cognitive Bias Index between amphetamine and a vehicle drug, **B**: percentage of trials when HA was chosen for each cue, **C**: response times for correct trials during 2 and 8 KHz and either responses during 5 KHz trials, **D**: omission percentages for each cue, **E**: premature response percentages during the session. Amphetamine dose was 0.3 mg/kg, delivered via i.p. injection, 15 minutes before the test sessions. Diamonds are cohort means, whiskers are SEM, circles are individual animals, N=15. $^*p < .05$. HA : High Action, HC : High Cue, AC: Ambiguous Cue, LC : Low Cue.

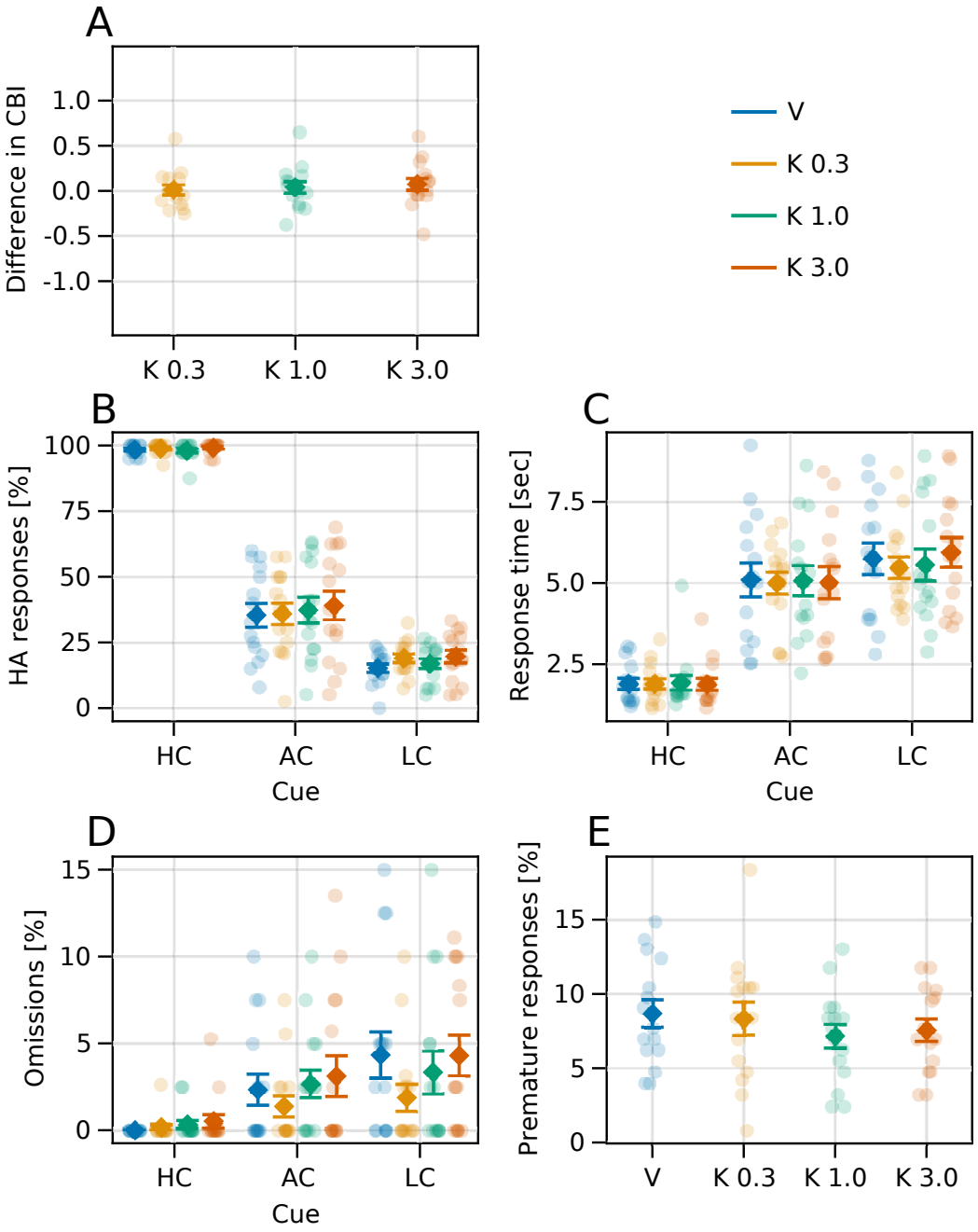


Figure 2.9: **Behavioural metrics during the ketamine dose-response study**: **A**: Difference in Cognitive Bias Index between each ketamine dose volume and a vehicle drug, **B**: percentage of trials when HA was chosen for each cue, **C**: response times for correct trials during 2 and 8 KHz and either responses during 5 KHz trials, **D**: omission percentages for each cue, **E**: premature response percentages during the session. HA : High Action, HC : High Cue, AC: Ambiguous Cue, LC : Low Cue, K : Ketamine. Ketamine doses (0.3, 1.0, 3.0 mg/kg) were delivered via i.p. injection, 60 minutes before the test sessions. Diamonds are cohort means, whiskers are SEM, circles are individual animals, N=15.

2.3.2 Cohort 2

Experiment 1: 1vs1 perceptual test

Both reward value after HA and LA were set to 1 pellet for this test, in order to investigate potential perceptual biases around the novel ambiguous cue. The CBI values during both sessions of this experiment were significantly negative (one-sample t-tests: $t(15) = -14.48, p < .001$, Figure 2.10A). Following the convention that HA action is paired with the compound cue that includes the 2 KHz tone component, negative CBI values indicate a bias towards the LA action, which is paired with the 8 KHz + LL cue. Most animals appeared faster at responding correctly to the 2 KHz + LL cue, compared to the other reference cue (Figure 2.10C).

Experiment 2: Effect of an acute ketamine treatment on the perception of the ambiguous cue

No shift in CBI (Figure 2.11A) was detected under ketamine (one-sample t-test: $t(15) = 1.13, p = .274$). Moreover, there were no changes to the other four behavioural metric, as shown in Figure 2.11.

Experiment 3: 1vs1 perceptual test after reference cue changes

The final experiment with cohort 2 involved another perceptual test with equal reward values, akin to the design of experiment 1. The difference between the two experiments was that the loudness of the 8 KHz tone was changed to 75 dB from its previous value of 64 dB. This adjustments led to significantly positive CBI values in both sessions, as shown in Figure 2.12A (one-sample t-tests: session 1 $t(15) = 23.86, p < .001$ and session 2 $t(15) = 27, p < .001$). This observation reflected a bias during the ambiguous cue trials towards the HA response, which was paired with the 2 KHz + LL cue throughout training. Interestingly, the rats appeared to choose the HA lever press more frequently during ambiguous trials, compared to trials of the 2 KHz + LL reference cue (Figure 2.12B). There was a small decrease in the sample mean of correct response times during the 8 KHz + LL cue, compared to the other reference cue, as depicted in Figure 2.12C.

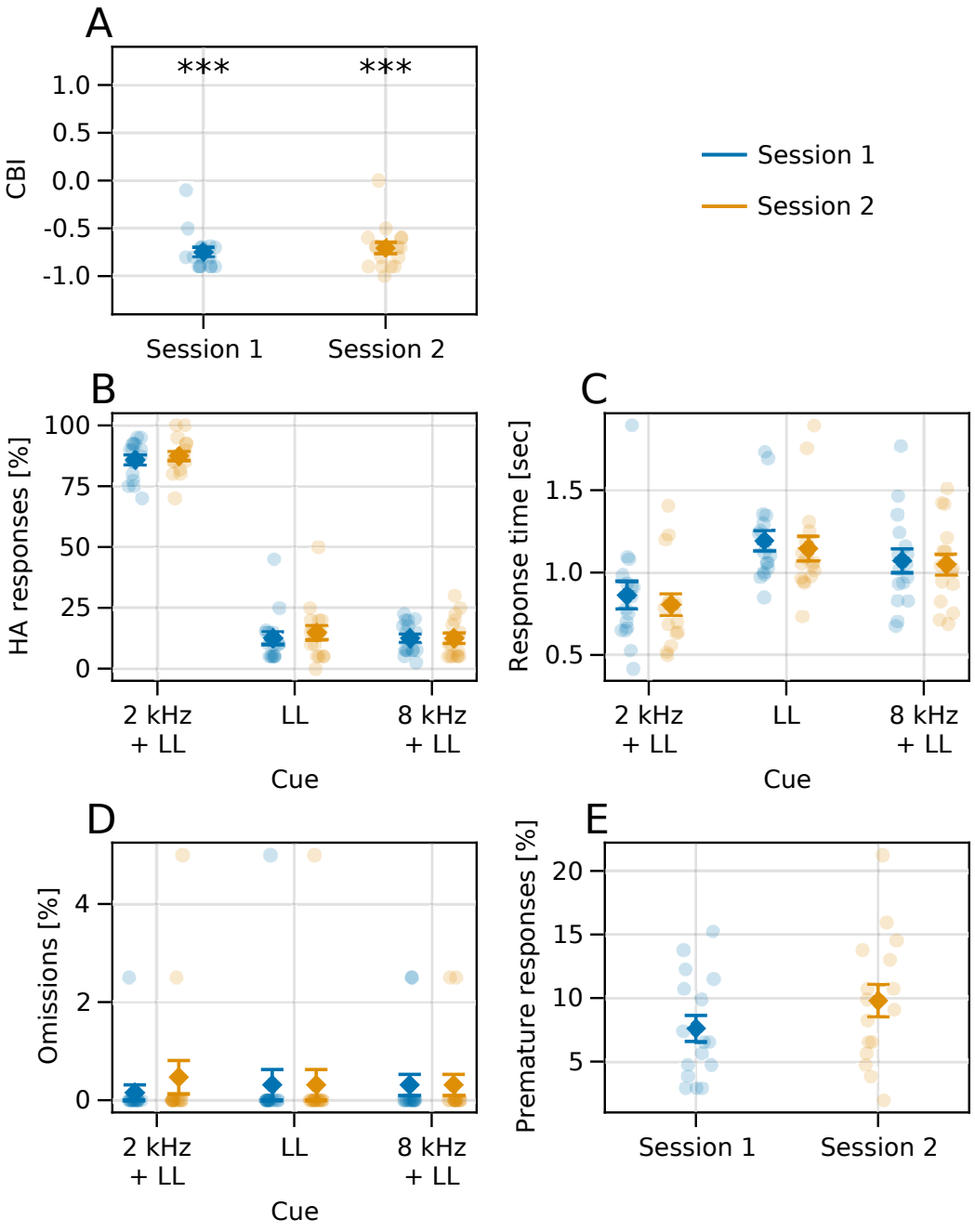


Figure 2.10: **Behavioural metrics for the 1vs1 perceptual test**: **A**: Cognitive Bias Index, **B**: percentage of trials when HA was chosen for each cue, **C**: response times for correct responses during reference cues and either response during the AC, **D**: omission percentages for each cue, **E**: premature response percentages during the session. Diamonds are cohort means, whiskers are SEM, circles are individual animals, N=16. $^{***}p < .001$. LL : lever lights switched on, HA : High Action, corresponds to the correct action for the 2 KHz + LL cue

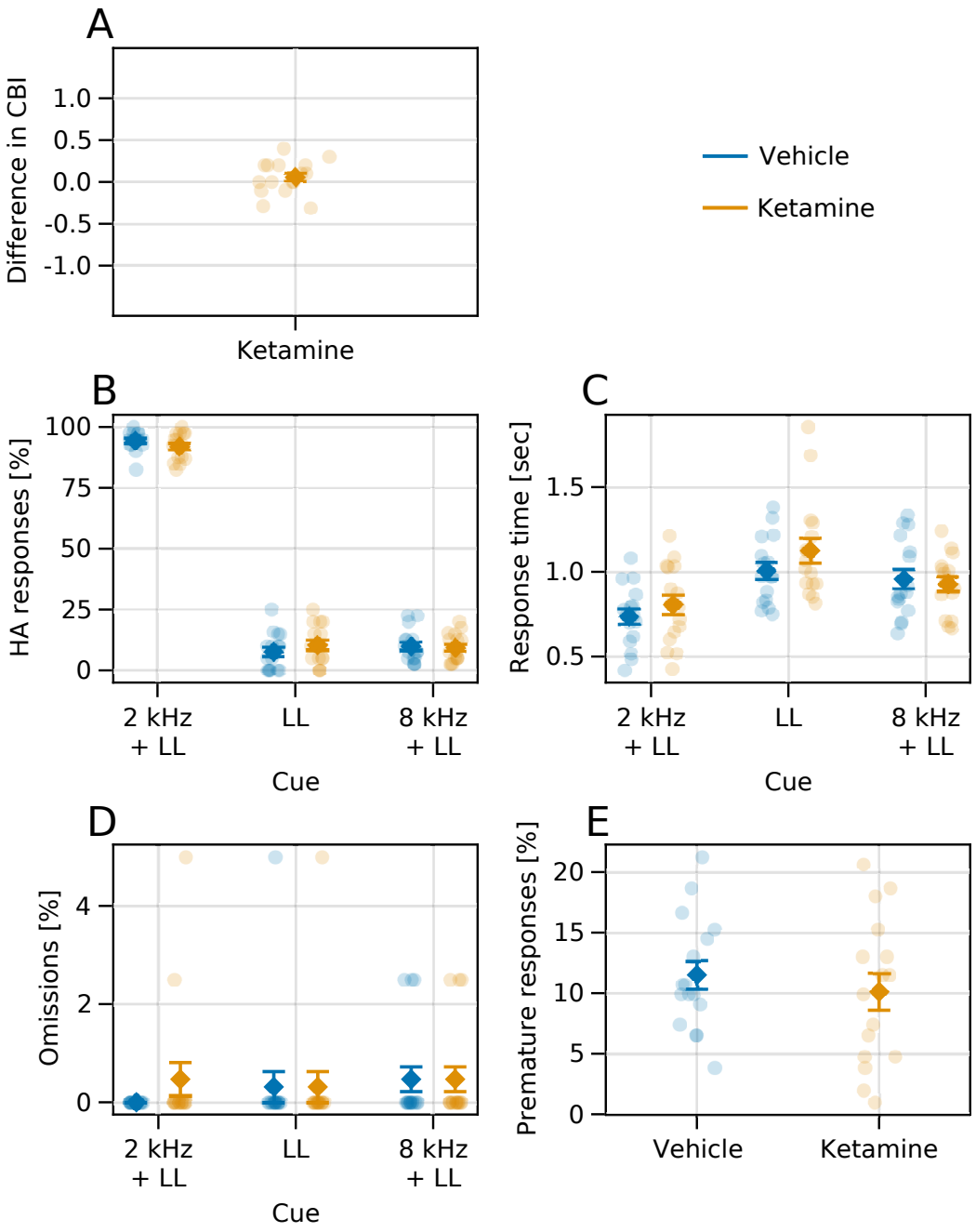


Figure 2.11: **Behavioural metrics for the acute ketamine treatment on 1vs1 sessions**: **A**: Difference in Cognitive Bias Index between ketamine and a vehicle drug,**B**: percentage of trials when HA was chosen for each cue, **C**: response times for correct responses during reference cues and either response during the AC, **D**: omission percentages for each cue, **E**: premature response percentages during the session. Ketamine dose was 1.0 mg/kg, delivered via i.p. injection, 1 hour before the test sessions. Diamonds are cohort means, whiskers are SEM, circles are individual animals, N=16. LL : lever lights switched on, HA : High Action, corresponds to the correct action for the 2 KHz + LL cue

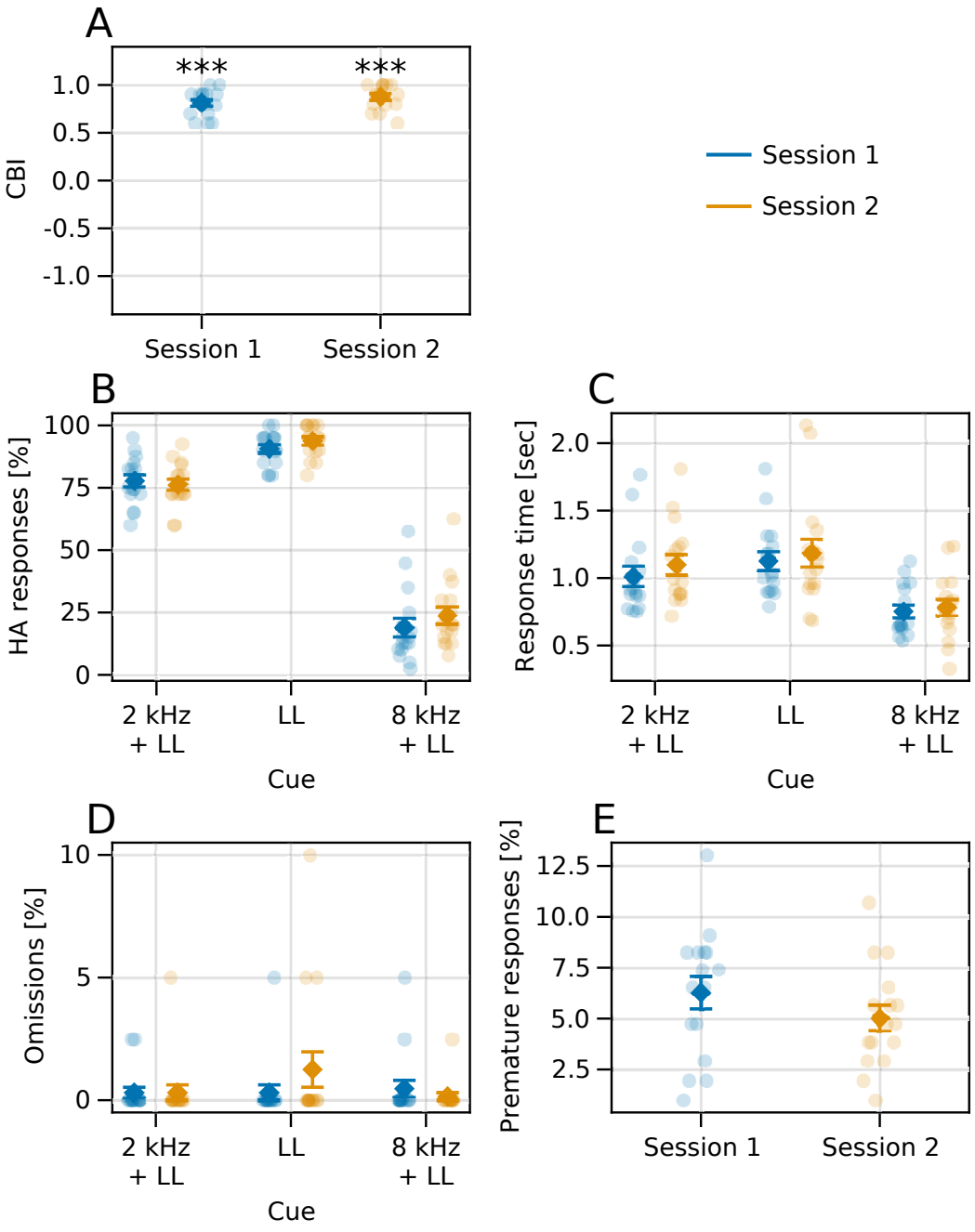


Figure 2.12: **Behavioural metrics for the 1vs1 perceptual test with altered tone loudness**: **A**: Cognitive Bias Index, **B**: percentage of trials when HA was chosen for each cue, **C**: response times for correct responses during reference cues and either response during the AC, **D**: omission percentages for each cue, **E**: premature response percentages during the session. Diamonds are cohort means, whiskers are SEM, circles are individual animals, N=16. $^{***}p < 0.001$. LL : lever lights switched on, HA : High Action, corresponds to the correct action for the 2 KHz + LL cue

2.4 Discussion

The present chapter aimed at expanding upon the JBT in rats, which involved exclusively rewarded outcomes and measured interpretation biases, driven by differential expectations of reward. In cohort 1, replication attempts of positive findings with ketamine were conducted, along with an effort to investigate the delayed effects of the drug for the first time in JBT. Subsequently, cohort 2 was part of a novel JBT variant, designed to address several confounding factors of the original task.

The direct replication attempt of a positive shift in interpretation bias under ketamine (Hales et al. 2017) was not successful. A second study that repeated the ketamine dose under investigation (1.0 mg/kg), along with two other ones (0.3 and 3.0 mg/kg), also did not show an effect for any of them, suggesting that the issue was not a different sensitivity to the drug between cohort 1 and the rats of the original ketamine studies. Moreover, none of the secondary behavioural metrics were disrupted under ketamine (Figures 2.6 and 2.9), thus confirming that the drug did not have any off-target effect. Furthermore, the baseline behaviour of the animals in cohort 1 was stable across sessions (Figure 2.5), and similar to the animals' behaviour in the original ketamine studies. The CBI was centered around zero, while the omissions and premature actions were below the exclusion criteria.

Two more direct replication attempts of the effect of ketamine, within the same facility as cohort 1, by the authors of the published study, were unsuccessful as well (unpublished data). The only known difference between the published work and these subsequent replication attempts was the animal supplier. This factor is a subtle one but could have undocumented effects on the animals' behaviour. Both genetic and behavioural differences have been reported in inbred mice across breeding facilities (Bryant 2011; Jaric et al. 2022). Moreover, outbred rats, which originated in different facilities, have exhibited differences in risk-taking and exploratory behaviour (Palm et al. 2011). A factor that could vary across breeding facilities is the extent at which animals are handled by people. Rats, which are not handled regularly during their first month, exhibit elevated anxiety-like behaviour in multiple tests, such as novelty-suppressed feeding and open-field exploration (Caldji et al. 2000).

The present study considered low doses of ketamine, in the range of 0.3-3 mg/kg, as previous work has shown positive shifts in interpretation bias without any behavioural side-effects at this range (Hales 2016). Ketamine and its main metabolite,

norketamine, are non-competitive NMDA receptor antagonists (Ebert et al. 1997), particularly in the medial prefrontal cortex (mPFC) and dorsal hippocampus (HPC) of rats in the considered dose range (Mion et al. 2013). The concentration of the drug in plasma and brain tissue, extracted from mPFC and HPC, increases rapidly within a few minutes of systemic administration of an 2.5 mg/kg dose, but diminishes within the first hour (Saland et al. 2018). The concentration of norketamine, however, remains more elevated within the same time window, and this effect is particularly pronounced in female rats. A similar systemic dose of 5 mg/kg in rats has been shown to disrupt communication between the two brain sites, by causing synaptic depression of the HPC-mPFC pathway, an effect that is likely mediated by the availability of dopamine D_1 receptors in mPFC (Kamiyama et al. 2011).

Amphetamine previously caused the largest positive effect on CBI (Hales 2016). This effect was replicated successfully, in cohort 1, as was the increase in premature responses (Figure 2.8), which is thought to be an outcome of the drug's stimulation of motor activity in general. Another JBT study with amphetamine observed a significant effect only at a higher dose (1.0 mg/kg, compared to our 0.3 mg/kg) (Rygula et al. 2014). However, the task design of that study involved a reward and a punishment as the potential outcomes, not rewards of different amount. Amphetamine is known to increase the release of monoamine neurotransmitters, primarily targeting dopamine and noradrenaline (Heal et al. 2013). The effects of dopamine and noradrenaline on responding to an ambiguous stimulus could be different, when the potential outcomes are rewards or when a punishment is also probable.

The failed replication attempts of ketamine motivated a closer look into the task design. One design factor that could be a caveat is the choice of the AC parameters, when the AC is on the same scale as the reference cues. In cohort 1, the AC was thought to be truly ambiguous due to a combination of tone frequency and loudness. However, the results of experiment 1, which was designed to assess the perception of this AC without different reward amounts, were inconclusive as one session had to excluded, due to an elevated number of premature responses (Figure 2.4). The included data also revealed an increased number of premature actions, relative to past and future sessions. Interestingly, all animals did fulfill the same criterion of having more correct responses than premature ones in the Discrimination Training stage, exactly before moving on to the test week for this experiment. This disparity in behaviour could only be attributed to the presence of the ambiguous cue.

Having ambiguous cue trials during the two test sessions of experiment 1 decreased

the number of rewards earned, since the presentations of reference cues, which were 100% rewarded after correct responses, were diluted by the presence of 50% rewarded ambiguous trials. Responses on reference cues were more than 70% accurate, given the criteria of Discrimination Training (2.3). Thus, rats were used to a higher reward rate during training. Consequently, during the test session, the lower reward rate due to unrewarded ambiguous trials could have caused an agitation that led to the increase in the premature percentage. The observations of more vigorous actions due to higher food deprivations have been reported across a range of species (Hull 1943). In Pavlovian conditioning, an increase in response vigor was observed when reward probability was decreased, thus reducing reward rate (Anselme 2015). The authors suggested that this effect arose because of an elevated appetite. Here I similarly assume that a decrease from the accustomed reward rate led to a higher appetite and consequently higher vigor, which overrode the animal's inhibitory control to withhold responding before the ITI ends, as per the task rules.

In addition to a possible effect of reward rate, experiments 1 and 2 of cohort 1 imply that there was a potential perceptual bias driving the rats' actions. Experiment 2 differed from experiment 1 only in the reward amounts after correct actions to the reference cues. The CBI values of experiment 2 (2.5A), when the outcomes were four or one pellets, did not change with respect to the CBI measured in the perceptual test of experiment 1 (Figure 2.4A), when both rewards were equal to one pellet. According to the hypothesised process behind AC responses, as described in Figure 2.1, this observation implies that rats had no prior expectation of any particular reward amount. However, this is unlikely to be the only interpretation of the data, since rats of the same strain, under the same food restriction protocol, were willing to work for a similar amount of reward pellets, even when multiple lever presses led to a single pellet reward (Griesius et al. 2020). It could be possible that the rats valued the four pellets more than the one, yet they were primarily driven by a perceptual interpretation of the AC, rather than one reflecting their reward expectations.

The novel JBT design of cohort 2 primarily aimed at addressing the potential perceptual bias of the original task design. However, the goal of removing the AC from the same scale as the one containing the reference cues, which is the frequency scale, was not successful in this initial attempt. The animals of cohort 2 exhibited a significant bias in experiment 1, with smaller inter-individual variance compared to the original task (Figure 2.10A). The sign of this extreme CBI value was reversed in experiment 3 (Figure 2.12), when the tone loudness of one of the cues was changed.

Rats are able to distinguish between tones within the range of loudness values that I used between experiments 1 and 3 (Akrami et al. 2018). Thus, it is likely that the rats were matching the AC to the reference cues using the tone loudness as the common scale. The AC contained no explicit tone, so it is probable that the cue was interpreted as the reference cue that was least perceptible to the animals.

Ketamine did not shift the animals' CBI under equal reward values in cohort 2 (Figure 2.11A). Nevertheless, the apparent perceptual bias during experiments 1 and 3, due to the potential use of tone loudness as the scale of reference, is a limiting factor when interpreting the effect of ketamine. The animals did not appear to treat the AC as an ambiguous stimulus, thus the experiment could not provide any insight in terms of the influence of ketamine on processing uncertain sensory information.

Overall, both JBT versions have provided evidence in favour of the existence of a perceptual bias, due to the AC not being truly ambiguous. Nevertheless, any interpretation of the animals' actions as being driven by perception is limited. This limitation stems from the increased number of premature actions and the high exclusion rate in the perceptual test of cohort 1 and the extreme CBI values of cohort 2 that rendered further experiments impossible.

Future JBT studies could reassess the possibility of perceptual interpretations of the AC by introducing probabilistic rewards to lower the reward rate that the animals are trained on. Probabilistic rewards could consequently lower the number of premature responses during the perceptual test with equal rewards. Moreover, using multiple ambiguous stimuli could further inform the inference on potential perceptual and reward-driven effects. Psychometric curves of the animals' actions during each cue could then be better fitted to the data (Carandini et al. 2013).

One hypothesis could be that intermediate tones closer to the reference ones are more likely to be perceptually matched to the closest reference cue, whereas tones that are equidistant from the reference cues are more ambiguous and thus more likely to lead to actions driven by reward expectations. A meta-analysis of JBT studies has shown that responses to intermediate cues are differentially modulated, depending on their distance to the reference cues (Neville et al. 2020a). Multiple tones of intermediate frequencies have been used in JBT studies with a reward and a punishment as the potential outcomes (Harding et al. 2004; Jones et al. 2018; Neville et al. 2020b). However, these studies only compared actions during the intermediate tones under different outcomes. In our JBT variant with exclusively rewarded outcomes, a perceptual test with multiple cues and equal outcomes could precede a

baseline test with different reward amounts, in order to dissect any perceptual and reward-driven effects on actions.

Future JBT studies could additionally benefit from testing for a potential effect of ketamine on the perceptual processing of the AC. A human study has shown that a comparable low dose of ketamine (0.4 mg/kg) caused a disruption in the integration of sensory information in a perceptual decision-making task without different reward values for each one of the two available actions (Salvador et al. 2022). Particularly in the original JBT task, where ketamine has previously shown promising results of translational value, it would be beneficial to investigate perceptual biases as a potential confounding factor. The use of multiple intermediate cues could further enhance the inference capabilities of perceptual effects.

Besides a new definition of ambiguity as missing information, the novel JBT design of cohort 2 included probabilistic rewards and fewer AC trials. These design choices were made to address potential learning effects and effects of the average reward rate, as animals were used to more frequent rewards and less instances of negative feedback in the training sessions than in the test sessions of the original JBT.

Given the extreme bias in most animals' actions in the novel JBT, the effect of these changes in the task design could not be assessed. However, chapter 4 will address the effects of past reinforcement and the negative feedback due to the 50% rewarded AC in the original JBT using a model-based approach. A generative model of the animals' actions will be embedded within a hierarchical statistical model, which will further account for any potential differences between rats originating in different breeding facilities. Thus, a large-scale analysis, including animals from both the original successful ketamine studies and the replication attempts, will be conducted to investigate the animals' behaviour in more detail.

Chapter 3

Conditioned suppression task: reward-seeking under threat

3.1 Introduction

Animals act to fulfill their needs for primary reinforcers, such as food and water, and to avoid aversive experiences, such as pain. Studies on the Affective Bias Test (ABT) in rats have provided evidence that the animal's affective state, during learning of an action-reward association, influences the memory of the learned reward and consequently leads to biases either towards or away from taking the same action, depending on the valence of the affective state (Stuart et al. 2013; Stuart et al. 2015b; Stuart et al. 2019; Hinchcliffe et al. 2022). However, less is known about how affective state during an aversive event might influence the animal's memory of it and bias future actions.

Fear conditioning studies have provided evidence for exacerbated responses to unconditioned aversive stimuli after the induction of a negative affective state during training, by serotonin receptor agonist drugs (Bauer 2015). Conversely, anxiolytic, serotonin receptor antagonists have resulted in lower levels of fear expression in the same protocol. These fear conditioning studies involve Pavlovian associations of a conditioned stimulus with an aversive event and require no action from the animal. The main behavioural metric is the amount of time an animal freezes during conditioned stimuli presentations. There has not been a reliable way to measure freezing as of now, as measurements by manual and automated methods do not agree to a satisfactory degree (Anagnostaras et al. 2010). Moreover, the elicitation of freez-

ing requires inescapable electric shocks of high intensity, which compromises animal welfare and does not enable extended studies with chronic treatments.

The present chapter involves the design of a novel behavioural task, with the aim of measuring affective biases on actions that are associated with the avoidance of an aversive outcome. This design was based on the conditioned suppression task (CST) (Ayres 1968). In CST, animals are trained to associate a conditioned stimulus (CS) with an aversive outcome. During subsequent test sessions, the animals are free to act in order to receive a reward and they suppress this reward-seeking behaviour when the CS is presented. The task has been used in the past as an animal model of fear (Kamin et al. 1963) and multiple studies introduced pharmacological manipulations (Davis 1968 for a review) or altered Pavlovian conditioning protocols (Thomas et al. 2005) to mitigate the suppression of the reward-seeking behaviour.

The aversive outcome typically used in published CST studies has been an electric shock, either constant or in a pulse pattern, of intensities more than 0.5 mA (Davis et al. 1979). Moreover the shock was unavoidable across all studies. This design is more relevant for animal models of post-traumatic stress disorder (PTSD) (Bali et al. 2015). However, we aimed at investigating how affective state influences memories of aversive events, that are not traumatic. Thus, we introduced an escape platform that was always available to the animals and considerably reduced the shock intensity to the range of 0.2-0.3 mA. Additionally, these choices were made to improve animal welfare for our task.

Maintaining a stable baseline suppression behaviour is essential for the design of studies akin to the ABT protocol. The effect of affective state manipulations on the memory of an aversive stimulus could then be measured by comparing the consequent suppression behaviour with the baseline. Thus, making the aversiveness of the shock stimulation smaller would enable longer within-subject studies without compromising animal welfare. This way, prolonged or chronic effects of affective state manipulations could also be evaluated. This is unlike published conditioned suppression experiments, where animals typically took part in only one or two week-long studies with inescapable shocks of higher intensity.

Making the aversive stimulus avoidable could induce a conflict in the rat between staying to press for reward and potentially experience the electric shock or flee. Other tasks, such as the Geller-Seifter (Geller et al. 1960) and the Vogel test (Vogel et al. 1971) showed that rats suppress a reward-seeking behaviour when such a conflict is present. However, in these tasks, the reward-seeking action also caused

the delivery of the electric shock. This design does not allow for a manipulation on the memory of the aversive stimulus specifically, which was the intention of our task. Therefore, we introduced an escape platform to the CST in order to have aversive conditioning and avoidable shocks.

The work described in this chapter involved a pilot study of our novel CST. It was a proof of concept for our design, as it assessed the baseline suppression behaviour of the animals without any affective state manipulations. Two experiments were conducted using different types of conditioned stimuli, in order to assess whether the option to flee from the imminent shock and the lower shock intensity could still lead to reliable conditioned suppression measurements. A hierarchical statistical model was designed to capture the effects of each shock condition on the cohort as a group and on individual animals. Sampling-based approximate Bayesian inference was employed to fit the model to data from both experiments and calculate the posterior distributions of the possible effects.

The ultimate goal of this line of work is to conduct the Pavlovian conditioning part of the protocol in different contexts, while manipulating the animal's affective state during the same days. Following this training process, the memory of the aversive outcomes would be tested by reintroducing the same contexts along with the matching CS, while the animals would be free to act for reward. Previous experiments have indicated several contextual elements, such as odour of the operant chamber and visual patterns on its walls, that do not allow for generalisation across contexts during Pavlovian conditioning (Thomas et al. 2003).

3.2 Methods

3.2.1 Animals

The experiments in this chapter were conducted on one cohort of male Lister Hooded rats (N=16), supplied by Envigo UK. Rats weighed approximately 260-300 g at the start of training. They were housed in groups of two in cages with environmental enrichment, which consisted of a red 3 mm Perspex house (30 x 10 x 17 cm), a cardboard tube and a cotton rope suspended across the cage lid. Temperature was controlled within the range of 19-23 ^{o}C and humidity at 45-65%. Rat cages were placed in rooms with a twelve-hour reverse light cycle (lights off at 0800 hours)

to ensure behavioural testing was carried out during their active period (between 0800 and 1800 hours). Water was available ad libitum in the home cage. During training and testing, rats were maintained at no less than 90% of their free-feeding body weight with restricted access to laboratory chow (LabDiet, PMI Nutrition International) to 18 g per rat per day. All procedures were carried out under local institutional guidelines (approved by the University of Bristol Animal Welfare and Ethical Review Board) and in accordance with the UK Animals (Scientific Procedures) Act 1986. During experiments all efforts were made to minimise suffering, and at the end of experiments rats were killed by giving an overdose of sodium pentobarbitone.

3.2.2 Hardware

Standard rat operant chambers (Med Associates, Sandown Scientific, UK) were used to conduct all experimental studies. The dimensions of the chambers were 30.5 x 24.1 x 21.0 cm. Each chamber was positioned inside a light-resistant and sound-attenuating box that contained a ventilation fan. All chambers included two retractable response levers, which were located on each side of the food magazine. Two lever lights were positioned above each lever (1" White Lens, 28 V). A rubber tube connected the food magazine to the pellet dispenser, while a house light (28 V, 100 mA) was located above the magazine. The chamber floor was a grid of parallel stainless-steel bars (ENV-005A), connected to programmable shock boxes that could deliver scrambled electric shocks in the range of 0 to 10 mA. A removable stainless-steel waste pan (ENV-007A3, dimensions 29.1 x 25.48 x 1.91 cm) was placed underneath the floor bars, containing absorbent paper. The pellets used as reward in all sessions were 45 mg grain-based sweetened food pellets (5TUL Test Diet, Sandown Scientific, UK). An audio generator (ANL-926, Med Associates, Sandown Scientific, UK) produced tones that were delivered to each chamber via a speaker placed at the top of the chamber, above the left lever.

After the preliminary training stages, an escape platform was place on the grid floor. It was located adjacent to the wall, opposite of the levers and the food magazine. The platform was made of acrylic perspex and covered approximately 30% of the grid floor. Additionally, a camera (Microsoft LifeCam HD-3000) was attached to the ceiling of the operant chamber in order to both monitor the animals' behaviour during the experiments and record it in video files.

3.2.3 Software

The K-Limbic software (Conclusive Solutions Ltd., UK) operated as an interface between the operant chamber and a computer. It managed all input parameters to the chamber, such as dispensing pellets, extending and retracting levers and switching lights on and off. The software further controlled the audio generator and the shock boxes. It received chamber output through lever presses and magazine entries, which it turned into a Microsoft Excel file (.xlsx format) after the end of each session. The output file was a form of no consistent structure across entry rows. Custom code was developed in Julia (v1.6) (Bezanson et al. 2015) to read through the output form and extract the relevant entries (`https://github.com/harisorgn/CST`).

Video acquisition from all cameras was synchronised via a custom Python (v3.8) script that interfaced with the cameras simultaneously, using parallel processing (`https://github.com/harisorgn/multi_webcam`). Video frames were both displayed on a PC screen for online monitoring and saved as video files.

3.2.4 Task Design

The task was designed to measure the extent at which an animal suppresses a reward-seeking behaviour in the presence of an anticipated aversive stimulus. The two components of the task, reward seeking and being able to predict an imminent negative experience, were trained separately and finally combined in test sessions. The reward-seeking component was implemented as multiple lever presses leading to a reward pellet. This was achieved during the preliminary training. The aversive conditioning was done differently in the two experiments that the rat cohort took part in. Both experiments had a common Pavlovian conditioning basis with the same unconditioned stimulus, the delivery of an electric shock. However, the experiments had different types of CS that that the animals were trained to associate with the shock.

Preliminary training

The conditioned suppression protocol was initiated by training rats to respond to levers, as shown in Figure 3.1A. The first training stage was CRF (Continuous Reinforcement) Training, where one lever remained extended throughout a session

and each lever press was rewarded with one sugar pellet. Active levers alternated in consecutive sessions.

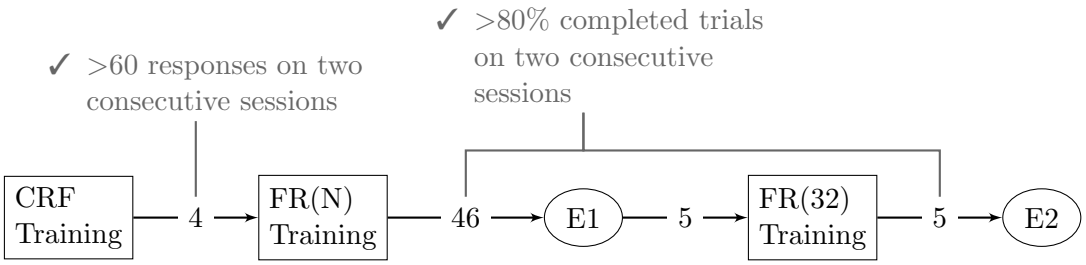
The following stage introduced the trial structure in the task, by keeping a lever extended for up to 60 seconds. Rats were required to press this lever N times, where N depended on the Fixed Ratio (FR) schedule, to gain one reward pellet. Consecutive presses had to be within 10 seconds of each other and all N presses had to be performed within the 60 seconds. If the N lever presses were not fulfilled in time, the lever was retracted and a reward pellet was not delivered. After the end of a trial, regardless of the outcome, there was an ITI of 20 seconds until the same lever was extended again for the next trial. The position of the active lever for each session alternated. The sequence of FR schedules was FR(1), FR(2), FR(4), FR(8) FR(16) and finally FR(32), where the number in the parenthesis denotes the number of lever presses required to receive a reward. The FR(32) schedule was used on all future sessions with active levers. The FR criterion in Figure 3.1A had to be met for each schedule before the next one started.

After successfully being trained to press a lever 32 times for a pellet reward, all rats progressed to the experimental phase of the protocol, including two different versions of evaluating the conditioned suppression behaviour. For all experiment sessions the escape platform was placed inside the operant chamber. Both experiments were designed around a five-day schedule, running one session per day, every day from Monday to Friday.

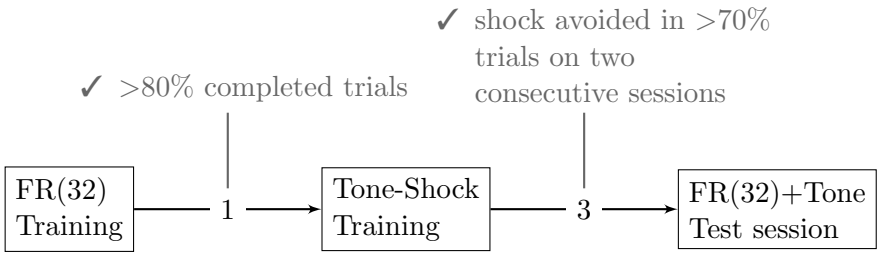
Experiment 1

The first conditioned suppression experiment is depicted in Figure 3.1B. Its first session was identical to FR(32) training, which was implemented to ensure that the rats' performance was stable. Following this session, 3 sessions of Pavlovian conditioning were conducted, training the rats to associate a CS with the delivery of an aversive unconditioned stimulus (US) and flee from it onto the escape platform. The CS in experiment 1 was a tone of 8 KHz and 65 dB, playing for 5 seconds. The US was a pulsed shock, delivered through the metal grid floor. The pulse pattern consisted of a time period of 0.2 seconds of active shock and 1.8 seconds of no shock, repeated 15 times. This pulse pattern was previously used in tasks involving active avoidance, as a less aversive alternative to continuous shock (Hinchcliffe 2019). Moreover, by including periods of no shock, we aimed at giving rats a better chance of

(A) **Preliminary training**



(B) **Experiment 1 (E1)**



(C) **Experiment 2 (E2)**

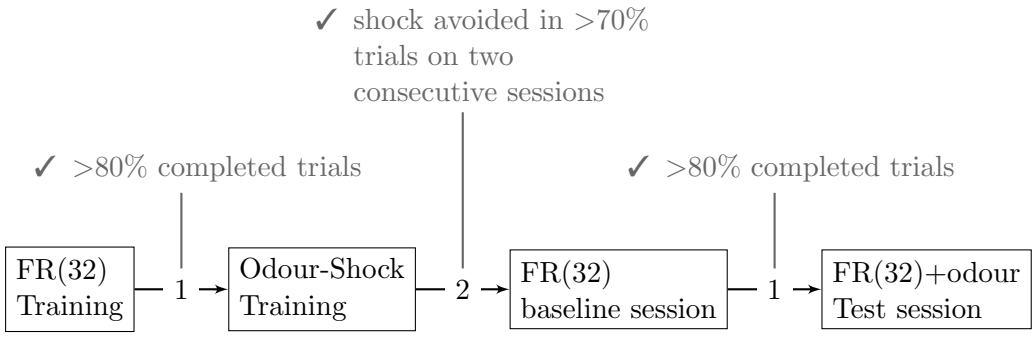


Figure 3.1: **Conditioned suppression protocol**: Preliminary training stages (**A**) were followed by two types of conditioned suppression tests, one involving tones as the CS (**B**) and the second one where CS was an odour, sprayed on the paper inside the waste pan **C**. Consecutive stages are linked by arrows, which contain within them the number of sessions animals spent on the first stage. Above each session number there is a list of criteria, in grey and denoted by a tick symbol, that animals had to fulfill to progress to the next stage. Squares represent individual stages and ellipses are experiments that contain a sequence of different types of stages.

escaping to the platform, without any disrupting motor effects of shock stimulation. There were 10 trials of tone-shock pairs with an ITI of 150 seconds between them. The ITI was set to be long enough in order to avoid associations being built between the context of the operant chamber and the shock stimulus, as it was shown for similar CS-US profiles in fear conditioning (Detert et al. 2008). Successful avoidance of the shock was measured manually, through the recorded videos.

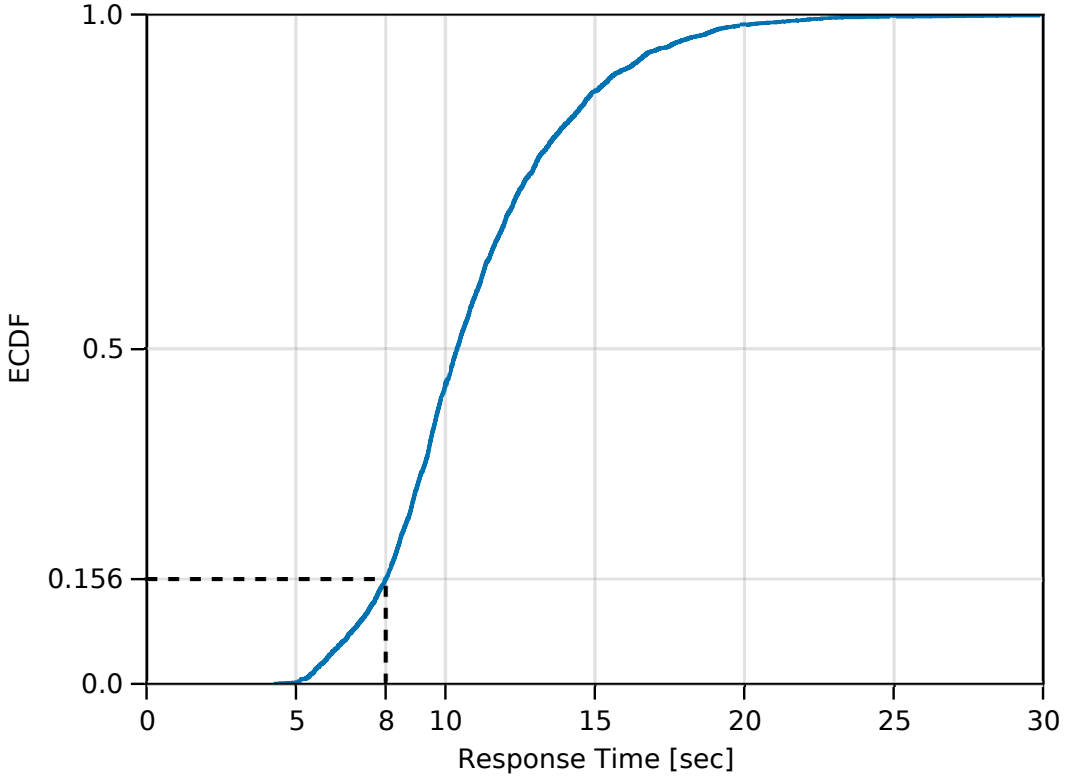


Figure 3.2: **Empirical cumulative density function of response time during training**: Response times from all animals were collated for the final week of FR(32) training and the empirical cumulative density function (ECDF) was calculated to estimate the proportion of trials that could be completed up to any number of seconds. The time point of interest was the time of shock onset in future test sessions. Given the chosen time delay between lever extension and tone onset, the tone duration and the delay between tone offset and shock onset, the first shock pulse would be delivered 8 seconds after lever extension. Within this time only 15.6% of trials could be completed, as shown by the dashed lines. Therefore it was unlikely for the rats to be able to finish an FR(32) schedule and flee from the imminent shock in time.

The last session of an experiment 1 week was the conditioned suppression test session. There were repetitions of a bank of trials, consisting of 8 trials identical to the FR(32) training trials and 1 trial, where an FR(32) trial and a CS tone presentation were ran in parallel. During the compound, FR(32) and tone, trials, the session

lever was extended and 2 seconds later the CS tone started playing for 5 seconds, as it was during Tone-Shock training. The delay between lever extension and tone presentation was chosen to encourage the rats to start pressing the lever before the tone that predicted an aversive event would be played. The delay of 2 seconds was chosen based on the rat's response times during the last week of FR(32) Training, shown in Figure 3.2. Approximately 15.6% of all trials were completed after 8 seconds, which is the time when the first shock is delivered (2 seconds lever-tone delay + 5 seconds tone presentation + 1 second tone-shock delay). Thus, the rats would have to decide on whether to carry on responding for food and risk shock exposure or flee, since it would be unlikely that they would manage to do both. This conflict was the purported driving force behind suppression behaviour. The assumption behind the chosen delay value was that the rats would start pressing a lever as soon as the lever was extended, which was a safe assumption to make as judged by qualitatively observing their behaviour via the cameras. Looking at the curve of Figure 3.2, it is evident that even after allowing for a couple of seconds more before the first lever press, up to 10 seconds since lever extension, still half of all trials would not have finished.

The week schedule of Figure 3.1B was repeated for two consecutive weeks, for two different values of shock intensity during the Tone-Shock Training sessions. These were 0.2 and 0.3 mA.

Experiment 2

At the end of experiment 1 there was a week of daily FR(32) sessions; this was used to verify that the animals' lever-pressing performance had not changed because of the aversive training and test session. It was noted during Tone-Shock training in experiment 1 (Figure 3.1B) that all of the animals spent most of the session time staying on the escape platform and not fleeing onto it during tone presentations. This observation suggested that the rats might have been conditioned on the chamber context during the Tone-Shock sessions, instead of on the tone exclusively. An example of a clear contextual difference between conditioning and testing sessions was the presence of a lever. In Tone-Shock training no lever was extended, while the FR(32)+Tone test session included an active lever on every trial. Rats could have used the absence of a lever as a cue that shocks could be delivered in the current session and thus chose to stay on the platform.

Experiment 2 was designed with the hypothesis of context conditioning in mind. It involved an explicit contextual element, which was present in both the shock conditioning and the test sessions. The aim was to promote stronger associations between the chamber context and the shock, which could in turn lead to higher suppression.

There were five sessions of different types, occurring daily from Monday to Friday, as shown in Figure 3.1C. The first session of experiment 2 was an FR(32) Training session, identical to that of experiment 1. The next two sessions were the conditioning sessions for this variant of the task. During Odour-Shock Training sessions, an odour was sprayed on the sheet of paper, which was placed inside the waste pan on the chamber floor. Essentially this session was identical to the Tone-Shock training of experiment 1 with the only exception being that there were no tones in Odour-Shock training. The conditioned stimulus now was the odour of the chamber. The pulsed shocks were the same as in experiment 1, being delivered with an ITI of 150 seconds. The papers where odour was sprayed were discarded after each session and the chamber was thoroughly cleaned.

Given that the animals were trained to associate the chamber odour with an electric shock, suppression was measured across two sessions. First, there was an FR(32) baseline session, when there was no added odour to the chamber. The next day, an FR(32)+odour test session took place and the conditioned odour was present throughout the session. The two sessions were identical in trial structure, matching that of FR(32) training, while the only difference was the presence of the conditioned odour in the final test session.

The session sequence of experiment 2, as depicted in Figure 3.1C, was repeated for four consecutive weeks. The first and second week contained electric shocks of 0.2 and 0.3 mA intensity respectively. At the beginning of the weekend between the second and third week the animals were taken off their food-restricted diet and given access to free food. The hypothesis behind this manipulation was that by increasing the rats' satiety, they would devalue the pellet reward, given the effort required to acquire it and consequently they would suppress or flee to the platform more often during the test session. The third and fourth week of experiment 2 tested this hypothesis with shock intensities of 0.3 and 0.2 mA respectively for all animals. The shock intensity was not counterbalanced within either pair of weeks that included the same dietary restrictions, as this was a pilot study to assess the animals sensitivity to the given range of shock stimulation. All animals experienced 0.2 mA on the

first week and 0.3 mA on the second. However, in order to account for confounding factors, such as re-exposure to shock causing stronger Pavlovian associations or sensitisation, the two shock intensities were presented in reverse order during the third and fourth week of the experiment, after the dietary manipulation.

The conditioned odour for the first and second week of experiment 2 was a bergamot essential oil, while the odour for the third and fourth week was a eucalyptus essential oil. Both oils were diluted in water, with 5 drops of oil in a 100 ml water bottle.

3.2.5 Model-based analysis

Inference on suppression behaviour given the data from experiments 1 and 2 was performed using a hierarchical Bayesian model. The model described our hypothesis about how the response time during each trial was generated. Its structure was

$$T_{s,c} \sim \text{LogNormal}(\alpha_{s,c}, \rho) \tag{3.1}$$

$$\alpha_{s,c} = \mu_c + \bar{\alpha}_{s,c}\sigma_c \tag{3.2}$$

$$\bar{\alpha}_{s,c} \sim \text{Normal}(0, 1) \tag{3.3}$$

$$\mu_c \sim \text{Normal}(2.5, 0.5) \tag{3.4}$$

$$\sigma_c \sim \text{Exponential}(0.5) \tag{3.5}$$

$$\rho \sim \text{Exponential}(0.5) \tag{3.6}$$

The first line of the model equations represented the likelihood density function that generated the response time data, T. The LogNormal distribution is a common choice for response time data, that can capture observed relationships between mean response time and variance in decision-making tasks (Wagenmakers et al. 2007). Response time in the current chapter was calculated as the time between the first and the last lever press in a trial. This definition differs from those in standard decision-making tasks, where response time is the time of deliberation before an action is taken. Despite this semantic difference, the LogNormal was chosen as the likelihood density function, as it can qualitatively match the distribution of observed response times and its parameters are easily interpretable.

All other lines containing a $\sim$ symbol represented model parameters to be inferred, given the observed response times, T, and the generative model. The $\alpha_{s,c}$ variable was an auxiliary quantity, representing the effect of a condition c on a subject s. The hierarchical structure of the model was based on the assumption that all

subject effects for a condition c were samples out of a Normal(μ_c, σ_c) distribution, where μ_c and σ_c were the mean and standard deviation of effect of condition c on the entire cohort respectively. Thus, subject-level parameters were sampled from cohort-level distributions. This sampling was indirect, by first sampling $\bar{\alpha}$ out of a standard normal distribution and consequently constructing α, in equation 3.2. This parameterisation was implemented to avoid pathological behaviour during the inference process, known as *funneling* (Betancourt et al. 2013). This way $\bar{\alpha}$ measured each subject's deviation away from the cohort mean μ_c in units of standard deviation σ_c, for each condition c.

The indices of the data and the model parameters were

$$\begin{aligned} \text{normalised trial time} &: t \in [0, 1] \\ \text{subject} &: s \in \{1, 2, \ldots, 16\} \\ \text{condition} &: c \in \{1, 2, \ldots, N_{\text{conditions}}\} \end{aligned}$$

where normalised trial time, t, was calculated by dividing all trial indices with the final trial index for each session of each animal. The other two variables were categorical and represented the subject and condition identities. All 16 subjects were included in the model fitting process. The number of conditions varied across the two experiments and are summarised in Table 3.1.

Experiment 1 involved comparisons of response time across different trial types, within the same session. The cohort-lever condition distributions of the model corresponded to the presence or absence of the tone, which was conditioned to predict the electric shock. In experiment 2, the list of conditions describes different sessions, marked by the presence or absence of the odour cue, which was the conditioned stimulus. The animals' diet was manipulated in the middle of experiment 2 and the same shock intensities were repeated with a new odour, in the reverse order with 0.3 mA preceding 0.2 mA. Behaviour during baseline sessions within the food restricted and the free-food weeks of experiment 2 was assumed to be the same. This assumption was based on the observed absence of difference between baseline trials during the 0.2 mA and the 0.3 mA weeks of experiment 1. However, it was assumed that baseline behaviour could be different after the change to free food, thus making it a separate condition in Table 3.1.

Approximate Bayesian inference was used to fit the present hierarchical model to the response time data. Given our prior beliefs about model parameters, the LogNor-

Experiment 1	Experiment 2
- Baseline trials, 0.2 mA week - Tone trials, 0.2 mA week - Baseline trials, 0.3 mA week - Tone trials, 0.3 mA week	- Baseline sessions, 0.2 and 0.3 mA weeks - Odour session, 0.2 mA week - Odour session, 0.3 mA week - Baseline sessions, 0.2 and 0.3 mA weeks, free food - Odour session, 0.3 mA week, free food - Odour session, 0.2 mA week, free food

Table 3.1: **List of conditions for both experiments :** Each condition described an experimental manipulation and corresponded to a cohort-level distribution of effects in the hierarchical model, out of which the individual, subject-level effects were being sampled. The conditions are listed in the same order at which the corresponding manipulations took place.

mal likelihood function and the observed response times, the posterior distribution of model parameters was approximated by sampling from it. This was achieved using the No-U-Turn-Sampler (NUTS) algorithm (Hoffman et al. 2014), a variant of the Hamiltonian Monte Carlo family of sampling methods (Betancourt 2018). Four independent chains of samples were used, each one containing 3000 samples. All sample values were then collated to approximate the posterior distribution of model parameters. The analysis was performed in Turing (v0.19.4), a probabilistic programming language based on the Julia programming language (Ge et al. 2018).

Prior specification

The choice of prior distributions for the model parameters is an integral part of Bayesian inference. These prior distributions are combined with the generative model of the data, as it was presented in the previous section, in order to lead to updated beliefs about the distribution of model parameters. In order to specify informative priors, that represent realistic expectations about response time data, the model parameters need to have an intuitive explanation. This is possible with our current model structure, equations 3.1-3.6. Our choice for the likelihood density

function of response time data, equation 3.1, was

$$T \sim \text{LogNormal}(\alpha, \rho) \tag{3.7}$$

by ignoring the subject and condition indices. The median of the LogNormal distribution is

$$m = e^{\alpha} \tag{3.8}$$

which after solving for α becomes

$$\log(m) = \alpha \tag{3.9}$$

Given training data that was not used in the fitting process, for example the one in Figure 3.2, I expected the median, m, of the response time distribution to be around 12 seconds. Then, $log(12) = 2.4849$, so centering the distribution of α around 2.5 was a reasonable expectation. The α parameter was a subject-lever parameter that was assumed to be sampled out of a Normal(μ_c, σ_c) distribution for each subject, within each condition c, as equations 3.2 and 3.3 indicate. Thus, its mean, μ_c was assigned the prior

$$\mu_c \sim \text{Normal}(2.5, 0.5) \tag{3.10}$$

with 2.5 reflecting the prior expectation for the median of response times under no effect of time and the standard deviation of 0.5 representing my uncertainty around the choice of the mean value. Most samples out of this distribution will be contained within $2.5 \pm 3 * 0.5$, which is the range $[1, 4]$. This range can be mapped to the more meaningful, real time scale, by transforming it with equation 3.8, resulting in $[2.7, 54.9]$ seconds. The range is quite large, however it contains not too unrealistic response time values even at its extremes. It should be noted that the animals had up to 60 seconds to fulfill the FR schedule and receive a pellet, thus the maximum value of 54.9 seconds for the median of the response time distribution is deemed a good starting point.

The remaining model parameters that required prior distributions were ρ and σ_c. Both parameters affect the variance of the response time distribution in non-linear ways, either directly (ρ) in the likelihood density or indirectly as the standard deviation of the cohort-lever distributions (σ_c). The same prior distribution or Exponential(0.5) was chosen for both parameters, as a prior that was not too restrictive.

A prior predictive test was conducted as a way of verifying the choices for the prior distributions of all model parameters, and particularly the choices for σ and σ_α that

have less intuitive interpretations. This test involved sampling parameter values from the respective prior distributions and then using those values to sample fictive response time data, from the generative model of equations 3.1-3.6. The output is plotted in Figure 3.3.

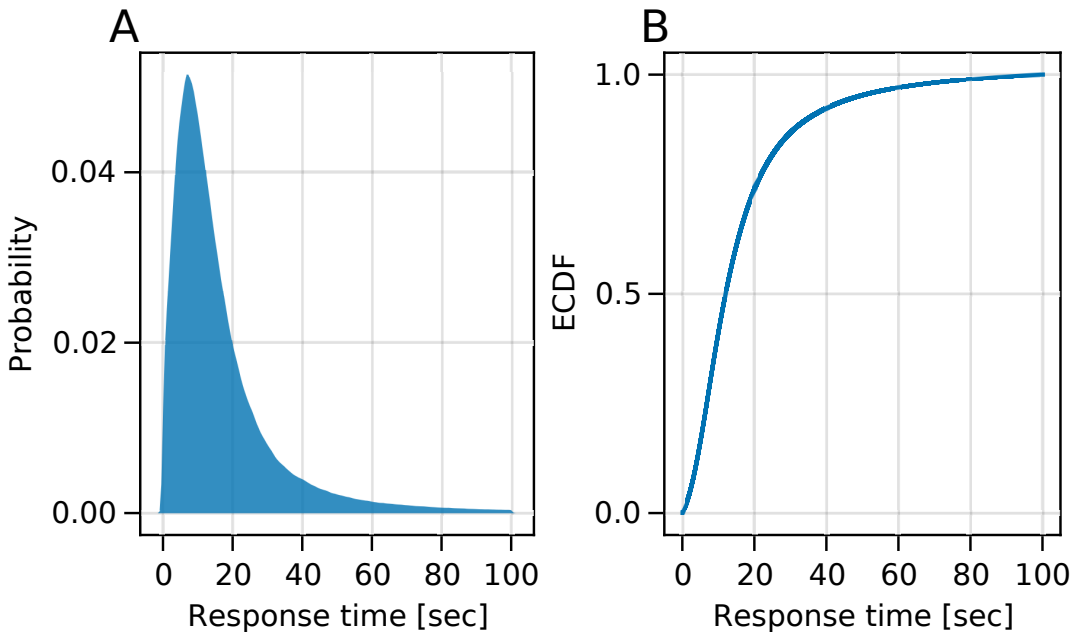


Figure 3.3: **Prior predictive test :** Fictive response time data was sampled using the proposed generative model. Values for all model parameters were first sampled from their prior distributions and consequently they were used to calculate the parameters of the LogNormal likelihood density and sample fictive response time values from it. **A**: the density of all fictive samples, across subjects and conditions, **B**: the empirical cumulative density function.

The density function of all fictive response time data in Figure 3.3A, collated across subjects and conditions, represents a reasonable expectation for the data. Additionally, it is not too restrictive by having a heavy tail towards larger response time values. Thus, the prior distributions for all model parameters were deemed informative for the subsequent inference process. The shape of the Empirical Cumulative Density Function (ECDF) in Figure 3.3B does not match that of the data from the final stage of training, as shown in Figure 3.2, however the median values are very close to each other, around 12 seconds.

3.2.6 Model diagnostics

The diagnostic checks of sampling algorithms of the Hamiltonian Monte Carlo family provide indications about the quality of the samples drawn during the inference process. The approximation of the posterior distribution is as good as the quality of the samples that are drawn from it, thus it is important to validate them. Ideally, the drawn samples will be independent from each other, both within and across sampling chains. Moreover, they should explore the entirety of the typical set of the posterior, that is the region where most of the probability mass is concentrated (Betancourt 2018). Three of the most common diagnostic plots will be used, that have proven to be useful particularly when performing inference on behavioural data from learning and decision-making tasks (Baribault et al. 2021).

The first diagnostic check that was implemented was the rank plot, as shown in Figures 3.4 and 3.5 for experiments 1 and 2 respectively (Vehtari et al. 2021b). This is a representative example of the rank plot on the mean effect of condition on the cohort level of the model, one of the main parameters that we were interested in. The plot is constructed by ranking the combined sampled values from all four chains into bins and consequently drawing histograms for each chain separately with the number of their samples that belong to each rank bin. Ideally the histogram of each chain should be close to a uniform distribution, indicating that the entire range of the sampled parameter was equally sampled by all chains. This was the case for the present inference process in both experiments.

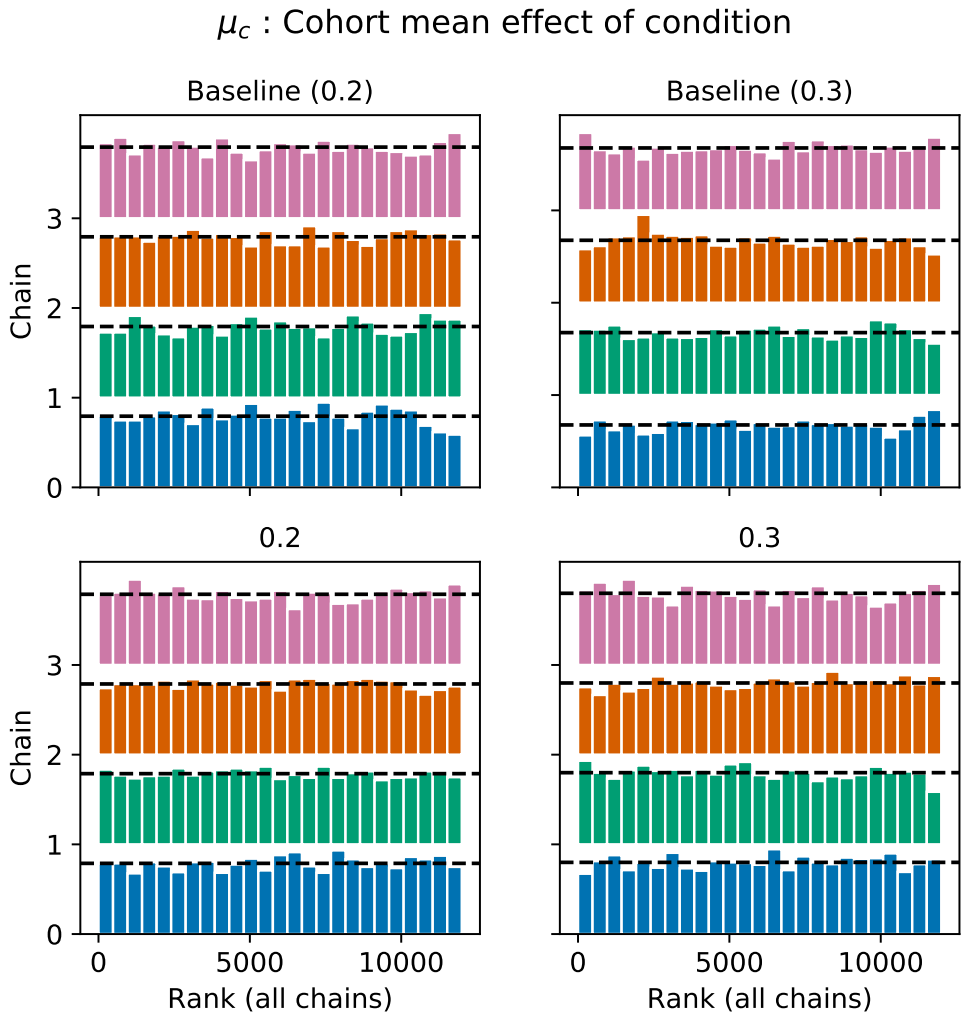


Figure 3.4: **Rank plot for the cohort mean parameter in experiment 1:** Samples of the μ_c parameter, collated across all four sample chains of the algorithm, were ranked according to their value. Consequently the ranks for each chains were drawn in separate histograms. The dashed lines represent a uniform distribution of ranks within each chain. This is the ideal distribution, as it implies that all chains sampled equally from the entire range of the posterior distribution of μ_c. The four panels correspond to rank plots for each condition. 0.2: 0.2 mA shock, 0.3: 0.3 mA shock

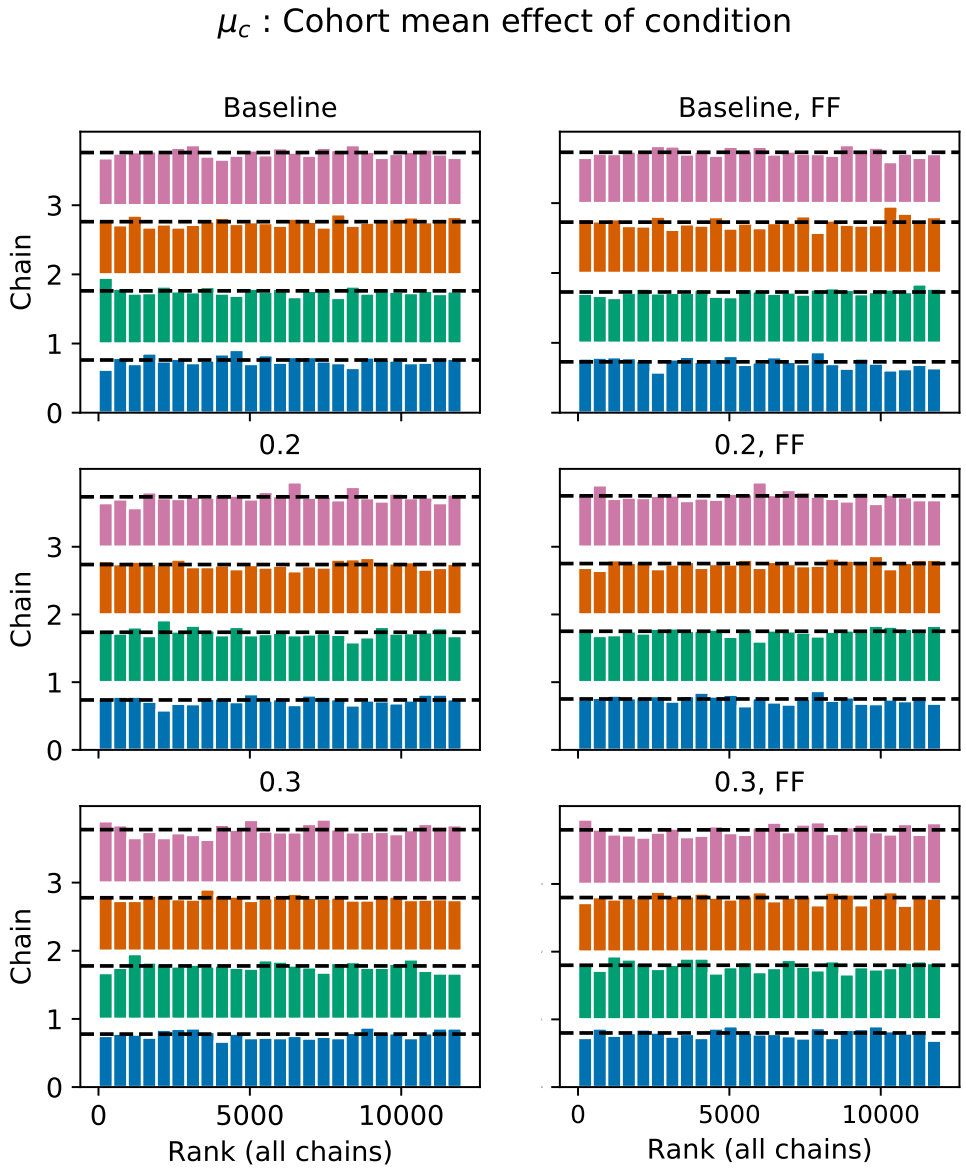


Figure 3.5: **Rank plot for the cohort mean parameter in experiment 2:** Samples of the μ_c parameter, collated across all four sample chains of the algorithm, were ranked according to their value. Consequently the ranks for each chains were drawn in separate histograms. The dashed lines represent a uniform distribution of ranks within each chain. This is the ideal distribution, as it implies that all chains sampled equally from the entire range of the posterior distribution of μ_c. The six panels correspond to rank plots for each condition. FF: Free Food, 0.2: 0.2 mA shock, 0.3: 0.3 mA shock

The effective sample size (ESS) of the chain samples was additionally calculated, as a way to assess whether there were enough independent samples drawn from the posterior distribution to approximate it. The ESS calculation combines samples drawn from all chains to estimate the effective number of samples coming from each quantile of the investigated parameter (Vehtari et al. 2021b). A small interval of 20 samples was used around each μ_c quantile. Figures 3.6 and 3.7 depict the results for the mean effect of cohort, μ_c, in both experiments. The dashed lines represent a recommended minimum of 400 samples, which is greatly surpassed across all quantiles for all conditions of both experiments.

Finally an energy plot was drawn as an additional measure of the extent at which the sampling chains explored an important part of the posterior distribution in an unbiased manner. The marginal energy distribution represents the distribution of energy levels that were visited during sampling, while the energy transition distribution is the distribution of energy level that could be reached between two consecutive samples (Betancourt 2018). Ideally both distribution should be matched. The marginal distribution was calculated as all visited energy levels minus the average energy level, so it contained a difference in energy values like the energy transition distribution did, thus making both of them directly comparable. This qualitative check is augmented by the Bayesian Factor of Missing Information (BFMI), which is the ratio of the variance of the energy transition distribution over the variance of the marginal energy distribution. Values of more than 0.3 are recommended, a criterion that was fulfilled for all four chains of both inference processes in Figure 3.8.

The rank and ESS plots were checked for the posterior distributions of all model parameters, which resulted from the inference processes in both experiments. Additionally the $\hat{R}$ metric was calculated (Gelman et al. 1992). This metric is a formula containing the ratio of the sample variability within each chain over the variability across chains. Values of $\hat{R} < 1.01$ are indicative of a healthy *mixing* behaviour in the chains, which means that all chains, across their sequence of samples, explored the regions of high posterior probability mass well enough, without getting stuck in any particular subregions.

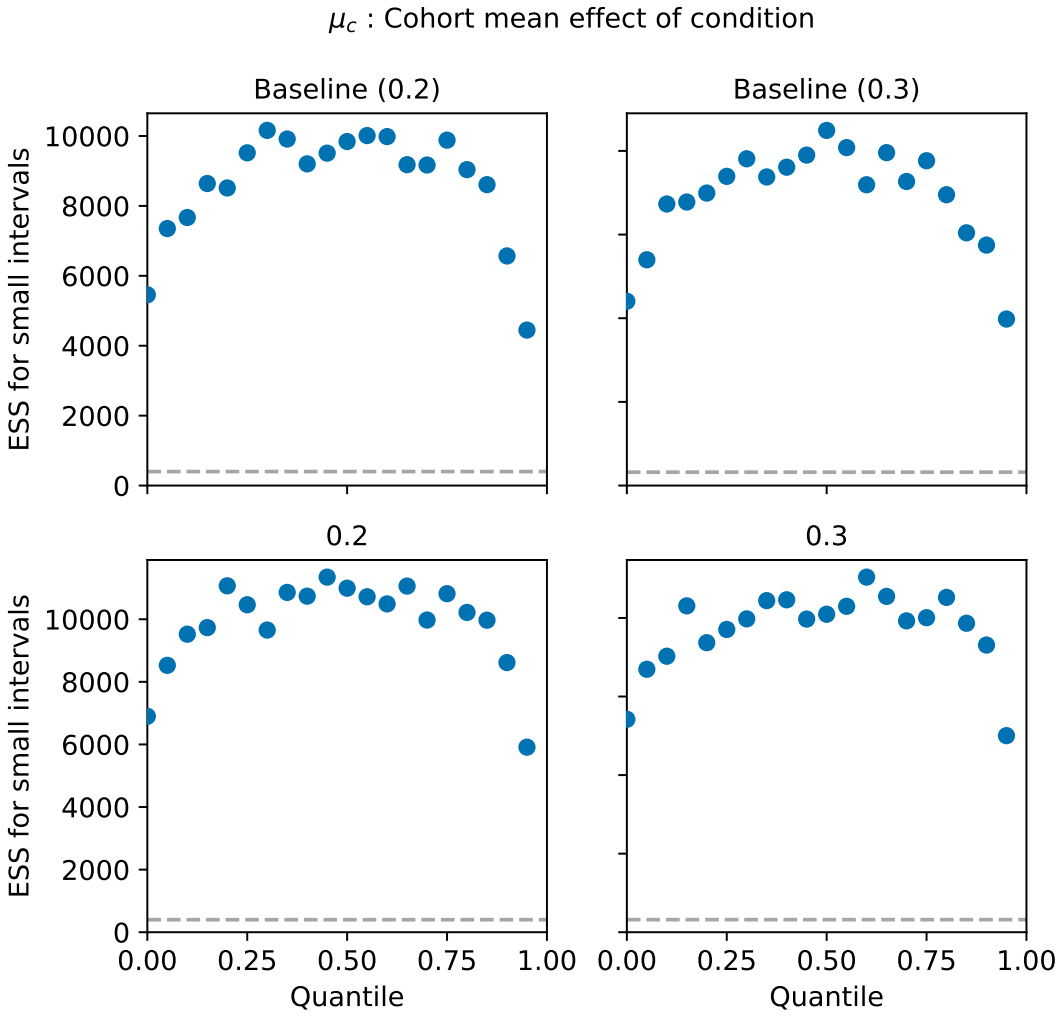


Figure 3.6: **Effective sample size plot for the cohort mean parameter in experiment 1:** The effective sample size (ESS) was calculated on the combined samples from all four chains, for a small region around each quantile of μ_c, the cohort mean effect of condition. A dashed line is drawn along the minimum recommended sample size of 400 samples. The ESS values are well above the minimum for all quantiles of μ_c. The four panels correspond to ESS plots for each condition. 0.2: 0.2 mA shock, 0.3: 0.3 mA shock

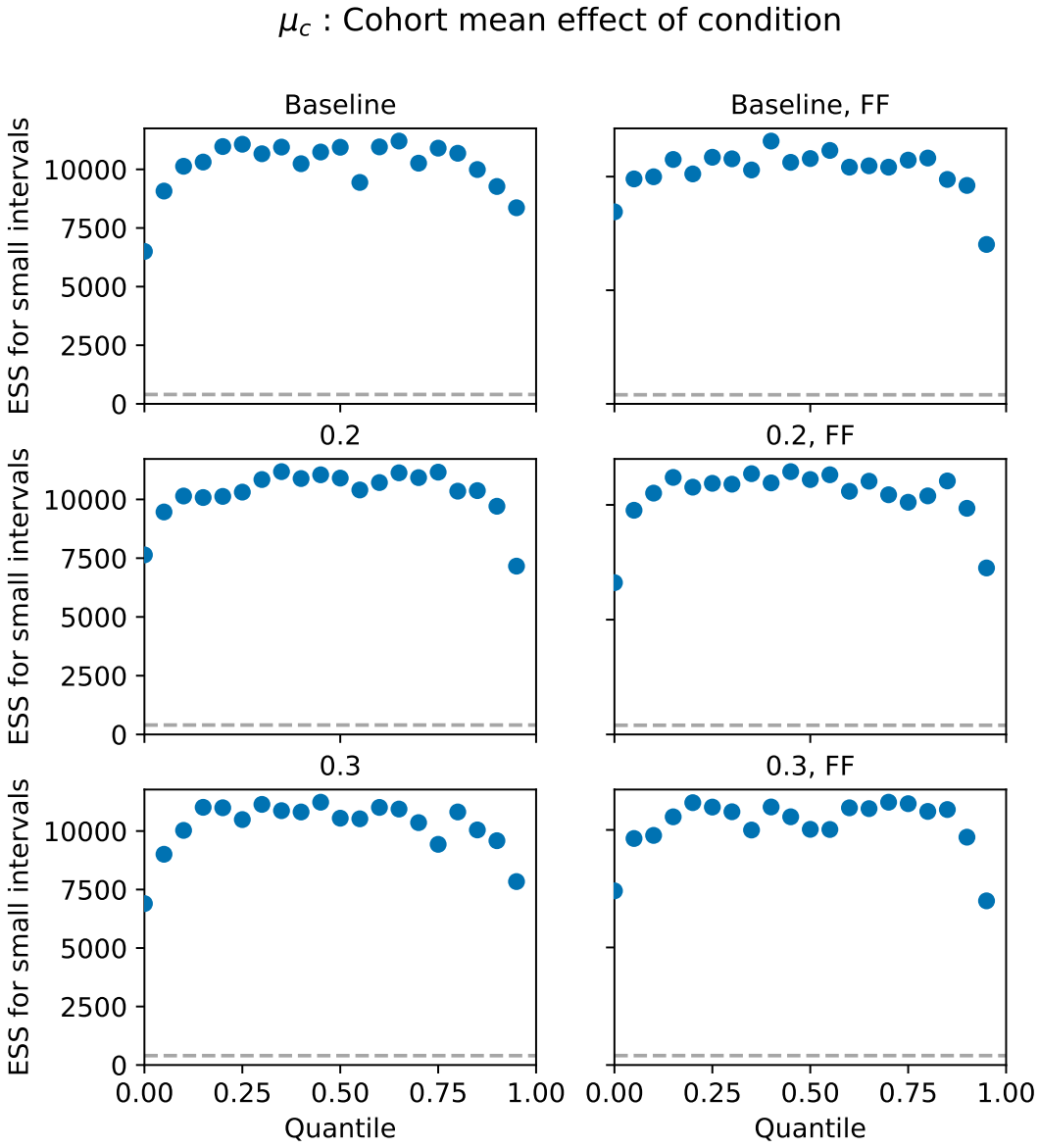


Figure 3.7: **Effective sample size plot for the cohort mean parameter in experiment 2:** The effective sample size (ESS) was calculated on the combined samples from all four chains, for a small region around each quantile of μ_c, the cohort mean effect of condition. A dashed line is drawn along the minimum recommended sample size of 400 samples. The ESS values are well above the minimum for all quantiles of μ_c. The six panels correspond to ESS plots for each condition. 0.2: 0.2 mA shock, 0.3: 0.3 mA shock

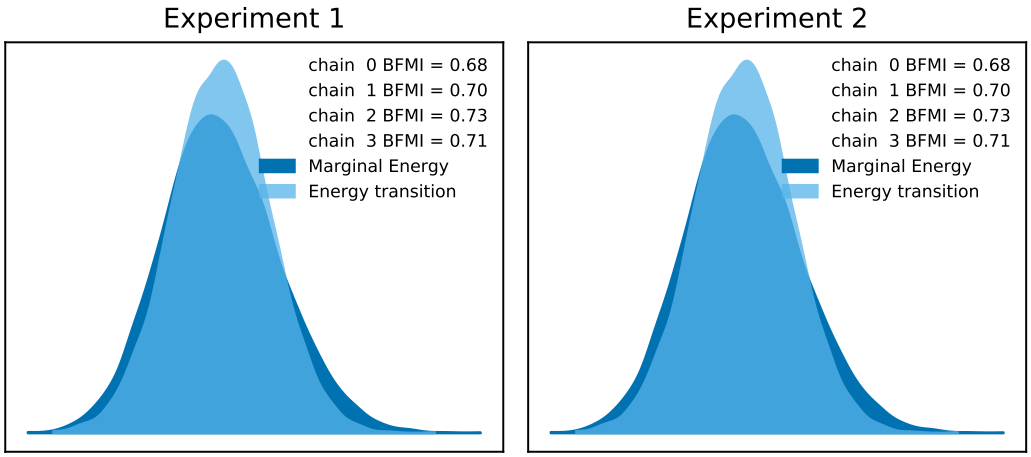


Figure 3.8: **Energy plot of the sampling process in experiments 1 and 2:** The marginal distribution of all visited energy levels during sampling is plotted in dark blue, along with the distribution of the transition between energy levels corresponding to two consecutive samples of the algorithm in light blue. The average energy value, among those visited, was subtracted from all visited energy values to calculate the marginal energy distribution. The Bayesian Factor of Missing Information (BFMI), the ratio of the variance of the energy transition distribution over the variance of the marginal energy transition, is shown in the legend for each chain of the algorithm. The recommended minimum value of BFMI was 0.3. The left panel depicts the energy plot for the sampling process using the experiment 1 data and the right panel is the same plot for the experiment 2 data.

3.2.7 Statistical analysis of model results

As discussed in section 3.2.5, the result of performing Bayesian inference on the hierarchical model was a posterior distribution over the model parameters, informed by the observed data. The original hypothesis of this chapter was that the animals would suppress their responses during the presence of a shock-predicting cue. In order to quantify the effect of the experimental manipulations on suppression, I calculated the probability that the difference in response time between a shock condition and its respective baseline was positive. This was equivalent to measuring the cumulative probability of the posterior distribution of this difference being positive. Thus, only the right tail of the distribution was considered, as it reflected the original hypothesis. A lower bound of 0.2 seconds was used instead of the zero value, assuming that differences in response time that were smaller than 0.2 seconds could have been attributed to motor effects relating to executing the lever-pressing movement.

This probability, named P_+, was the cumulative density function value of the posterior of the difference between shock and baseline conditions. It was defined as

$$P_+(x > 0.2) = \int_{X_+} P(x|D)dX_+ \approx \frac{|\tilde{X}_+|}{N} \tag{3.11}$$

where x is the difference in the median of the response time distribution between shock and baseline conditions, either on the group or the subject level. The posterior $P(x|D)$ of the difference in medians X given data D is integrated across

$$X_+ = \{x : x > 0.2\} \tag{3.12}$$

to result in the desired cumulative density function value, representing the probability of suppression. However, since the posterior $P(x|D)$ is only approximated in a sample-based manner, this cumulative probability was also approximated by dividing the cardinality of the set $\tilde{X}_+$, representing the number of samples for which $x > 0.2$ holds, by the total number of samples drawn, N.

3.3 Results

Bayesian inference was performed on the model of section 3.2.5, by fitting it to response time data from experiments 1 and 2. Omission trials, when the FR(32) schedule was not fulfilled were excluded from the fitting process. Inference results are distributions on the model parameters, representing the updated knowledge that was gained by observing the data, relative to the prior expectations before the experiments.

The main parameter of interest was the effect of shock intensities on the entire cohort. Thus, the posterior distribution of the cohort-level mean of the distribution of possible individual effects under each condition, μ_c, was transformed by the exponential function, in order to map it to the scale of response times in seconds. This result is presented in Figure 3.9A for experiment 1 and in Figure 3.11A for experiment 2. Both figures include the transformed prior distribution of μ_c, common for all conditions c. This transformed parameter could be interpreted as a distribution of the median of the response time distribution of each subject under each condition. The prior and posterior distributions are centered around similar values of approximately 12 seconds, however the uncertainty around this value was greatly decreased in the posterior, as the result of the inference process.

In order to better visualise the effect of each shock condition, the difference in the posterior distributions of e^{μ_c} between tone and baseline trials is plotted in Figures 3.9B and 3.11B for both experiments. These differences are centered close to zero, with the probability of suppression P_+ being not higher than 0.6 for any shock condition across the two experiments. Only in the case of the Free Food (FF) manipulation did the peak of the distribution of cohort effects shift towards positive values in Figure 3.11B, resulting in a probability of suppression $P_+ \approx 0.6$.

Looking within each shock condition, a posterior distribution of subject-level effects is drawn in Figures 3.10 and Figure 3.12 for experiments 1 and 2 respectively. This is the difference in the exponential transformation, f, of the intercept

$$f(\mu_c, \bar{\alpha}_{s,c}, \sigma_c) = e^{\mu_c + \bar{\alpha}_{s,c}\sigma_c} \tag{3.13}$$

between a shock and its respective baseline condition, for each subject s. Conceptually each distribution represents the uncertainty about a subject's median response time, given that this median was a sample out of the cohort-level distribution of the

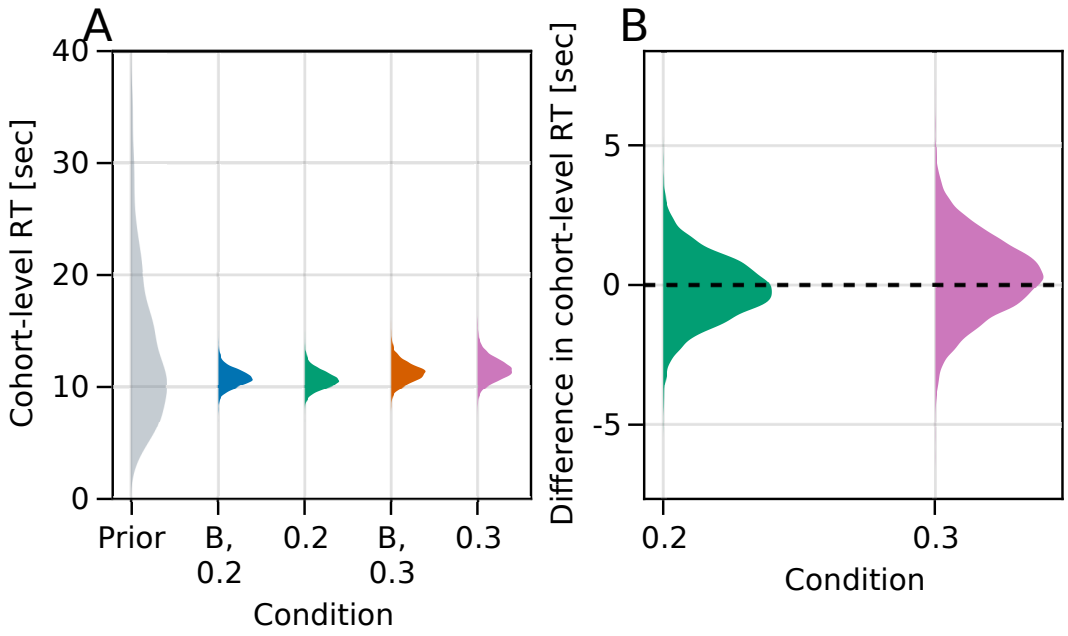


Figure 3.9: **The effect of shock conditions on the cohort level in experiment 1: A**: The posterior distribution of the exponent of the cohort-level mean parameter, e^{μ_c}, under each condition c is plotted against the common prior distribution, transformed by exponentiation as well, which reflected the expectation before the inference process. **B**: The effect of condition was calculated as the difference between a shock condition and the respective baseline. Baseline *B, 0.2* was subtracted from *0.2* and *B, 0.3* from *0.3*, resulting in the two distributions of differences. A dashed line across 0 indicates the absence of an effect. $P_+ = 0.381$ for the *0.2* condition and $P_+ = 0.535$ for the *0.3* one. B: Baseline, 0.2: 0.2 mA shock, 0.3: 0.3 mA shock.

same condition. The condition colours of cohort-level and subject-level distributions are consistent between Figures 3.9 and 3.10, and between Figures 3.11 and 3.12.

The subject-level effects fluctuate around zero, which explains the similarity between the cohort-level posterior distributions in Figures 3.9B and 3.11B. It should be noted though that there are subjects for whom the effect of shock is very likely to be non-zero. Such effects exist on either side of the zero line, indicating that some animals were going faster during tone trials that were associated with the electric shock, while others were responding more slowly, as our original suppression hypothesis predicted. The probabilities of suppression, P_+, for each subject are summarised in Table 3.2 for experiment 1 and in Table 3.3 for experiment 2. This change in the median of a subject's response time distribution between a shock and a baseline condition is more apparent in experiment 2, as seen both in the figures and the tables. Moreover, the FF manipulation on the last two weeks of experiment 2 resulted in

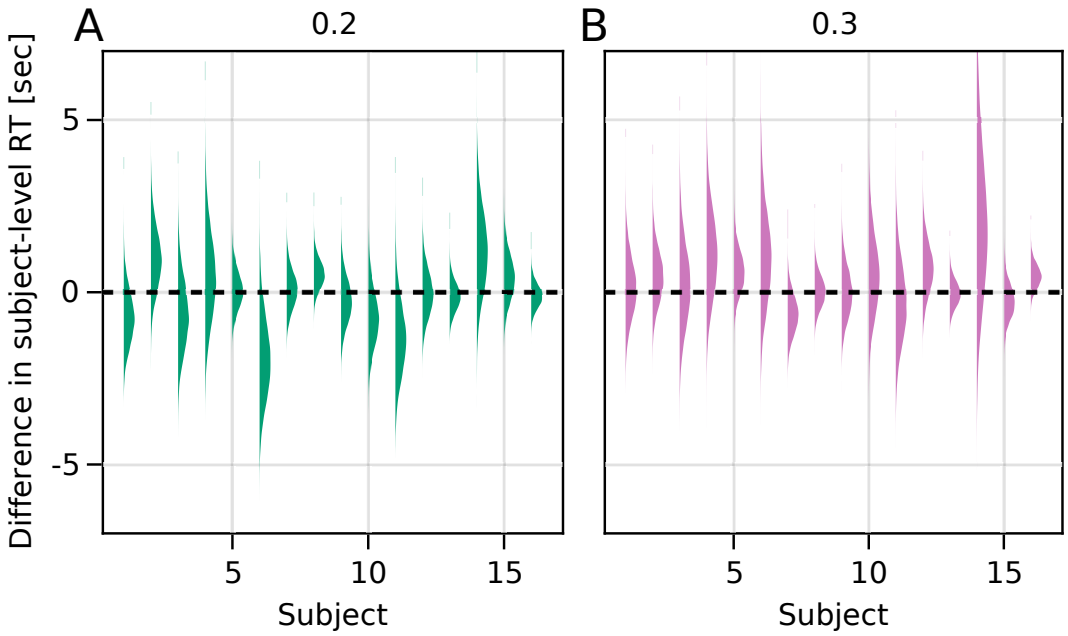


Figure 3.10: **The effect of shock conditions on the subject level in experiment 1:** Posterior distributions of the difference in the exponent of the subject-level intercept parameter, $e^{\mu_c + \bar{\alpha}_{s,c}\sigma_c}$ between shock and baseline trials, **A**: for the 0.2 condition and **B**: for the 0.3 condition. A dashed line across zero indicates the absence of an effect. The probability of suppression for each subject is presented in Table 3.2. 0.2: 0.2 mA shock, 0.3: 0.3 mA shock.

larger fluctuations for some animals, given the larger scale of the y-axis in Figures 3.12C and 3.12D.

	Conditions	
Subject	*0.2*	*0.3*
1	0.122	0.465
2	0.8	0.624
3	0.21	0.5
4	0.572	0.76
5	0.39	0.734
6	0.042	0.725
7	0.41	0.08
8	0.79	0.353
9	0.254	0.382
10	0.13	0.6
11	0.075	0.33
12	0.36	0.755
13	0.217	0.266
14	0.8	0.8
15	0.632	0.2
16	0.2	0.785

Table 3.2: **Probability of suppression for each subject in Experiment 1 :** The probability of a suppression P_+, that is a positive shift in RT, was measured for each subject individually, using the data from Figure 3.10.

	Conditions			
Subject	*0.2*	*0.3*	*0.3, FF*	*0.2, FF*
1	0.0	0.99	0.0	0.99
2	0.914	0.003	0.0	1.0
3	0.158	0.008	0.99	0.042
4	0.0	0.99	0.0	1.0
5	0.0	0.977	0.372	0.77
6	0.12	1.0	0.39	0.58
7	0.0	1.0	0.0	1.0
8	1.0	0.946	0.0	
9	1.0	0.0	1.0	0.0
10	1.0	0.008	0.993	0.717
11	0.431	0.69	0.39	0.436
12	0.812	0.146	0.458	0.305
13	0.0, 1.0	0.017	0.0	
14	0.137	1.0	0.127	0.98
15	0.0	0.161	0.008	0.096
16	1.0	0.0	1.0	0.0

Table 3.3: **Probability of suppression for each subject in Experiment 2 :** The probability of a suppression P_+, that is a positive shift in RT, was measured for each subject individually, using the data from Figure 3.12.

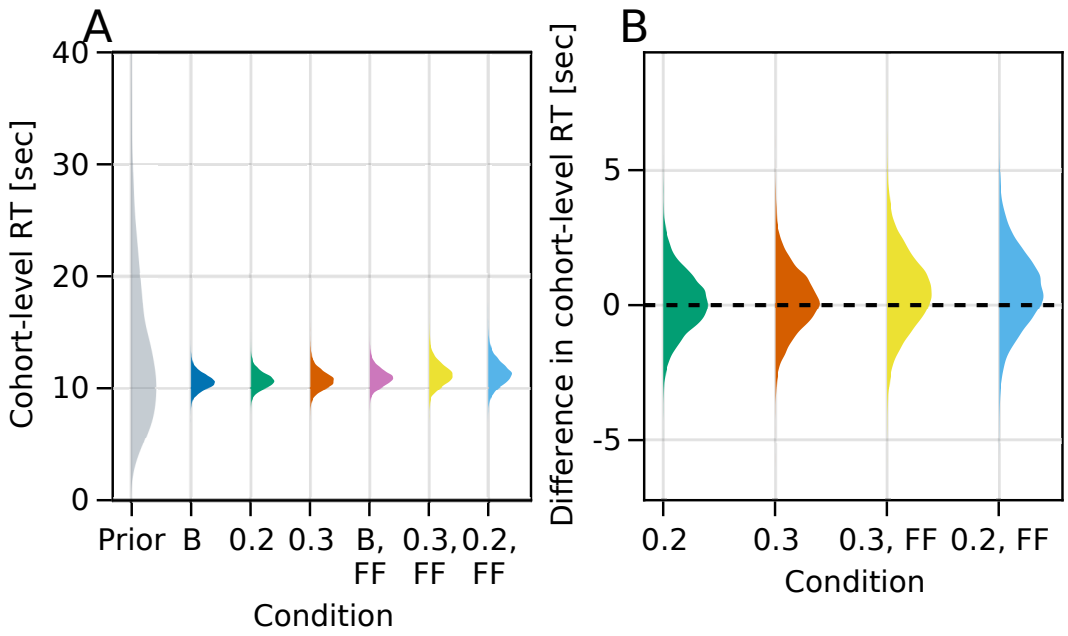


Figure 3.11: **The effect of shock conditions on the cohort level in experiment 2:** **A**: The posterior distribution of the exponent of the cohort-level mean parameter, e^{μ_c}, under each condition c is plotted against the common prior distribution, transformed by exponentiation as well, which reflected the expectation before the inference process. **B**: The effect of condition was calculated as the difference between a shock condition and the respective baseline. Baseline B was subtracted from the *0.2* ($P_+ = 0.467$) and *0.3* ($P_+ = 0.481$) shock conditions and *B, FF* from the *0.3, FF* ($P_+ = 0.587$) and *0.2, FF* ($P_+ = 0.6$) conditions. The result was the four distributions of differences. A dashed line across 0 indicates the absence of an effect. B: Baseline, 0.2: 0.2 mA shock, 0.3: 0.3 mA shock, FF: Free Food.

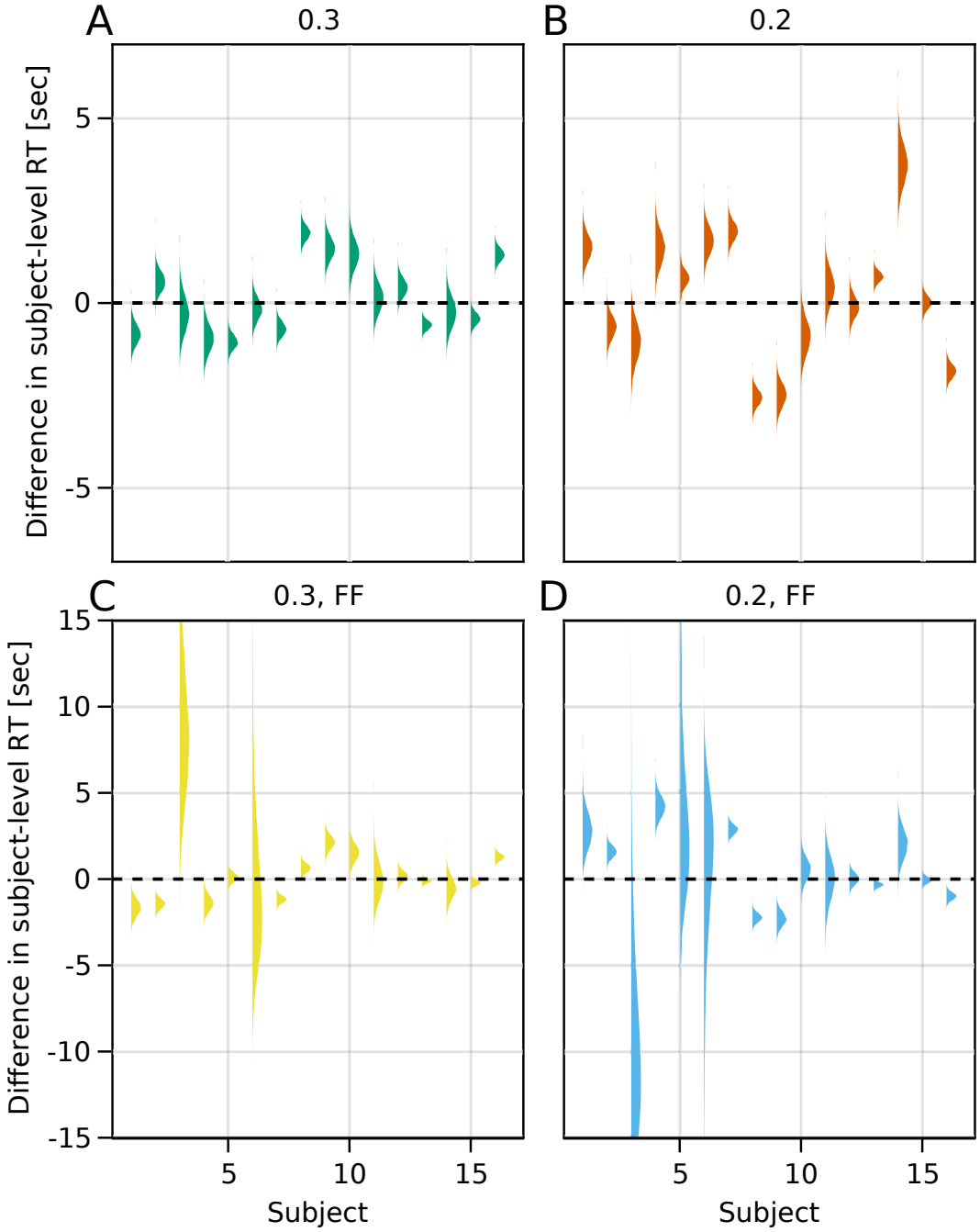


Figure 3.12: **The effect of shock conditions on the subject level in experiment 2:** Posterior distributions of the difference in the exponent of the subject-level intercept parameter, $e^{\mu_c + \bar{\alpha}_{s,c}\sigma_c}$ between shock and baseline trials **A**: for the 0.2 condition and **B**: for the 0.3 condition. A dashed line across zero indicates the absence of an effect. 0.2: 0.2 mA shock, 0.3: 0.3 mA shock, FF: Free Food.

3.4 Discussion

The experimental work of this chapter was a proof of concept for a novel task design based on the conditioned suppression task. Two versions of our task, both involving escapable shock stimuli of low intensities, were tested with regard to their ability to elicit suppression of a reward-seeking behaviour during conditions that predicted the shock stimuli. The versions differed in terms of the condition that was paired with the delivery of the pulsed shock. The conditioned context of experiment 2, using an odour cue, was more successful in causing a change in the animals' response time, compared to the more classical tone-shock association of experiment 1. The direction of the effect however varied across individual animals.

A hierarchical statistical model was designed to infer the effects of different shock intensities on the animals' suppression behaviour, on individual subjects and on the group level of the entire cohort. The results of a Bayesian inference process revealed that the shock-predictive condition did not have a highly certain effect on the cohort level, as measured by the cumulative probability of suppression P_+. This observation held for both experiments, with either one of the 0.2 and 0.3 mA shock intensities. Reducing the animals motivation for reward by putting them onto free food, during the third and fourth week of experiment 2, revealed only a small positive shift towards suppression for the same shock intensities, as was shown in Figure 3.11B. The potentially more sated state of the animals during these weeks did not lead to elevated suppression due to the hypothesised decrease in the value of the pellet rewards.

The subject level of the model provided a more detailed view of the animals' behaviour. Conditions associated with shock had a more certain non-zero effect on a few animals in experiment 1 (Figure 3.10 and Table 3.2) and even more animals in experiment 2 (Figure 3.12 and Table 3.3). Particularly in the presence of the odour cue in experiment 2, individual animals exhibited shifts in the median of their response time distribution, with extreme values of P_+. Interestingly, such effects were observed on either side of the zero line, implying that some animals were responding more quickly during the shock-associated condition, while others showed signs of classic suppression by responding more slowly. The effects were more pronounced in the food-restricted condition, in Figure 3.12A and 3.12B. The estimated subjects' effects during the free-food weeks, in Figure 3.12C and 3.12D, were more variable, probably reflecting the increase in omission trials, which resulted in less data to inform the model, thus making inference more uncertain.

It should also be noted that the shocks during the Odour-Shock training sessions of experiment 2 were presented at seemingly random times from the animals' point of view, as opposed to experiment 1, when each shock was preceded by a conditioned tone. The higher unpredictability of shock delivery in experiment 2 could have exacerbated the individual differences within the cohort, thus resulting in more non-zero individual effects. Nevertheless, a conditioned suppression study with inescapable shocks found that rats were suppressing more after a tone predicting a certain shock, compared to a tone predicting an uncertain shock (Wright et al. 2019). Thus, the individual differences of experiment 2 could relate to the nature of the CS. A herbal odour might is a more naturalistic stimulus for a rat as opposed to a pure tone. This difference in CS could have caused the underlying difference among individual animals between experiments.

Facilitation, as opposed to suppression, of the reward-seeking behaviour has been previously reported in a variant of the conditioned suppression task with inescapable shocks and conditioned stimuli instead of a conditioned context and only after extensive training (Strickland et al. 2021). Despite the task design differences between this study and ours, it is possible that a common underlying factor is the individuals' ability to adapt to the threat of shock, as suggested by the authors of the study. For instance, rats that sped up their responding while the odour was present could potentially have done so in order to acquire the reward and flee to the platform before the shock is presented.

Individual differences when responding to aversive outcomes have additionally been reported for fear conditioning (Bush et al. 2007; Ji et al. 2018 Jan-Dec), where 25% of rats were considerably slower at reducing their freezing during extinction sessions. This result was correlated with increased anxiety-like behaviour in secondary tasks. Moreover, when animals were being trained to act in order to avoid an incoming shock, approximately 25% of them were unable to meet the training criteria, because they resolved to freezing instead of fleeing to safety (Laughlin et al. 2020). Thus, individual differences are apparent in different tasks that involve both avoidable and inevitable punishment and they potentially indicate inherent differences around the processing of punishment. It is therefore very likely that such differences were present within our cohort and they led to the divergence of effects of shock conditions in experiment 2.

Increasing the intensity of the electric shock has been shown to result in smaller individual differences (Pietersen et al. 2006). This reduction in inter-individual

variability is associated with more traumatic experiences, which typically elicit a freezing response. This behaviour is more relevant for an animal model of PTSD. However, we aimed at designing a behavioural assay that could work as an animal model of anxiety, by inducing and manipulating aversive memories, that are not traumatic.

Given all evidence about individual differences in experiencing and responding to shock stimulation, particularly one of low intensity, a hierarchical model was deemed the most beneficial approach to analysing the experimental data. By structuring the model parameters on a cohort- and a subject-level, we allowed for variations in the response times of each individual both within and across experimental conditions, while constraining the amount of variation between individuals that were assigned to the same experimental condition. Our model was an initial attempt to capture effects of shock on both of these levels in our new task. It could be expanded to include more degrees of freedom, for example by allowing the second parameter of the LogNormal likelihood, ρ to vary across subjects and conditions in order to investigate changes in the variability of response times under the threat of shock.

Overall, neither the conditioned tone nor the conditioned context resulted in clear effects on our cohort, in experiments 1 and 2 respectively. We aimed at increasing the welfare standard of behavioural tasks that measure responses to aversive, yet not traumatic, events by designing a task with escapable, low-intensity shocks. However, these features did not cause enough conflict in our rats to make them more indecisive about whether to flee or press for reward, so that they would eventually respond more slowly. On an individual level though, the second version of the task, involving odour cues as the conditioned context, did result in clear individual differences. Given that manipulations of contextual elements were part of the original plan, this version would seem the most promising for future studies. Any future work should include a fully counterbalanced design in terms of the order to presentation of each shock intensity to avoid issues around differential sensitisation within the cohort.

Another element of our task that needs to be addressed in future work is the Pavlovian conditioning sessions, which are essential for training the animals to associate the conditioned context or stimulus with the electric shock. In both experiments of our task, these sessions were conducted under escapable shocks. It was observed through the recorded videos that all 16 animals spent most of the session time on the escape platform, as early as the second training session. Thus, we did not have a reliable measure of the extent at which they were trained to associate the condi-

tioned context or stimulus with the shock. A different training protocol could be implemented in future studies, that involves Pavlovian conditioning sessions with inescapable shocks in a first phase. The escape platform could then be introduced during further Pavlovian conditioning sessions. This way the first phase would enable more reliable conditioning and the second phase would involve avoidance training. This is akin to the training process of active avoidance tasks, where the reward-seeking component is not present (Bravo-Rivera et al. 2014; LeDoux et al. 2017).

Chapter 4

A model-based analysis of the judgement bias task

4.1 Introduction

The primary statistical approach to the Judgement Bias Task (JBT) is to measure the Cognitive Bias Index (CBI) and test whether the average CBI in a cohort of animals has been significantly altered under some affective state manipulation. This inference is based on the assumption that CBI is exclusively driven by the animal's prior expectation of an outcome, which in turn is modulated by affective state. However, it could be possible that multiple factors are driving the animals' choices during ambiguous trials. These factors could themselves be modulated by affective state. For example, past experiences of reward have been shown to lead to a more positive affective state during a subsequent decision task (Nygren et al. 1996; Iigaya et al. 2016). Conversely, the omission of an expected reward could lead to a negative affective state. This has been corroborated by Pavlovian conditioning studies where a cue, that predicted the omission of reward with certainty, blocked the conditioning to a new cue predicting an electric shock, thus suggesting generalisation across punishment and reward omission (Dickinson et al. 1980). Additionally, the influence of reward omission on affective state is present in situations of uncertain rewards and relates to their predictability (Rutledge et al. 2014; Blain et al. 2020).

Effects of the history of reinforcement are not limited to studies that explicitly investigate affective state. There are evidence from both sensory decision making tasks in rats and humans (Akrami et al. 2018; Roy et al. 2021) and value-based tasks

in humans (Bornstein et al. 2017) and macaques (Wittmann et al. 2020), suggesting that the outcomes of past trials influence future actions. Importantly, these studies included fully-trained subjects and learning of cue-action-outcome contingencies was not required during their sessions.

The JBT is based upon both sensory- and value-based decisions. The experience of reward or its omission is an integral part of the task, occurring multiple times within a session. The current chapter used a large dataset of JBT sessions to visualise the CBI across trials and the probability of taking an action conditioned on the feedback of the most recent trial. These visualisations suggested that reinforcement history affected the animals' actions. Thus, accounting for these factors could enable a finer look into the influence of affective state on actions under ambiguity. Hierarchical statistical models were employed in order to dissect the effects of feedback and the interpretation bias, which the task was originally designed to measure. The models encompassed the effects of past actions, the reinforcement history, perceptual and reward-driven biases and lapses to explain the animals' choices.

The effect of reinforcement history was split into two separate factors, each one operating on a different timescale. Firstly, the immediate past was considered to capture the animals' sensitivity to the most recent feedback given their chosen action. This effect could be interpreted as the extent at which the animals employ a *win-stay, lose-shift* strategy (Robbins 1952). This strategy suggests that animals are more likely to repeat an action that has been recently rewarded and switch to a different action after the omission of reward. It has been suggested that animals have a propensity to utilise this heuristic rule because it can lead to optimal resource acquisition in certain foraging scenarios (Charnov 1976). Supporting evidence for this strategy have been observed across species and decision-making tasks (humans: Worthy et al. 2012; Ivan et al. 2018, monkeys: Medin 1972, rats: Reed 2016; Rayburn-Reeves et al. 2013).

The second factor, that was linked to reinforcement history, considered a longer time window into the past to capture the effect of the average reward earned. Besides the influence on affective state mentioned above, the average reward of the past, or reward rate, has been shown to affect animals' decisions in tasks that were not designed to promote this effect (Wittmann et al. 2020; Scholl et al. 2015). Here, this effect was incorporated into a generative model of decision making in two ways; as the average reward of past ambiguous trials, in the learning model, and as the average reward of all past trials, which was part of the reward rate model. An influ-

ence of past reinforcement that was specific to ambiguous trials could indicate that the animals were trying to learn new ambiguous cue-action-outcome contingencies, whereas the influence of all past outcomes could modulate the motivational state of the animal by altering its state of hunger.

Data from multiple past studies, including cohort 1 of chapter 2, was used to perform inference on the model parameters. The data included baseline sessions, a ketamine and a vehicle treatment. Approximate Bayesian inference resulted in posterior distributions for model parameters, with each parameter reflecting the effect of one of the considered factors, either on a population or a subject-specific level. Subsequently the learning and reward rate models were compared in order to estimate which one described the animals' data more accurately.

The animals were also split into two populations, according to the breeding facility that supplied them. As discussed in chapter 2, the animal supplier was the main difference between cohorts of rats, where ketamine had successfully induced a positive shift in CBI (Hales 2016; Hales et al. 2017), and cohorts that subsequently failed to replicate this result. After leaving the bredding facility, the animals were housed in the same facility during the experiments and were handled by the same people. Thus, my current investigation constituted a more detailed analysis of the animals' actions by considering potential effects from past reinforcement and differences in their strategies, both within and across the two populations of distinct origins. The positive effect of amphetamine on the CBI of cohort 1 in chapter 2 was additionally reexamined in light of the two proposed models.

4.2 Methods

4.2.1 Model definition

A hierarchical Bayesian model was designed to estimate the effects of past trials, perceptual and interpretation biases on the animals' behaviour during the ambiguous cue trials.

There were three levels to our model. First was the data level, that corresponded to the subjects' choices. Next was the subject level containing parameters for each individual animal that were used to generate that animal's choices for each trial of each condition. Subject-level parameters were sampled from distributions that were pa-

rameterised by the population-level parameters of the top level. The population-level parameters were the mean and standard deviation of the distributions of possible subject-specific effects for each condition.

The choice data was generated by each subject in each trial of each condition, according to

$$A_{c,s,t} \sim \text{Binomial}(P_{c,s,t}) \tag{4.1}$$

for each trial t of subject s within each condition c. The range of each index was

$$\begin{aligned} \text{trial} &: t \in \{1, 2, \ldots, N_{\text{trials}}\} \\ \text{subject} &: s \in \{1, 2, \ldots, N_{\text{subjects}}\} \\ \text{condition} &: c \in \{1, 2, \ldots, N_{\text{conditions}}\} \end{aligned}$$

with N_{trials} being the number of trials completed by an animal in each condition. One condition corresponded to a single session, which could be a baseline, a vehicle, a ketamine or an amphetamine session. The number and types of conditions used in each inference process is mentioned in the following Results section, along with the number of subjects.

The Binomial distribution of equation 4.1 was the likelihood of observing each action given a model. Actions were encoded in a binary format with 0 corresponding to choosing the high-reward lever and 1 representing low-reward lever choices. The probability of choosing the low-reward lever was defined as

$$P_{c,s,t} = \frac{\epsilon_s}{2} + \frac{1 - \epsilon_s}{1 + e^{-K_{c,s,t}}} \tag{4.2}$$

which had the form of a logistic function with an added lapse rate ϵ_s. The lapse rate was the probability of a subject choosing at random between the two levers. We assumed that it only varied across subjects and not across conditions, as the drug studies used to fit the model contained a low dose of ketamine or amphetamine that has not been associated with the dissociative effects that are typical in higher dose volumes.

When choices were not made at random, with probability $1-\epsilon_s$, they were generated according to a logistic function, parameterised by

$$K_{c,s,t} = b_{c,s} + w_{c,s} I_{c,s,t} + a_{c,s} \hat{R}_{c,s,t} \tag{4.3}$$

The three terms that made up $K_{c,s,t}$ in equation 4.3 corresponded to three types of effects we were expecting to be important in driving animals' behaviour. The bias variable, $b_{c,s}$, was a subject-specific intercept term for each experimental condition. In other words, it represented the bias that an animal had towards selecting an action, irrespective of past reinforcement. Given the JBT design, it was interpreted as the combined perceptual and interpretation biases. Perceptual biases described the tendency to match the sound of the ambiguous cue to one of the reference cues. The prior expectation for either a small or a large reward that was originally hypothesised to drive ambiguous cue responses in chapter 2 was the interpretation bias component of $b_{c,s}$.

The $w_{c,s}$ variable was added in order to estimate the extent at which animals used a *win-stay, lose-shift* strategy. Its regressor $I_{c,s,t}$ was an indicator function of the choice and the outcome of the previous trial. It was defined as

$$I_{c,s,t} = \begin{cases} +1, \text{if} \quad A_{-1} = \text{LA and } R_{-1} = 1 \\ -1, \text{if} \quad A_{-1} = \text{LA and } R_{-1} = 0 \\ -4, \text{if} \quad A_{-1} = \text{HA and } R_{-1} = 4 \\ +4, \text{if} \quad A_{-1} = \text{HA and } R_{-1} = 0 \end{cases} \tag{4.4}$$

where HA and LA were the actions paired with the high- and low-reward levers respectively. Since low-reward actions were coded as 1 in the Binomial distribution of the likelihood (equation 4.1), positive values of $I_{c,s,t}$ indicate a tendency to choose the low-reward lever. We assumed that this *win-stay, lose-shift* strategy was modulated by the reward magnitude that was experienced (positive $I_{c,s,t}$ values) or missed (negative $I_{c,s,t}$ values).

The remaining variable in the model of an animal's actions was $a_{c,s}$. This was the weight that animals put in the average past reward $\hat{R}_{c,s,t}$. Two different models were fit to JBT data. Their structure was identical to the model presented here, with the only difference being how the $\hat{R}_{c,s,t}$ regressor was calculated. In one model, titled learning model, $\hat{R}_{c,s,t}$ was the average reward of past trials when the ambiguous cue was presented, whereas in the other model, called reward rate model, it was calculated as the average reward across all types of past trials. The $\hat{R}_{c,s,t}$ value of each trial was calculated for a time window containing the 12 trials that preceded it, for both models. This window length was chosen as JBT sessions were split into 10 banks of 12 trials each. Each trial bank was guaranteed to contain four trials of each one of the two reference cues and four ambiguous trials. Thus, the chosen time

window was most likely to contain an equal number of trials of each cue.

All subject-level variables, $b_{c,s}, w_{c,s}, a_{c,s}, \epsilon_s$, followed the non-centered parameterisation. Random parameters were sampled from a standard Normal prior distribution as

$$\bar{b}_{c,s} \sim \text{Normal}(0, 1) \tag{4.5}$$
$$\bar{w}_{c,s} \sim \text{Normal}(0, 1) \tag{4.6}$$
$$\bar{a}_{c,s} \sim \text{Normal}(0, 1) \tag{4.7}$$
$$\bar{\epsilon}_s \sim \text{Normal}(0, 1) \tag{4.8}$$

which were then used to construct the subject-level variables

$$b_{c,s} = \mu_c + \bar{b}_{c,s}\sigma_c \tag{4.9}$$
$$w_{c,s} = \xi_c + \bar{w}_{c,s}\gamma_c \tag{4.10}$$
$$a_{c,s} = \eta_c + \bar{a}_{c,s}\rho_c \tag{4.11}$$
$$\epsilon_s = \Phi(\kappa + \bar{\epsilon}_s\omega) \tag{4.12}$$

based on these subject-level parameters and the population-level mean and standard deviation parameters. The inference method employed required that all model parameters were sampled from unconstrained distributions, defined on the real number axis. However, the lapse rate represented a probability, so it had to be mapped to the [0,1] interval. This was achieved by transforming the unconstrained $\kappa + \bar{\epsilon}_s\omega$ through the use of the cumulative density function of a Normal(0,1) distribution, $\Phi(x)$, in equation 4.12. This strategy has been used in similar hierarchical models for learning and decision-making (Ahn et al. 2017).

The population-level mean parameters were sampled from their respective prior distributions according to

$$\mu_c \sim \text{Normal}(0, 1) \tag{4.13}$$
$$\xi_c \sim \text{Normal}(0, 1) \tag{4.14}$$
$$\eta_c \sim \text{Normal}(0, 1) \tag{4.15}$$
$$\kappa \sim \text{Normal}(-1.5, 0.2) \tag{4.16}$$

while the population-level standard deviation parameters were sampled from their

prior distributions as

$$\sigma_c \sim \text{Exponential}(1) \tag{4.17}$$

$$\gamma_c \sim \text{Exponential}(1) \tag{4.18}$$

$$\rho_c \sim \text{Exponential}(1) \tag{4.19}$$

$$\omega \sim \text{Exponential}(0.5) \tag{4.20}$$

Constructing the subject-level variables according to equations 4.9-4.12 was equivalent to directly sampling them from the population-level distributions, for example $b_{c,s} \sim \text{Normal}(\mu_c, \sigma_c)$. However, the indirect parameterisation employed here, via the $\bar{b}_{c,s}, \bar{w}_{c,s}, \bar{a}_{c,s}, \bar{\epsilon}_s$ parameters was implemented to avoid pathological behaviour during the inference process, known as "funneling" (Betancourt et al. 2013). This way, $b_{c,s}, w_{c,s}, a_{c,s}, \epsilon_s$ were not directly inferred. They were auxiliary variables that were used internally by the model.

The subject-level parameters and the population-level mean and standard deviation parameters containing a $\sim$ comprised the model parameters. These were assumed to be sampled from the respective prior distributions, which in turn represented our belief about each parameter before observing the choice data. Inference was then conducted to update these prior distributions to posterior ones given the data from ambiguous trials. Model lines containing a = are definitions of auxiliary variables that were used internally by the model.

Approximate Bayesian inference was used to fit these hierarchical models to the choice data. Given our prior beliefs about model parameters, the likelihood function and the actions, the posterior distribution of model parameters was approximated by sampling from it. This was achieved using the No-U-Turn-Sampler (NUTS) algorithm (Hoffman et al. 2014), a variant of the Hamiltonian Monte Carlo sampling method (Betancourt 2018). Four independent chains of samples were used, each one containing 2000 samples. All sample values were then collated to approximate the posterior distribution of model parameters.

Prior specification

The prior distributions for the population-level mean effects of bias (μ_c), average reward of the past (η_c) and *win-stay, lose-shift* behaviour (ξ_c) were centered around zero to be unbiased towards either lever. The standard deviation for these mean

parameters was chosen as one. This choice, and the choice of Exponential(1) for the prior of all the population-level standard deviation parameters $(\sigma_c, \rho_c, \gamma_c)$ are common choices for prior distributions that we do not have previous knowledge of, yet they are not completely uniform and uninformative (McElreath 2016).

The prior for the mean and the standard deviation of the distribution of possible lapse rates were more informed, since we expected the animals to not behave at random during these test sessions after successfully passing all training stages. The population-level mean, κ, was centered around -1.5, which after the transformation gives $\Phi(-1.5) \approx 0.065$. The $\Phi(x)$ function used to map unconstrained real values to [0,1] is a continuous and monotonic function, thus the transformed parameter ϵ was expected to be centered around 0.065. This corresponded to a 6.5% rate of lapsing and choosing randomly. The uncertainty around this mean value, reflected by the standard deviation of the κ prior distribution, and the standard deviation of the population-level distribution, ω were tuned in order to attain a reasonable prior distribution. In order to assess the effect of the two standard deviation terms, the distribution of subject-level lapse rates was plotted, by sampling from the population-level priors and transforming through $\Phi(x)$. The eventual choice of prior distributions for both κ and ω resulted in the distribution of lapse rates for individual animals that is shown in Figure 4.1. Most of the probability mass was concentrated in small values of ϵ. This expectation was reasonable, given the fact that animals were highly trained on the task. Moreover, the distribution had a long tail to allow for deviations towards large probabilities of lapsing. As such, this choice of population-level priors made modest assumptions about the data.

4.2.2 Model diagnostics

Diagnostic checks were performed to ensure that the four chains of consecutive samples, as calculated by the NUTS algorithm, were independent of each other and covered all regions of the posterior distribution equally well. The energy plot, the effective sample size for each quantile plot and the rank plot were visually inspected for all parameters. A more detailed presentation of these plots and the check that is performed in each one was presented in chapter 3. Moreover, the $\hat{R}$ metric was calculated for each parameter of both models and it was verified that no values were larger than 1.01, as recommended in Gelman et al. 1992. Both the plot-based diagnostic checks and the $\hat{R}$ metric were successfully fulfilled, thus validating that the sampling chains of both models indeed contained representative samples out of

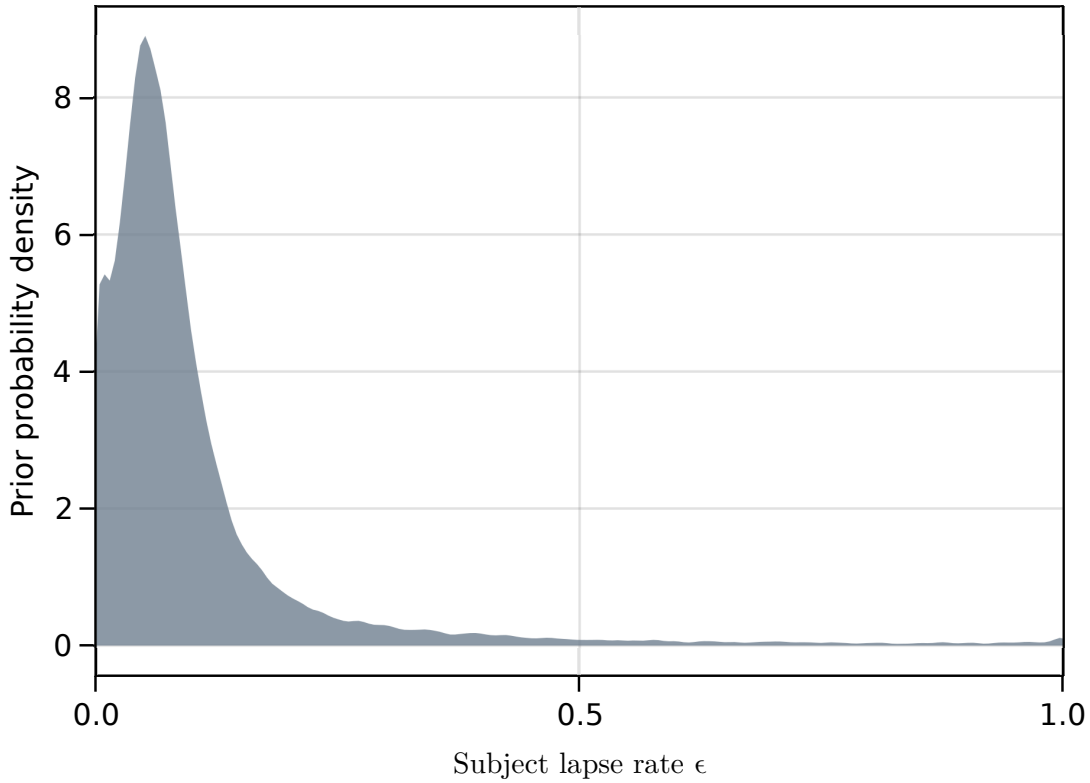


Figure 4.1: **Prior distribution of subject-level lapse rates:** Lapse rates for individual animals were calculated using the transformation $\Phi(\kappa + \bar{\epsilon}_s\omega)$ after sampling κ, ω and $\bar{\epsilon}_s$ from their respective prior distributions. This check was performed in order to evaluate the distribution of lapse rates that were expected prior to observing the data. The depicted prior distribution of ϵ was considered a reasonable expectation.

the posterior distribution.

4.2.3 Statistical analysis of model results

A similar approach to Section 3.2.7 was taken for effect quantification by calculating cumulative probability values from the posterior distribution of the difference between a manipulation and its baseline condition. Unlike Chapter 3 though, there was no a priori expectation about the direction of an effect. A region of practical equivalence (ROPE) was chosen around zero to indicate values that are equivalent to zero (Kruschke 2018). A difference between treatment and baseline parameter values that falls within the ROPE is interpreted as there was no effect of treatment on the particular parameter.

The entire posterior of difference between treatment and baseline for each parameter was used to calculate the ROPE. This metric has exhibited satisfactory robustness to noise and minimal false positive results, relative to other established Bayesian metrics for measuring effects, in simulation studies (Kelter 2020).

The choice of ROPE was based on the effect that the difference in a parameter would have on the probability of choosing the low-reward action, $P_{c,s,t}$ (equation 4.2). All model parameters are centered around zero, which leads to $P_{c,s,t} = 0.5$. I treated probability values in the interval $(0.45, 0.55)$ as equivalent to 0.5. Thus the ROPE for the logistic regression parameters would be

$$|x| < \frac{\text{logit}(0.55) - \text{logit}(0.45)}{4} \approx 0.1 \tag{4.21}$$

so the ROPE interval would be $(-0.1, 0.1)$ for any parameter difference x. The logit function in equation 4.21 is the inverse of the logistic function of equation 4.2. The division-by-four is an approximation to the actual change in probability for logistic regression models, which is more accurate when there are similar number of both choices in the data (Kruschke 2018). Rats in the present experiments have not shown a strong bias for either the high- or low-reward choices during ambiguous trials, thus this approximation was considered appropriate.

The cumulative probability that a difference x falls inside the ROPE is

$$P_0(-0.1 \leq x \leq 0.1) = \int_{-0.1}^{0.1} P(x|D)dx \approx \frac{\tilde{X}_0}{N} \tag{4.22}$$

where $P(x|D)$ is the posterior of the difference x, given data D. Similar to Section 3.2.7, since the posterior was approximated using a chain of samples, the cumulative probability of being within the ROPE is also approximated in a Monte Carlo manner, where $\tilde{X}_0$ is the number of posterior samples that are within the ROPE and N is the total number of samples.

The metric that is reported in this chapter when posterior distributions are compared is the complementary probability to the ROPE probability

$$P_* = 1 - P_0(-0.1 \leq x \leq 0.1) \tag{4.23}$$

representing the probability of a non-zero effect, that is a difference in a parameter that falls outside the ROPE.

4.2.4 Model comparison

The learning and reward rate models were evaluated in terms of their predictive accuracy after successful inference of the posterior distribution of their parameters. The Leave-One-Out Information Criterion (LOOIC) was used for this comparison (Vehtari et al. 2017). The LOOIC is based on cross-validation and evaluates the probability of a model observing some held-out data point, after it is fitted on the rest of the available data. The benefit of this metric over alternative information criteria, such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC), is that LOOIC is a pointwise estimate of the result of leaving out each data point, resulting in a distribution of values, unlike AIC and BIC which are scalar estimates. Thus, model comparison uses the variance of the LOOIC values as an estimate of uncertainty about the accuracy of each model. Typically, the difference in the mean of LOOIC between models is accompanied by the standard error of this mean.

Since multiple iterations are require, one for each choice of excluded data point, the complete cross-validation process is, in general, computationally expensive, often prohibitively so. The Pareto-Smoothed Importance Sampling (PSIS) is an approximation to the complete cross-validation process that greatly reduces the computational cost of calculating LOOIC. Moreover, PSIS offers a diagnostic check of how well it approximated the cross-validation, with suggested guidelines on how to interpret indications of a bad approximation (Vehtari et al. 2021a).

4.2.5 Software

The analysis was performed in Turing (v0.19.4) (Ge et al. 2018), a probabilistic programming language based on the Julia programming language (v1.6) (Bezanson et al. 2015). The implementation of PSIS-LOOIC from the ArviZ package for Julia (Bezanson et al. 2015) was used for model comparison (Kumar et al. 2019). All code to extract the data, run the analysis and produce the following figures is freely available online (`https://github.com/harisorgn/JBT`).

4.3 Results

4.3.1 Raw data visualisation

The influence of past trials on the rats' behaviour during ambiguous trials was first visualised using data from baseline sessions for 199 animals. This group of animals included the rats of both the original (N=31) and the replication (N=52) ketamine studies, corresponding to the two breeding facilities. The rest of the animals (N=116) originated from either one of the two breeding facilities and took part in JBT experiments within the same experimental facility as the animals of the ketamine studies. There were two baseline sessions per animal.

Firstly, each session of each animal was split into five blocks, with each block containing approximately 24 trials. Each block contained two banks of 12 trials each, thus it was almost guaranteed to contain 8 trials of each reference cue and eight trials of the ambiguous cue, given the bank structure of the session. The Cognitive Bias Index (CBI) within each block is shown in Figure 4.2A. Individual animals span the entire range of CBI values from -1 to 1 across all blocks. However, animals exhibit shifts in their CBI values as trials progress. This shift is illustrated in Figure 4.2B, which contains the difference in CBI between consecutive blocks of trials. Specifically early trials, between the first and second block, exhibit a decrease of CBI for more animals compared to later blocks. Individual animals are spread on both sides of the zero line in Figure 4.2B suggesting an effect of past trials on animals choices during the ambiguous trials.

The effect of the immediate past was investigated as well. The extent to which the action and the outcome of the most recent trial affected the current action during an ambiguous trial was measured in 4.3 as the ratio

$$\text{Conditional probability ratio} = \frac{P(A|C = \text{AC}, A_{-1}, R_{-1})}{P(A|C = AC)} \tag{4.24}$$

where A was the current action, C was the current cue that was equal to the ambiguous cue AC and A_{-1}, R_{-1} where the previous action and outcome respectively. Ratio values equal to 1 suggest that the choice of the current action was independent of the previous action and outcome. Values larger than 1 suggest that the animal was more likely to choose action A after A_{-1} and R_{-1} than it did overall in the session. Conversely, ratio values smaller than 1 imply that the animal was less likely to choose A given the past action and outcome compared to how often it chose A

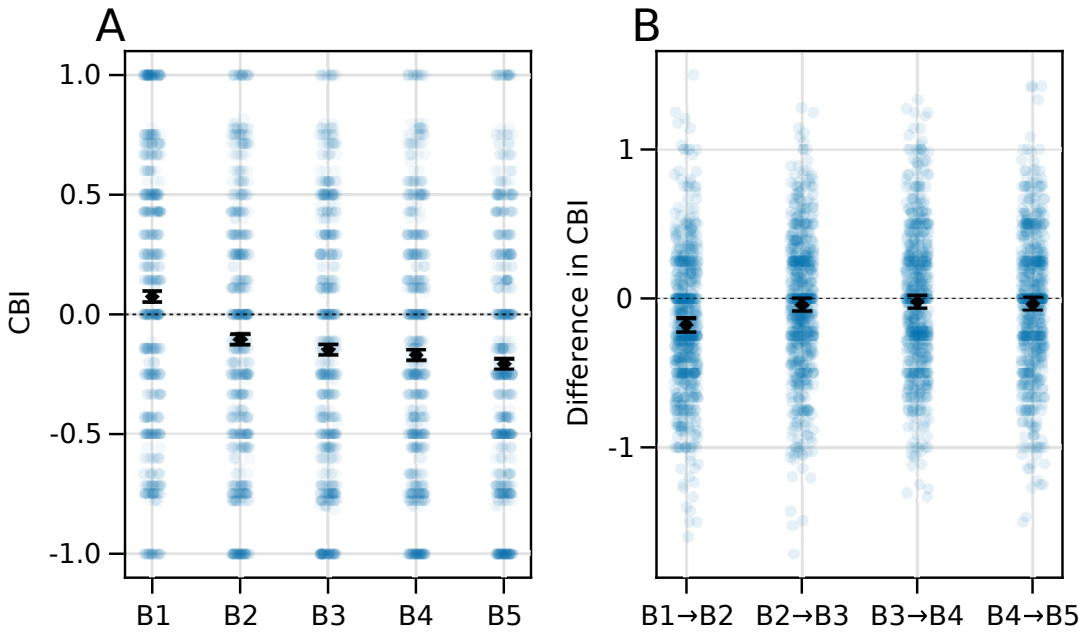


Figure 4.2: **Cognitive Bias Index changes over trials : A**: Trials of baseline test sessions were split into five blocks (B1-B5) and the Cognitive Bias Index (CBI) was calculated for all trials within each block. The CBI exhibited a shift towards negative values as trials progressed. **B**: Difference in CBI between consecutive blocks per animal. The CBI decreased the most in early trials, between the first and second block. Colored circles are individual animals, black diamonds and whiskers represent the mean and SEM respectively. N=199.

overall in ambiguous trials.

The distributions of the conditional probability ratio for each scenario were spread around the value of 1, with most of their density within the $[0, 2]$ interval. Particularly when the previous trial resulted in no reward, the distributions were more heavy-tailed. Specifically when an animal first chose a high-reward action (HA), received no reward ($A_{-1} = HA, R_{-1} = 0$) and then switched to a low-reward action (LA), the mode of the distribution was shifted towards more positive values. This shift suggested that the animals were more likely to switch to LA after an unrewarded HA, compared to how often they chose LA overall. In the same case of previous action and outcome ($A_{-1} = HA, R_{-1} = 0$), the distribution of ratio values for choosing the HA again was bimodal, with more values close to zero. This indicated that some animals would never repeat a HA that was unrewarded on the most recent trial. This result was specific to negative feedback on the HA. The distribution of choosing the LA again, after it previously resulted in no reward ($A_{-1} = LA, R_{-1} = 0$) was more concentrated around one, suggesting that animals

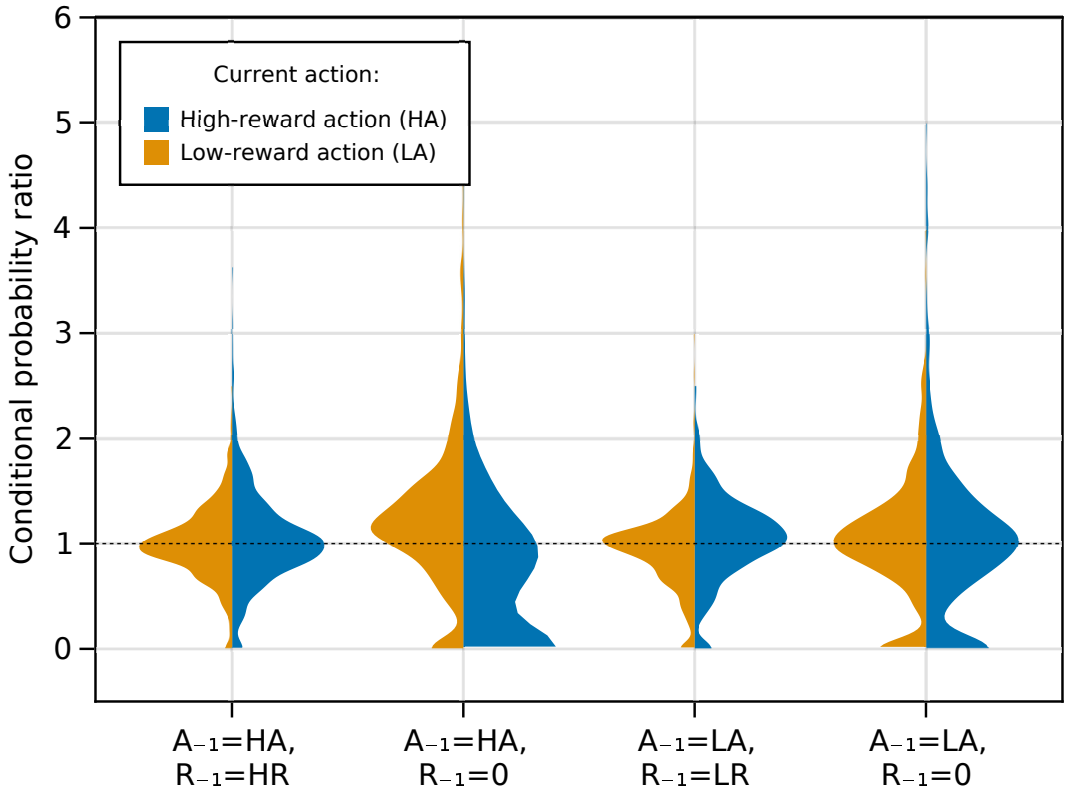


Figure 4.3: **The effect of the previous action and outcome on the current action during ambiguous trials:** The dependence of an action during an ambiguous cue presentation on the most recent action and outcome was assessed for each possible combination of previous action and outcome and current action. The independence probability ratio was calculated as the ratio of the probability of taking an action given a current ambiguous trial and a previous action and outcome over the probability of taking the same current action given only a current ambiguous trial. A value of the ratio equal to 1 means that the current action does not depend on the previous action and outcome during ambiguous trials. Values greater than 1 mean that it was more likely for an animal to choose a current action after a specific previous action-outcome case than it chose the same action on average and conversely for values smaller than 1. The ratio was calculated for each animal and its distribution is presented for each case. Even though the mode of the ratio distribution lied near the value of 1 in most cases, it spread on either side of 1, with longer tails in cases when the previous outcome was no reward. A_{-1}: action on the previous trial, R_{-1}: outcome on the previous trial, HA: High-reward action, LA: Low-reward action, HR: High reward (4 pellets), LR: Low reward (1 pellet). N=199.

were less affected by the omission of the low reward.

4.3.2 Model comparison

The results of the model comparison using the PSIS-LOOIC are summarised in Table 4.1. The data used for comparison included two baseline test sessions, a ketamine (1.0 mg/kg) session and its vehicle control session. The comparison was performed separately for the two populations of animals, corresponding to the original ketamine studies and the replication attempts.

The learning model had higher predictive accuracy of the held-out data, indicated by the higher LOOIC value. The difference between the LOOIC of both models including its standard error was 8.44 ± 5.13 for the original studies and 8.85 ± 5.7 for the replication studies. Both difference values including the standard error did not overlap zero. Therefore the learning model was given a higher probability of being the model that animals employed when selecting an action during ambiguous trials. Finally the diagnostic check for the PSIS approximation indicated a good fit of the Pareto distribution, thus validating the comparison.

Original studies

model	LOOIC	SE	ΔLOOIC	ΔSE	weight
reward rate	-2989.55	26.2	8.44	5.13	0.18
learning	-2989.55	26.06	0	0	0.82

Replication studies

model	LOOIC	SE	ΔLOOIC	ΔSE	weight
reward rate	-4585.91	40.35	8.85	5.7	0.22
learning	-4577.05	40.52	0	0	0.78

Table 4.1: **Model comparison using LOOIC :** LOOIC: the sum of the logarithm of the pointwise predictive density values for each held-out data point, that is the probability density of a data point given a posterior distribution of the model parameters after observing all other data points, SE: standard error around the logarithm of the pointwise predictive density values for each held-out data point, ΔLOOIC : difference in LOOIC between the current model and the model with the highest LOOIC, ΔSE: standard error of the difference in LOOIC, weight: the probability of the model being the best model to predict the data among the considered models. The top table used data from the original studies (N=31), where there was a main effect of ketamine on CBI towards more positive values. The bottom table corresponded to model comparison after inference was conducted on data from the replication studies (N=52), where the effect of ketamine was not replicated.

4.3.3 Model-based analysis of baseline sessions and a ketamine study

Posterior distribution results are presented for the learning model, as model comparison suggested that it performed much better at predicting held-out trial data, compared to the reward rate model. The posterior distributions of population-level mean parameters is shown in Figure 4.4. The animals were divided into two populations; one corresponded to the original ketamine studies were a significant positive effect of ketamine on CBI was observed and the second one included the rats from the replication studies. Moreover the rats of the original and the replication studies were supplied by two distinct breeding facilities.

The bias term μ (Figure 4.4A), which encompassed both perceptual and interpretation biases, was comparable between populations for baseline sessions. Positive bias values describe a bias towards choosing the LA. The vehicle condition led to an increase in bias only for the population of the replication studies. During the ketamine condition, the bias of both populations was comparable.

The population-level effect of the previous trial (Figure 4.4B), denoted by ξ, was comparable for the two populations across conditions. Positive values of ξ correspond to a *win-stay, lose-shift* strategy. The posterior distributions indicate that ξ was very likely to be non-zero, particularly for the animals of the replication studies.

The effect of learning (Figure 4.4C) was comparable across all three conditions of the two populations. Particularly at the baseline condition, this mean parameter was most likely to be non-zero. Smaller effects were also found for the vehicle and the ketamine conditions. Negative values of η indicate that a higher average reward on past ambiguous trials made animals more likely to choose the HA on the current ambiguous trial.

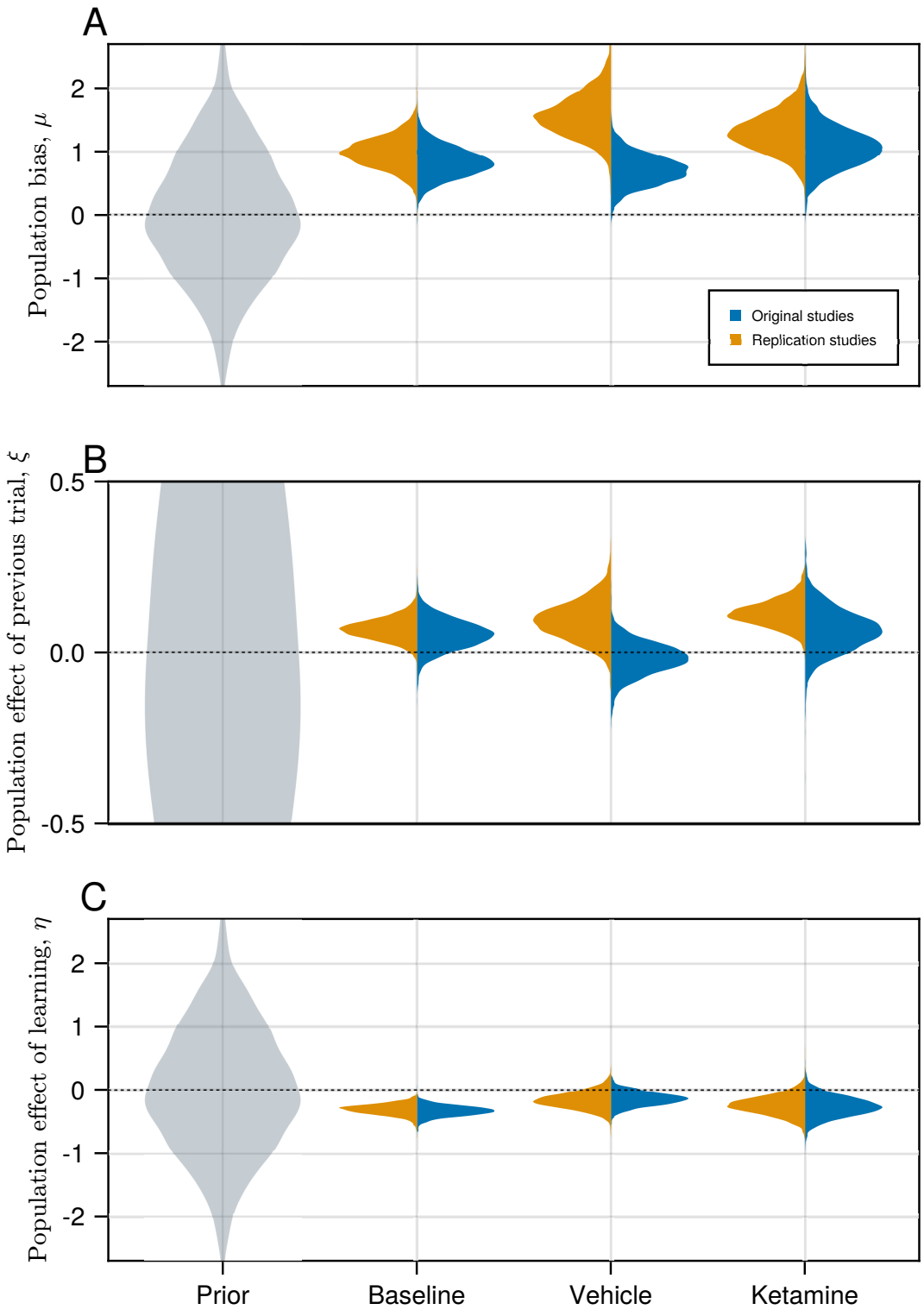


Figure 4.4: **Posterior distributions of population-level mean parameters for baseline sessions and a ketamine study:** Each effect considered by the learning model is shown as the posterior distribution of its population-level mean parameter across all conditions of the two populations. (Continued on the following page)

Figure 4.4: (Continued) The populations corresponded to the original studies, where ketamine caused a significant positive shift of CBI and the replication studies, where there was no effect. **A**: The combined perceptual and interpretation bias parameter. Positive values correspond to a bias towards the low-reward action. **B**: The effect of the previous trial, as a win-stay, lose-shift strategy. Note the different y-axis scale. The prior for this effect is the same as the other two plots of the figure, however the scale was shrunk in order to better visualise the posterior distributions. **C**: The mean effect of learning for the two populations. This was the weight that animals put on the reward of ambiguous trials, averaged over the most recent 12 trials. The grey-coloured distribution is the prior distribution for the respective parameter. Ketamine dose volume was 1.0 mg/kg). Original studies: N=31, Replication studies : N=52.

The effect of ketamine on each population-level mean parameter of the learning model is summarised in Figure 4.5. The depicted distributions are the posterior of the difference between each parameter during the ketamine condition and the same parameter during vehicle treatment. Ketamine had opposite effects on the bias term μ (Figure 4.5A); it shifted the bias towards the HA in the replication studies and it increased the bias towards LA in the original studies. These effects were likely to be non-zero, as indicated by the P_* values ($P_* > 0.8$ for both studies). Ketamine led to a positive shift in the *win-stay, lose-shift* parameter ξ of the animals of the original studies (Figure 4.5B) and a negative shift in the learning parameter η (Figure 4.5C). However the probabilities of a non-zero effect P_* for these two effects, as shown in the figure label, did not indicate a very likely effect.

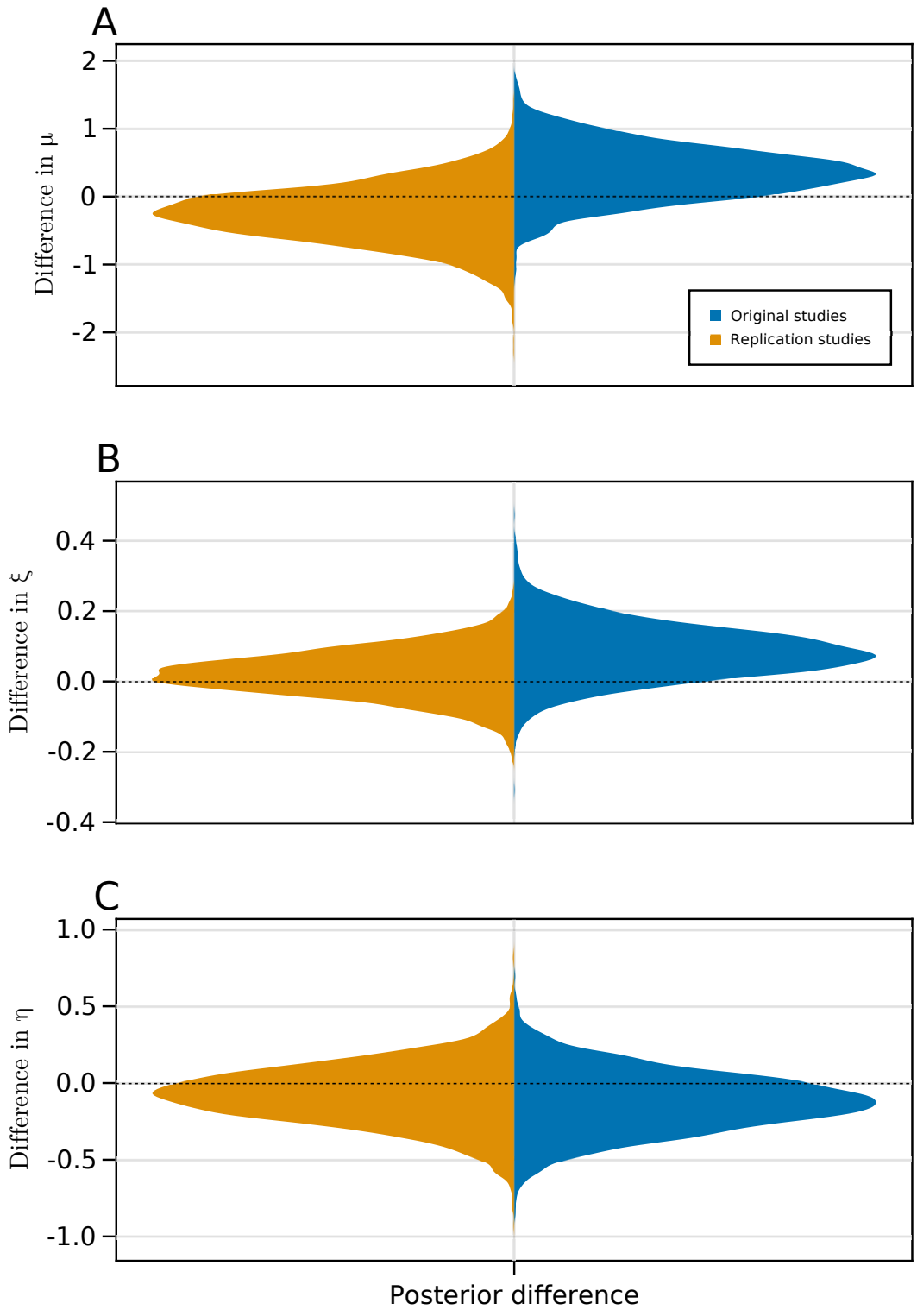


Figure 4.5: **The effect of ketamine on the original and the replication studies:** Population-level mean parameters for each one of the three considered factors are depicted as the posterior of the difference between the ketamine and the vehicle conditions. (Continued on the following page)

Figure 4.5: (Continued) **A**: Ketamine caused opposite effects on the combined perceptual and interpretation bias parameter. During the original studies, ketamine led to a shift in bias towards the low-reward action ($P_* = 0.857$). Conversely, in the replication studies, ketamine caused a shift towards the high-reward action ($P_* = 0.83$). **B**: The effect of the feedback of the most recent trial was slightly increased by ketamine in the original studies. However, a difference of zero has considerable probability mass around it, thus a clear positive change is uncertain ($P_* = 0.4$). In the replication studies, it is more certain that ξ did not change under ketamine ($P_* = 0.15$). **C**: Ketamine did not cause changes to the effect of learning with high certainty for either population of animals (Original studies: $P_* = 0.666$, Replication studies: $P_* = 0.616$). Ketamine dose volume was 1.0 mg/kg). Original studies: N=31, Replication studies : N=52.

Subject-level parameters for each of the factors included in the learning model are shown in Figure 4.6. The expected value for each parameter was plotted in place of its posterior distribution to simplify the visualisation. This value was calculated as the mean over all samples from the four sampling chains of the corresponding parameter.

Animals of the replication studies appear to have greater inter-individual variance in the bias parameter b, compared to the animals of the original studies (Figure 4.6A). However, ketamine reduced the variability for the *win-stay, lose-shift* effect w (Figure 4.6B), specifically for the animals of the replication studies. The subject-level effect of learning, r, was comparable between the two populations (Figure 4.6C). Most animals had a negative expected value or r, suggesting that they were more likely to choose HA after a greater average reward from past ambiguous trials.

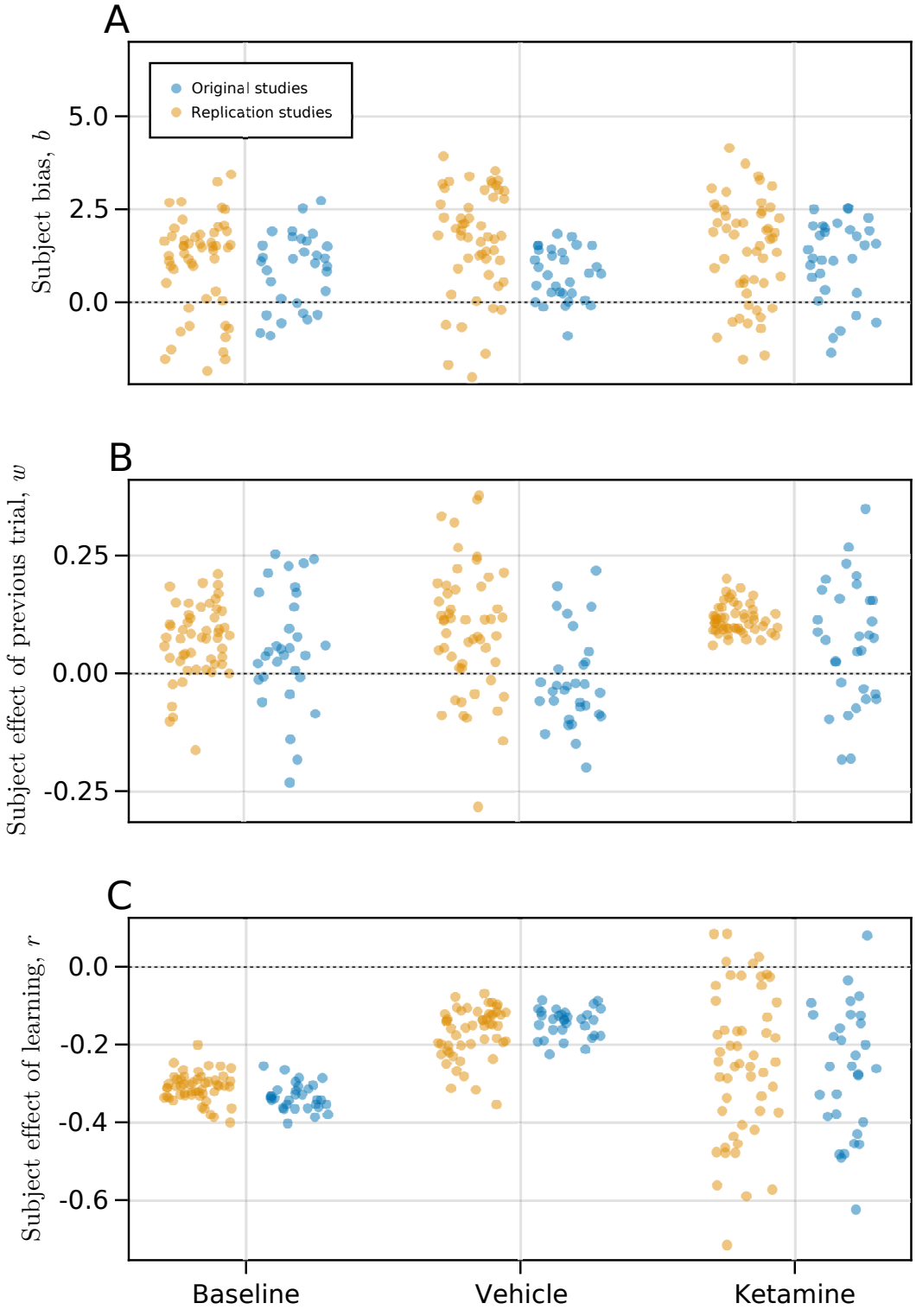


Figure 4.6: **Posterior distributions of subject-level parameters for baseline sessions and a ketamine study:** The expected value of each subject-level parameter is presented for all animals of the two populations. (Continued on the following page)

Figure 4.6: (Continued) The populations corresponded to the original studies, where ketamine caused a significant positive shift of CBI and the replication studies, where there was no effect. **A**: The combined perceptual and interpretation bias parameter. Positive values correspond to a bias towards the low-reward action. **B**: The parameter w corresponded to the effect of the previous trial, as a *win-stay, lose-shift* strategy. **C**: The effect of learning r was the weight that animals put on the reward of ambiguous trials, averaged over the most recent 12 trials. Each circle corresponds to the expected value of the respective parameter for an individual animal. Ketamine dose volume was 1.0 mg/kg). Original studies: N=31, Replication studies : N=52.

Finally, the lapse rates of individual subjects are summarised in Figure 4.7. The prior distribution of Figure 4.1 is shown as well for comparison. It is evident that high values of lapse rate were more likely in the posterior, after the inference process on either population of animals, compared to the chosen prior distribution.

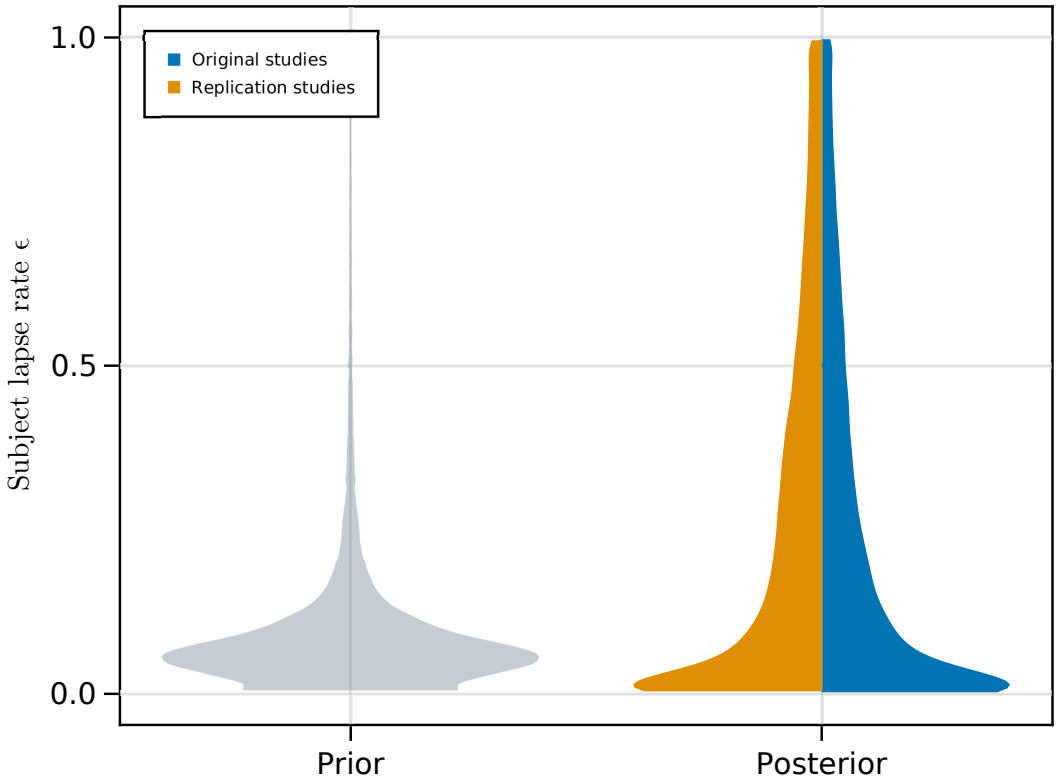


Figure 4.7: **Posterior distribution of subject-level lapse rates for baseline sessions and a ketamine study:** Lapse rates for individual animals of the two populations were calculated using the transformation $\Phi(\kappa + \bar{\epsilon}_s\omega)$ after sampling κ, ω and $\bar{\epsilon}_s$ for each subject s from their respective posterior distributions. The populations corresponded to the original studies, where ketamine caused a significant positive shift of CBI and the replication studies, where there was no effect. The posterior distributions of individual animals within a population were collated to produce a single distribution of subject-level lapse rate values. This was done so that the posterior would be directly comparable to the prior. The prior distribution is also shown in grey. The posterior distribution of both populations had a heavier tail towards larger lapse rate values, even though these values were not initially expected, as shown by the thinner tail of the prior. Individual animals' lapse rates were assumed to be sampled out of a common distribution for the baseline, vehicle and ketamine (1.0 mg/kg) conditions. Original studies: N=31, Replication studies : N=52.

4.3.4 Model-based analysis of an amphetamine study

Following the inference on baseline sessions and the ketamine study, both models were fitted on the amphetamine (0.3 mg/kg) study of cohort 1, presented in chapter 2. The model comparison in this case resulted in very similar differences between the models as Table 4.1, with the learning model being the best candidate model for the action selection process. Thus, only the posterior of the learning model is presented for the amphetamine study.

The effect of amphetamine on the population-level parameters was different from that of ketamine. Figure 4.8 contains a summary of the posterior distribution of these parameters. Amphetamine mitigated the bias towards the LA, setting the value of μ closer to zero (Figure 4.8B, $P_* = 0.974$). The drug did not influence the effect of the previous trial ξ with enough certainty ($P_* = 0.395$), suggesting no alterations to a potential *win-stay, lose-shift* strategy (Figure 4.8D). The effect of learning, η, was moved closer to zero under amphetamine, with a probability $P_* = 0.77$ of a non-negligible effect (Figure 4.8F).

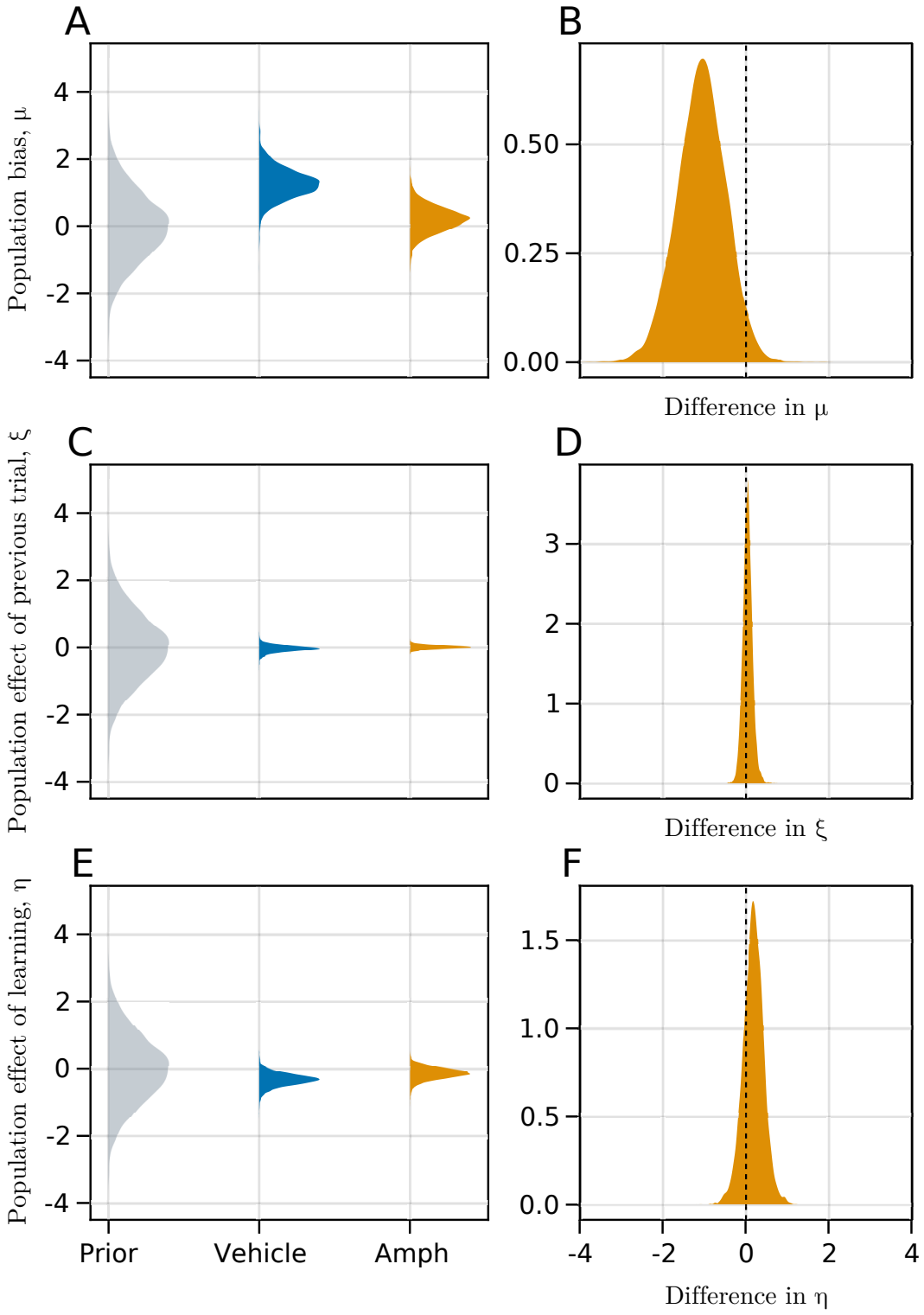


Figure 4.8: **Posterior distributions of population-level mean parameters for an amphetamine study :** Each effect considered by the learning model is shown as the posterior distribution of its population-level mean parameter across conditions and the difference between the amphetamine and vehicle conditions. (Continued on the following page)

Figure 4.8: (Continued) **A**: The combined perceptual and interpretation bias parameter. Positive values correspond to a bias towards the low-reward action. **B**: There was a considerable shift in bias towards the high-reward action under amphetamine ($P_* = 0.974$). **C**: The effect of the previous trial, as a *win-stay, lose-shift* strategy. **D**: Under amphetamine its value was concentrated around zero with only a small amount of variance, making the absence of an effect likely ($P_* = 0.395$). **E**: The mean effect of learning for the population. This was the weight that animals put on the reward of ambiguous trials, averaged over the most recent 12 trials. **F**: Amphetamine did not alter the effect of learning with high certainty ($P_* = 0.77$). Grey-coloured distribution in **A**,**C**,**E** is the prior distribution for the respective parameter. Dashed vertical lines at zero in **B**,**D**,**F**. Amph: Amphetamine (0.3 mg/kg). N=15.

The posterior distribution of individual animals' lapse rates had a heavier tail towards larger values than in the prior distribution (Figure 4.9). This mismatch between the prior and posterior distributions was similar to that observed in Figure 4.7 for the case of baseline sessions and the ketamine study of the previous section.

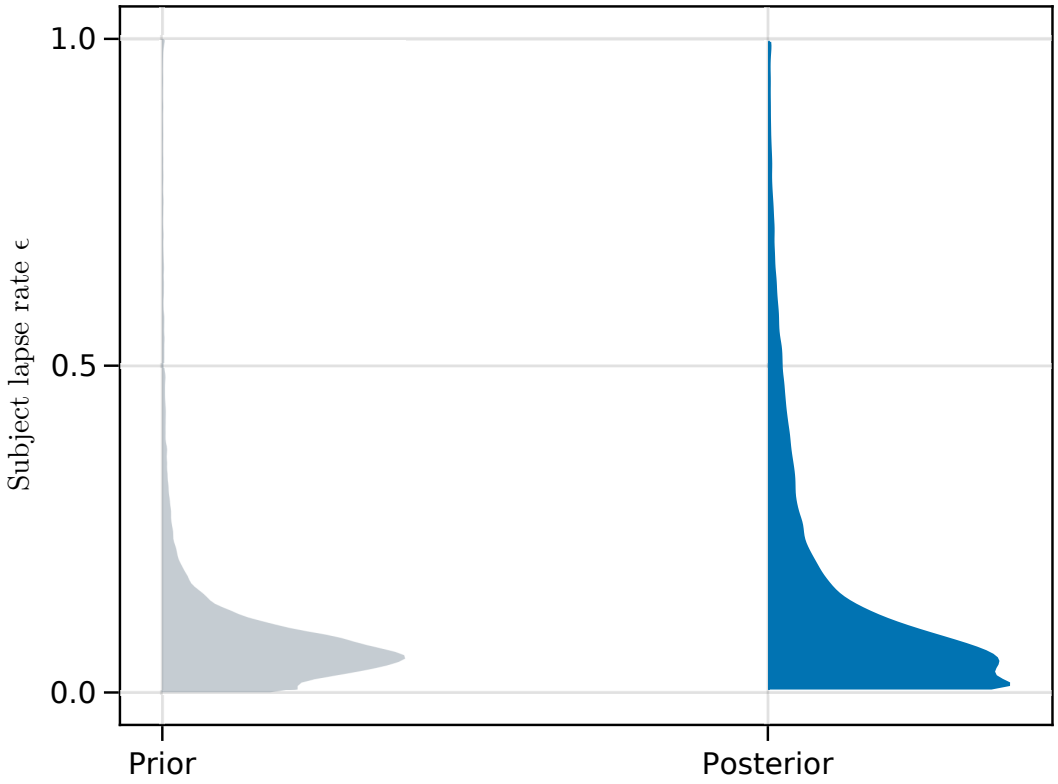


Figure 4.9: **Posterior distribution of subject-level lapse rates for an amphetamine study:** Lapse rates for individual animals were calculated using the transformation $\Phi(\kappa + \bar{\epsilon}_s\omega)$ after sampling κ, ω and $\bar{\epsilon}_s$ for each subject s from their respective posterior distributions. The posterior distributions of individual animals were collated to produce a single distribution of the subject-level lapse rate values. This was done so that the posterior would be directly comparable to the prior. The prior distribution is also shown in grey. The posterior distribution of both populations had a heavier tail towards larger lapse rate values, even though these values were not initially expected, as shown by the thinner tail of the prior. Individual animals' lapse rates were assumed to be sampled out of a common distribution for the vehicle and amphetamine (0.3 mg/kg) conditions.

4.4 Discussion

The JBT was designed as a decision-making task, where the behaviour at each trial was assumed to be independent of the trial history. However, visualisation of the animals' CBI across trials (Figure 4.2) and of the effect of the most recent trial (Figure 4.3), which included multiple animal cohorts, suggested that this assumption was not entirely accurate. A model-based analysis of the animals' actions was employed in order to account for such effects of past reinforcement. This analysis

revealed differences between two populations of animals, each one originating in a different breeding facility. Moreover, differences were observed in the effects of ketamine and amphetamine on the animals' responses to the ambiguous cue.

The bias parameter of the learning model revealed the most striking difference between the two populations of rats (Figure 4.4A). Even though the bias during baseline sessions was comparable between populations, the vehicle condition led to a shift in bias towards the LA for the animals of the replication studies. Previous work on a similar rat strain indicated that handling the animals during their development led to decreased anxiety-like behaviour in an open field and a novelty-suppressed feeding test, compared to unhandled animals (Caldji et al. 2000). The behaviour of the unhandled animals of this study was comparable to the behaviour of animals that had undergone maternal separation, early in their development. Although information about handling conditions or the weaning period in either breeding facility was not available, differences in handling could potentially explain this difference in bias. Moreover, it has been shown that maternal separation can cause early life adversity and consequently a heightened sensitivity to stressful conditions (Stuart et al. 2019). For instance, the elevated bias towards the least rewarding action in the animals of the replication studies could have been triggered by a more aversive response to the injection if these animals were less handled in their breeding facility and hence more anxious. Even though the bias parameter was a combination of both perceptual and interpretation biases, it is unlikely that the two rat populations differed in their perception of the ambiguous cue, as their bias values during baseline sessions were comparable (Figure 4.4A).

In the population of the original studies, ketamine led to a small increase in bias towards the LA (Figure 4.5A). A similar effect has been observed, when a subset of this population was used to fit the parameters of a drift-diffusion model (DDM) (Hales et al. 2017). The decision starting point, corresponding to the initial bias of the animals prior to any sensory evidence accumulation, was shifted closer to an LA decision under the drug. However, in the current analysis ketamine caused a small shift to the opposite direction for the population of the replication studies. If the above assumption about a more aversive response to the injection is true, then this effect of ketamine could be akin to the efficacy of the drug to mitigate negative biases (Stuart et al. 2015b; Carboni et al. 2021). A corollary of this hypothesis is that the animals need to first be at a negative affective state in order for the drug to reduce the associated negative bias.

Amphetamine caused a greater shift in bias towards zero, thus mitigating a bias towards the LA that was present during the vehicle session (Figure 4.8A). There is no evidence for disruptions in perception under amphetamine, thus its effect on bias was most likely a change in interpretation bias. In a different decision-making task with probabilistic rewards, rats showed a similar bias towards the more frequently rewarded action under amphetamine and when ambiguous stimuli were presented (Der-Avakian et al. 2013). Similar effects have been observed in human studies that pharmacologically increase dopamine, for example using the L-DOPA drug (Sharot et al. 2012). In the human experiments, dopaminergic drugs induce an "optimism bias", as participants tend to persevere on previously learned best actions, even after negative feedback.

Interestingly though, cocaine, which increases the levels of dopamine, did not lead to a significant change in CBI in the same JBT variant, as amphetamine did (Hales et al. 2017). Amphetamine increases the release of all monoamine neurotransmitters, so its effect could be further mediated via noradrenaline or serotonin. Data from the cocaine study can be used to fit the current learning model and compare the results with those of amphetamine to evaluate the differences between the two treatments in more detail.

Apart from their difference in bias, the posterior distributions of the other two model parameters, corresponding to the effects of the immediate and the longer-term past, were comparable for the two populations. The mean parameters ξ and η were likely to be nonzero in most cases (Figure 4.4B and C). This result supported the suggestion from the visualisations of CBI (Figure 4.2) and the conditional probability ratio (Figure 4.3) that past reinforcement was affecting the animals' actions during ambiguous trials.

Sensitivity to recent feedback has also been observed in rats performing a probabilistic reversal learning task (PRLT) (Noworyta-Sokolowska et al. 2019). This work observed inter-individual differences in sensitivity to the feedback of the previous trial, while the sensitivity of each rat was stable across time. Besides the spread of the conditional probability ratio distributions (Figure 4.3), our inference results support this claim as well. The expected values of the w parameter, corresponding to the effect of the most recent trial, for each rat covered both positive values and the region around zero (Figure 4.6B). This result indicates that some rats were sensitive to the most recent feedback and employed a *win-stay, lose-shift* strategy, while others did not exhibit this sensitivity.

On the population-level, the mean effect of the most recent trial, ξ, had a positive value for baseline sessions, the ketamine condition and its matching vehicle treatment, suggesting that a *win-stay, lose-shift* strategy was likely to be part of the animals' behaviour (Figure 4.4B). A similar effect of the previous outcome was also observed in a group of rats during a JBT task that involved both punishments and rewards (Neville et al. 2020b). Inference on the amphetamine study resulted in a distribution for ξ closer to zero (Figure 4.4B), however this inference had only a limited number of animals to inform the population-level parameters.

The dependency of actions on the most recent trial varied across animals and was particularly pronounced in in cases of negative feedback and even more so when the large reward was omitted (Figure 4.3). Assuming that the animals were holding some expectation for reward when they were choosing a lever, this observation could be interpreted as a disappointment-like effect due to the omission of reward.

The shift in CBI towards more negative values over trials (Figure 4.2), suggested that there was an effect of the longer-term past, which could not be captured by the sensitivity to the previous trial. Some animals showed an increase in CBI over trial blocks, while some others were choosing the LA more often as the session progressed (Figure 4.2B). These changes to CBI for individual animals spanned the entire range from -1 to $+1$. Thus, it was unlikely that this effect was driven by a factor related to the passage of time, for example satiety. I proposed that reinforcement history could be responsible, since one third of the trials were ambiguous ones and these trials were reinforced only half of the time with a reward that depended on the action taken. Model comparison indicated that the learning model, which included an average over the outcomes of past ambiguous trials within a time window, was much more likely to explain the animals' behaviour in both populations, compared to a reward rate model, which included an average over the outcomes of all trials within the same window.

Ketamine increased the variability of the subject-level learning parameter, r, compared to both the baseline and the vehicle conditions (Figure 4.6B). One interpretation is that ketamine causes a disruption to working memory, that is the capacity to temporarily store and manipulate information (Baddeley 2007). Even a smaller dose volume of the drug (0.4 mg/kg) has caused behavioural deficits in macaques during a visual working memory task, while it increased the variability of single-neuron firing within the lateral prefrontal cortex (Ma et al. 2015). Thus, it was possible that rewards further into the past, that contributed to the calculation of the average

reward, were not processed as reliably as during the baseline and vehicle conditions.

Previous model-based analysis of JBT experiments in humans with rewards and punishments as the potential outcomes found an effect of the average reward of all past trials on the participants' actions during ambiguous trials (Neville et al. 2021). However, this study did not directly compare the model that incorporated this average reward with one using the average of previous ambiguous trials, to test for learning-specific effects. Additionally the average reward acquired did not correlate with the participants' self-reported affect scores. The learning model of the current chapter could be fitted to the data from this human study and compared with the reward rate model in order to evaluate whether the current results are also found in humans and when punishments are present. Moreover, the difference between the reward of an ambiguous trial and the latest average estimate could be used as a prediction error, as it has been used in other studies, where it was successfully correlated with affective state (Rutledge et al. 2014; Blain et al. 2020).

Inference on the lapse rates of individual animals yielded partially unexpected outcomes. Even though most animals' probability of acting at random was concentrated between 0 and 0.1, there was a long tail and a considerable amount of probability mass in higher values. This result was common between the inferred lapse parameters of the baseline sessions and the ketamine study (Figure 4.7) and those of the amphetamine study (Figure 4.9). This could indicate an issue of unidentifiable parameters in the model. If all other subject-level parameters are close to zero, then an animal would select an action with probability 0.5, regardless of the value of the lapse rate. This is a common issue in behavioural models, particularly when the parameters are selected to be interpretable (Wilson et al. 2019). A parameter recovery study should thus be performed, where model parameters are sampled from their respective prior distributions in order to generate fictive choice data, which in turn will be used to fit the model and verify whether the original parameter values were recovered (Gelman et al. 2020).

Besides the potential methodological limitation, an elevated lapse rate could reflect exploratory behaviour rather than inattention, as it was shown in a perceptual decision-making task in rats (Pisupati et al. 2021). A lower average reward of the recent past could lead rats to use a more exploratory behaviour, particularly to the unfamiliar ambiguous cue. Examples of such behaviour include downshifts in reward value during a radial maze task in rats (Pecoraro et al. 1999) and a foraging task in humans (van Dooren et al. 2021). Recent model-based inference on a perceptual

task in mice proposed that the animals can behave according to multiple strategies and switch between them within a session, with exploratory, lapse-like strategies following bouts of biased responding (Ashwood et al. 2022). A similar behavioural motif could be underlying the rats' behaviour during JBT, where the animals switch to an exploratory strategy, for instance after a low average reward of the recent past. Incorporating multiple strategies within the same model, along with assumptions about the structure of transitions between strategies could elucidate the animals' behaviour even further. It could also circumvent potential issues with unidentifiable parameters, as only a subset of factors would be part of each strategy.

The current work provided evidence about learning from past reinforcement being an important factor in modulating responses to the ambiguous cue. This result justifies the suggestions of chapter 2 about introducing probabilistic rewards and reducing the number of ambiguous trials in the session (see also Roelofs et al. 2016 for similar suggestions for JBT task design). Moreover, the learning model could be expanded to include running estimates of the value of each lever-press action using the reinforcement learning framework (Sutton et al. 2020).

By extending the learning model to incorporate reinforcement learning rules (Sutton et al. 2020), the effects of past reinforcement can be described in more detail. For instance, a learning rate parameter provides temporal discounting, so that the weight that animals put on past rewards decays exponentially. This is a less limiting assumption than an average over a time window of 12 trials, $\hat{R}$, being multiplied by r, as the time window of the considered rewards depends on the learning rate, which is a model parameter to be inferred. Additionally, learning from positive and negative feedback could be separated by distinct learning rates. The conditional probability ratio (Figure 4.3) of cases of negative feedback was more variable across animals, whereas the positive feedback cases were more concentrated near the value of 1. This suggests that the previous feedback was more influential in cases of negative feedback. Previous studies on PRLT found that rats were less sensitive specifically to negative feedback under ketamine (Rychlik et al. 2017; Wilkinson et al. 2020). Thus separating the effects of wins and losses in conjunction with temporal discounting of the past could further elucidate the effects of the drug.

Finally, past model-based inference on the present version of JBT utilised a DDM, informed by both choices and response time distributions (Hales et al. 2016; Hales et al. 2017). These studies revealed differences between affective state manipulations, even when the CBI scores of the respective experiments were similar. However,

these studies assumed stationarity of the decision-making process across trials. In this chapter, I have argued that the past is an important determinant of the animals' actions. The generative model of choices presented here could be augmented by a generative model of response times, similar to the model in chapter 3 or a more theoretically inspired model that combines the evidence accumulation of the DDM with the reinforcement learning mechanisms discussed above (Fontanesi et al. 2019), in order to make inference more informative.

Chapter 5

Affective state in learning: a theoretical model & naturalistic simulation environments

5.1 Introduction

Typically in reinforcement learning (Sutton et al. 2020) - whether in a learning task for an animal or human participant or in a simulated task for an artificial agent - a single action leads to a matching outcome. This is the assumption which underpins a large number of successful reinforcement learning models such as Q-learning (Watkins et al. 1992) and temporal-difference learning (Montague et al. 1996) in explaining behavioural and neural data from experimental studies (Niv 2009 for a review).

Though common in laboratory experiments and simulations, temporally strict action-outcome contingencies are rarely encountered in the real world. An agent could perform several actions before receiving any reward from their environment, while actions might vary in terms of how long a reward caused by them is delayed. Therefore, attributing causes to outcomes is a key problem in learning. The implications of this mismatch between contrived and natural environments in mental disorder research have been discussed in a recent review (Scholl et al. 2018). More naturalistic tasks could elucidate learning and decision-making deficits that have been observed in patients. By emulating features of the real world in a task, subjects could exhibit behaviour that is closer to their everyday actions and processing of outcomes.

Thus, the observed behaviour would be a more representative example of the way patients interact with their environment. Consequently, computational models, able to process features of real environments, would be beneficial when doing inference on patient groups.

Two experiments in macaques with both instructed and uninstructed delays between actions and their causal rewards have measured correlations between actions and rewards that were not causally linked (Jocham et al. 2016). The results were in agreement with early observations of outcomes reinforcing non-causal actions, which occurred in temporal proximity to the causal action (Thorndike 1933). Importantly, in this experiment, actions were followed by immediate outcome feedback, yet the same outcome affected actions that occurred several trials both before and after subjects received it, resulting in higher repetition of false actions in future trials. A reward schedule task that required a sequence of correctly timed actions before the delivery of a reward revealed influences of the macaque's position in the sequence on error rates, that could not have been accounted for by the temporal distance between action and reward (Bowman et al. 1996).

Seemingly unrelated actions and outcomes have been shown to influence the behaviour of rats, even when subjects were not explicitly trained to associate them. These studies used multimodal outcomes, including primary reinforcers and affective state manipulations. The Affective Bias Test (ABT) was introduced as a behavioural assay in rats to test whether subjects have differential preference between two actions, when both actions result in the same learned reward but one of them had been paired with a manipulation on affective state during learning (Stuart et al. 2013). Manipulations of either positive or negative valence were equally effective in leading to an action bias, whether they were administered before or after the learning sessions. Another ABT study found that an anxiogenic drug, adminsitered shortly before a learning session, led to a negative action bias, while this bias was mitigated when a positive affective manipulation, in the form of social play, took place several hours after the learning session on the same day (Hinchcliffe et al. 2022). Both ABT studies involved manipulations that biased animals' choices during test day, even when they occurred several hours after a learning session. This observation is indicative of a memory consolidation process, which combines all experiences of the same day. Salient experiences in this case include learning action-reward contingencies during ABT sessions and being in different affective states during and after manipulations.

We propose here that seemingly undesirable influences of outcomes upon actions that are not causally linked with them could be beneficial in natural environments, where actions and outcomes do not obey a strict temporal order. In such cases, we claim that an agent would still seek to learn associations between its actions and rewards and we investigate how this challenge might be attainable when a model of the environment is unknown. Without structural knowledge of its environment, we suggested that an agent's affective state is an important component of its learning process. Affective state works as a contextual cue, aiding the agent as it is looking back into past experiences that have resulted in its current state, in an effort to account for it. The agent model we proposed, named Delta agent, could be seen as a functional account of affective state in reinforcement learning environments where actions lead to delayed outcomes.

We have augmented classical action-value learning models with two computational components that have been supported by experimental evidence, and expand on them in order to account for the novel observations in the ABT studies. These components are a leaky integrator of reward prediction errors (RPEs) and experience-based priority sampling during an offline replay phase. The Delta agent updates its expected reward prediction error (RPE) across all experiences of an episode by using a leaky integrator of RPEs, which we have called the episode's residue. Online the agent learns the immediate value of its actions, while offline it updates values, representing the expected, long-term RPE that the respective action will cause. The offline values correspond to the affective biases of the original ABT studies, as the online values of the two rewarded substrates are the same and represent the immediate pellet reward.

While it might seem detrimental for a reward-maximising agent to conflate actions and outcomes that are not necessarily causally linked, we argue that our offline learning rule is beneficial in environments where contingencies are unknown and do not conform to a trial-based structure of immediate rewards after actions within a trial. Building upon standard ten-bandit problems, we designed a number of novel tasks in order to capture features of naturalistic action-reward contingencies that are not immediate in (discretised) time. Two types of tasks were created: the Deferred Bandits (DB) and the Tardy Bandits (TB) tasks. Both environments include immediate and delayed outcomes, while differing in the generative processes that deliver both types of outcomes.

A comparison of the Delta agent was performed against the Monte Carlo agent

model (Sutton et al. 2020, section 5.4), which learns exclusively offline and holds all information about future rewards following every action. The Monte Carlo model was chosen as an ideal case of a tabular agent, since it learns the future gain after an action, including rewards that are delayed by any number of timesteps within an episode. Additionally, a version of the Delta agent without the leaky integrator and the offline sample-based learning was implemented as a way to assess the contribution of these proposed components in the agent's performance.

5.2 Methods

5.2.1 Simulated environments

Deferred Bandits

The basic idea behind the DB task was that a reward was delivered immediately after its causal action, while a larger reward was deferred until the end of the episode. The large reward was a function of all previous actions within an episode. There were three variants of the DB task, which differed in the immediate reward, while they utilised the same deferred reward rule.

In the first variant (DB_1, Figure 5.1A), rewards during all timesteps but the last one were set to 0, whereas the final action delivered the deferred reward, R_d. In the other two environments, all actions led to either a binary (DB_2, Figure 5.1B) or a continuous (DB_3, Figure 5.1C) reward, while the deferred reward was added on the final timestep.

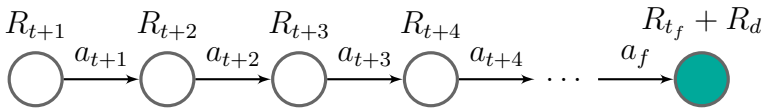
All three DB tasks modelled a "day and night" cycle; during the day there are multiple trials, but the substantial part of the reward for these trials was only received at night. In the simplest version, DB_1, this deferred reward was the only reward, so

$$R_t = \begin{cases} 0, & t \in [1, t_f - 1] \\ R_d, & t = t_f \end{cases} \tag{5.1}$$

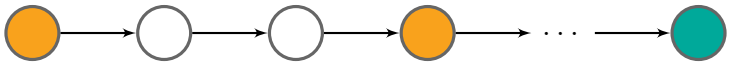
where the "night" trial was $t = t_f$ and the deferred reward was

$$R_d = 2 \sum_{a \in A} \rho_a I_a \tag{5.2}$$

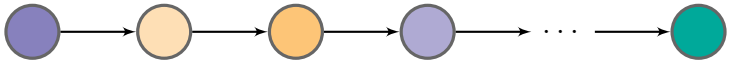
(A) **DB$_1$: no rewards**



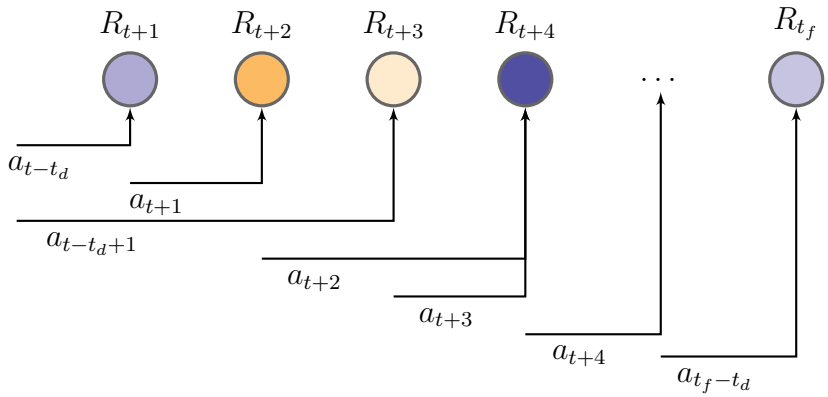
(B) **DB$_2$: binary rewards**



(C) **DB$_3$: continuous rewards**



(D) **TB**



Reward legend:

Figure 5.1: **Structure of the Deferred Bandits and the Tardy Bandits tasks**: In DB, **A**-**C**, rewards were sampled from different distributions and a deferred reward R_d was added to the reward of the final step of each episode. **A**: DB$_1$ generated no rewards during the first $t_f - 1$ timesteps, **B**: in DB$_2$ rewards were sampled from a Binary(0.5) distribution for the same timesteps, **C**: DB$_3$ used a Normal(0,1) distribution to generate them. **D**: In TB, rewards were sampled from Normal distributions with mean values dependent on the bandit index (1-10) and a common standard deviation of 1. They were delivered with a delay of t_d timesteps, which was sampled at each step from a Poisson distribution. The shading of blue and orange nodes marked the magnitude of the negative and positive outcome respectively. a_t: action taken at timestep t, R_t: reward received at timestep t

where

$$\rho_a = \frac{N_a}{t_f} \tag{5.3}$$

was the ratio of the number of times action a was chosen, N_a, over the total number of timesteps in an episode t_f and I_a was the bandit identity of the same action, which corresponded to a reward equal to i for the i^{th} bandit. The factor of 2 in equation 5.2 was applied to ensure that the deferred reward would be much larger than the immediate rewards, particularly in DB_2 and DB_3, where the immediate reward was nonzero.

The DB_2 and DB_3 variants included, in addition to the deferred reward R_d, an immediate reward $\hat{R}_t$. For DB_2 this was a random binary reward

$$R_t = \begin{cases} \hat{R}_t \sim \text{Bernoulli}(0.5), & t \in [1, t_f - 1] \\ \hat{R}_t + R_d, & t = t_f \end{cases} \tag{5.4}$$

whereas in DB_3 the immediate reward was sampled from a continous distribution:

$$R_t = \begin{cases} \hat{R}_t \sim \text{Normal}(0, 1), & t \in [1, t_f - 1] \\ \hat{R}_t + R_d, & t = t_f \end{cases} \tag{5.5}$$

Delayed Bandits

The second task was also an elaboration of the standard 10-arm bandit. This was the TB task (Figure 5.1D), where choosing a bandit resulted in a delayed reward. The delayed time of delivery t_d was being sampled from a Poisson distribution with an expected rate λ.

Reward magnitudes depended on the identity of the chosen bandit at each timestep, $I_{A(t)}$, which was the mean of a Normal distribution that generated them. Rewards from multiple actions were summed together when their sampled delays made them coincide at the same future timestep.

5.2.2 Delta agent

The proposed agent, which we called the Delta agent, learned in two phases. During the online phase, the model updated the action value Q_a, for an action a, taken at

timestep t, according to the delta rule

$$\delta_t = R_t - Q_{A(t),t} \tag{5.6}$$

$$Q_{a,t+1} = Q_{a,t} + \eta\delta_t, \qquad t \in [1, t_f], \quad a = A(t) \tag{5.7}$$

with learning rate η, $A(t)$ being an indicator function of the action taken at time t and t_f representing the final timestep of an episode. Q values represented the immediate value of taking an action.

The RPE δ_t was then being integrated to accumulate the episode's residue

$$\Delta_{t+1} = \Delta_t + \psi\left(\delta_t - b_{A(t),-} - \Delta_t\right) \tag{5.8}$$

Our choice for the episode's residue update rule was a recency-weighted average of RPEs relative to the initial bias value of the chosen action, $b_{A(t),-}$. The RPE, δ_t was compared to the bias of the chosen action $b_{A(t),-}$ during the update of equation 5.8. In section 5.2.2, it will be shown that bias learning driven by our episode's residue is equivalent to assigning a proportion of the integrated episode's RPE to bias values.

Leaky integrators of RPE values have been shown to correlate with mood in humans and influence decision-making in lottery tasks with no learning required (Eldar et al. 2015; Rutledge et al. 2014). In learning tasks, integrated RPEs were more important for predicting future choices of human subjects compared to reward magnitude (Blain et al. 2020). Here we treat affective state as a generalisation of mood for non-human species, in the sense that both concepts could be mapped on a 2-dimensional plane of valence and arousal (Russell 2003). Affective biases then represent the expected affective state that an action might lead the agent to experience. The values learned offline in our model corresponded to affective biases, as the ones observed in the original ABT studies.

Biases were only updated offline, during the replay phase, according to

$$b_{\mathrm{a},+} = b_{\mathrm{a},-} + \xi M_{\mathrm{a}} \Delta_{t_f}, \qquad 0 \leq M_{\mathrm{a}} \leq M \tag{5.9}$$

Thus, bias values were being updated by a fraction of the episode's residue, given a learning rate $0 \leq \xi \leq 1$ and M_{a} corresponding to the number of times action "a" was sampled during replay, out of a total of M replay samples. The definition of the episode's residue included the integrated value of past RPEs, along with the bias values of past actions. Given our assumption about integrated values of RPE

estimating the agent's current affective state, bias values could be interpreted as the expected affective state that the agent will experience after taking the corresponding action. The bias update was equivalent to an agent that tried to match its current value of integrated RPE, or affective state, to bias values of past actions by performing least-square regression. This equivalence will be shown in section 5.2.2.

During the offline phase of our agent, sample priority was given to less likely actions, with probabilities assigned by the agent's policy as

$$\boldsymbol{p} = \begin{bmatrix} 1 - \pi(A(t_1)) \\ 1 - \pi(A(t_2)) \\ \vdots \\ 1 - \pi(A(t_f)) \end{bmatrix} \tag{5.10}$$

where $\pi(a)$ was the probability of taking action a given the policy that the agent used when engaged with a task. This rule was inspired by the action typicality (or normality) bias, the elevated affective responses of humans after surprising events followed non-typical actions (Kahneman et al. 1986). The episode's residue was a measure of how much reward was not predicted by the agent; it therefore was likely that the events that contributed to the residue were ones that the agent had not encountered frequently enough in the past for them to have already formed part of the agent's model of the reward environment. Thus, unlikely actions were prioritised during the replay phase. During each one of the M steps of the replay phase an action is being sampled according to

$$\mathrm{a} \sim \mathrm{Categorical}(\boldsymbol{p}) \tag{5.11}$$

using the priority vector defined in equation 5.10. The final value $b_{a,+}$ for each action then became the value to be used during the next episode $b_{a,-}$. The episode index was dropped for simplicity of notation.

In the subsequent episode an agent would take actions according to its policy π by combining the learned online and offline values as

$$a \sim \mathrm{Categorical}(\pi(\boldsymbol{Q} + \boldsymbol{b}_-)) \tag{5.12}$$

where $\boldsymbol{Q}$ and $\boldsymbol{b}_-$ were the vectors of online and offline values respectively. By incorporating the bias values in the action selection process, the agent considered the expected RPE that it would experience as a result of the chosen action. Thus, the

agent was being influenced by estimates of its future affective state, when deciding how to act. The idea of expected RPE values driving action selection has been previously used to explain leave-stay behaviour during a foraging task in humans (Wittmann et al. 2016), where classical reinforcement learning models of pure reward learning failed. Furthermore, Mendl and Paul have presented a theoretical account of affect during decision making, which included, among other components, the expected affective state after an action influencing action selection (Mendl et al. 2020).

Equivalence between the offline update rule and least-square regression

During the replay phase, the agent looks back into its actions in an effort to account for its current affective state. This interpretation becomes more apparent by considering the similarity between the combined residue and bias update rules, equations 5.8 and 5.9 respectively, and the minimisation of the objective function

$$U = \left(\hat{\Delta}_{t_f} - \sum_{a \in \mathcal{A}} w_a b_a \right)^2 \tag{5.13}$$

by gradient descent. $\mathcal{A}$ was the set of all possible actions in the environment and $\hat{\Delta}_{t_f}$ is the integrated RPE values, δ_t, reflecting the agent's affective state at the end of an episode

$$\hat{\Delta}_{t+1} = \hat{\Delta}_t + \psi \left(\delta_t - \hat{\Delta}_t \right) \tag{5.14}$$

Minimising the objective function in equation 5.13 was done by gradient descent with respect to the bias values of each action i as

$$\frac{\partial U}{\partial b_i} = -2 w_i \left(\hat{\Delta}_{t_f} - \sum_{a \in \mathcal{A}} w_a b_a \right), \quad i \in \mathcal{A} \tag{5.15}$$

Updating the bias values according to this gradient had the form

$$b_{i,+} = b_{i,-} - \xi \frac{\partial U}{\partial b_i}, \quad i \in \mathcal{A} \tag{5.16}$$

with $b_{a,-}$ and $b_{a,+}$ being the bias values before and after the update respectively and ξ the step size along the gradient. Substituting the gradient function at 5.15 into

5.16 resulted in the bias update

$$b_{i,+} = b_{i,-} - 2\xi w_i \left(\hat{\Delta}_{t_f} - \sum_{a \in \mathcal{A}} w_a b_a \right), \quad i \in \mathcal{A} \tag{5.17}$$

This was the normative formula for bias updates in order to minimise equation 5.13. It enabled the agent to achieve the hypothesised goal of holding actions accountable for the eventual episode's residue. Each bias update was then proportional to

$$\Delta b_a \propto \hat{\Delta}_{t_f} - \sum_{a \in \mathcal{A}} w_a b_a \tag{5.18}$$

It can be shown that the proposed Δ_t update of equation 5.8 is an approximation to the above update rule. First by expanding the recursive formula in equation 5.14, while assuming that its initial value was 0, it followed that

$$\begin{aligned} \hat{\Delta}_{t+1} &= (1-\psi)^t \cancelto{0}{\hat{\Delta}_0} + \sum_{i=1}^{t} \psi(1-\psi)^{t-i}\delta_i \\ &= \sum_{i=1}^{t} \psi(1-\psi)^{t-i}\delta_i \end{aligned} \tag{5.19}$$

Performing the same expansion of the delta rule in equation 5.8, Δ_{t+1} is written as a function of all past RPEs and biases

$$\begin{aligned} \Delta_{t+1} &= (1-\psi)^t \cancelto{0}{\Delta_0} + \sum_{i=1}^{t} \psi(1-\psi)^{t-i} \left(\delta_i - b_{A(i),-}\right) \\ &= \sum_{i=1}^{t} \psi(1-\psi)^{t-i} \left(\delta_i - b_{A(i),-}\right) \end{aligned} \tag{5.20}$$

Consequently, by substituting equation 5.19 into equation 5.20 we got

$$\Delta_{t+1} = \hat{\Delta}_{t+1} - \sum_{i=1}^{t} \psi(1-\psi)^{t-i} b_{A(i),-} \tag{5.21}$$

Changing the summation over timesteps to a summation over possible actions yielded

$$\Delta_{t+1} = \hat{\Delta}_{t+1} - \sum_{a \in \mathcal{A}} b_{a,-} \sum_{j \in T(a,t)} \psi(1-\psi)^{t-j} \tag{5.22}$$

where $T(a,t)$ was the set of all timesteps when action a was taken until and including timestep t. Rewritting equation 5.22 for the final timestep t_f and simplifying the notation of the weights of each $b_{a,-}$ term resulted in

$$\Delta_{t_f} = \hat{\Delta}_{t_f} - \sum_{a \in \mathcal{A}} k_a b_{a,0} \tag{5.23}$$

where

$$k_a = \sum_{j \in T(a,t_f)} \psi(1-\psi)^{t-j} \tag{5.24}$$

Substituting the above definition of Δ_{t_f} into the bias update rule (equation 5.9) led to

$$b_{\mathrm{a},+} = b_{\mathrm{a},-} + \xi M_{\mathrm{a}} \left(\hat{\Delta}_{t_f} - \sum_{a \in \mathcal{A}} k_a b_{a,0} \right) \tag{5.25}$$

which was rewritten to show that the difference between consecutive bias values after an update was proportional to

$$\Delta b_a \propto \hat{\Delta}_{t_f} - \sum_{a \in \mathcal{A}} k_a b_{a,-} \tag{5.26}$$

This gradient matched the normative gradient that the least-square problem should follow, in equation 5.18, with the bias weights being equal to

$$w_a = k_a = \sum_{j \in T(a,t_f)} \psi(1-\psi)^{t-j} \tag{5.27}$$

Therefore an agent that integrated the difference between RPEs and the bias of the chosen action as Δ_t, in equation 5.8, and performed bias updates according to equation 5.9 approximated a normative solution to the regression problem of equation 5.13. Thus, the agent updated its bias values offline in order to match its affective state at the end of the latest episode.

The proposed bias update rule could also be seen as a delta rule, akin to the standard tabular learning of equation 5.7, with two exceptions: the same quantity, Δ_{t_f}, was being used to update all bias values and the update for each bias value additionally depended on the bias values of other actions that were taken during the episode.

Relation between the Delta agent and a contextual value agent

A model that has been used to explain experimental observations of how non-immediate rewards influence actions around them is the contextual value agent. Given that the Delta agent was motivated by similar behavioural observations, this section was concerned with the relation between the two agent models. It was shown that the Delta agent included a running estimate of reward rate, similar to the contextual value agent, although the reward rate affected the agent's value updates in the opposite way compared to the contextual value agent.

The contextual value agent model learns context-dependent state- or action-values by comparing the immediate reward received with a recency-weighted average of past rewards, or reward rate. In order to achieve this, the agent integrates all rewards received into an estimate of reward rate, using a delta rule with a learning rate, which modulates how far into the past rewards are being considered. Higher values of the learning rate parameter create an estimate for the reward rate by including rewards from the recent past and, conversely, lower values integrate rewards further into the past.

In a vigilance task, where rewards were delivered after a variable number of trials in different cued contexts, the error rate of macaques was dependent on the proximity of the current trial to the rewarded one and on the context itself. Temporal difference learning of immediate rewards offset by an average of reward rate could explain these results (Dayan 2009). Punishment avoidance is another topic that has been addressed with a similar model, offering an explanation about how avoiding a punishment could be rewarding and thus reinforced as an action (Palminteri et al. 2015). Macaques' choices in a probabilistic reward learning task also depend on the average reward rate of the past and the contextual value model could account for differences in stay/switch behavior, when recent reward history was either rich or poor (Wittmann et al. 2020).

A contextual value agent updates the average reward rate ρ_t as

$$\rho_{t+1} = \rho_t + \psi(R_t - \rho_t) \tag{5.28}$$

and its Q action-values as

$$Q_{a,t+1} = Q_{a,t} + \eta(R_t - \rho_t - Q_{a,t}) \tag{5.29}$$

Expanding on the recursive formula of equation 5.28 led to the reward rate as a function of all past rewards

$$\begin{aligned}\rho_{t+1} &= (1-\psi)^t \cancelto{0}{\rho_0} + \sum_{i=1}^{t} \psi(1-\psi)^{t-i}\left(R_i\right) \\ &= \sum_{i=1}^{t} \psi(1-\psi)^{t-i}\left(R_i\right) \end{aligned} \tag{5.30}$$

Applying the same expansion on the Q update rule resulted in

$$\begin{aligned}Q_{a,t+1} &= (1-\eta)^{N(a,t)} \cancelto{0}{Q_{a,0}} + \sum_{i=1}^{N(a,t)} \eta(1-\eta)^{N(a,t)-i}\left(R_{T(a,i)} - \rho_{T(a,i)}\right) \\ &= \sum_{i=1}^{N(a,t)} \eta(1-\eta)^{N(a,t)-i}\left(R_{T(a,i)} - \rho_{T(a,i)}\right) \end{aligned} \tag{5.31}$$

where the function $N(a,t)$ calculated the number of times action a was taken up to timestep t and $T(a,i)$ provided the timestep t, at which action a was taken for the i^{th} time.

The integrated reward rate, ρ_t appeared in Delta agent's updates, first by expanding the Δ update rule of equation 5.8,

$$\begin{aligned}\Delta_{t+1} &= \Delta_t + \psi\left(\delta_t - b_{A(t),-} - \Delta_t\right) \\ &= (1-\psi)^t \cancelto{0}{\Delta_0} + \sum_{i=1}^{t} \psi(1-\psi)^{t-i}\left(\delta_i - b_{A(i),-}\right) \\ &= \sum_{i=1}^{t} \psi(1-\psi)^{t-i}\left(R_t - Q_{A(i),t} - b_{A(i),-}\right) \\ &= \sum_{i=1}^{t} \psi(1-\psi)^{t-i} R_t - \sum_{i=1}^{t} \psi(1-\psi)^{t-i}\left(Q_{A(i),t} + b_{A(i),-}\right) \end{aligned} \tag{5.32}$$

the first summation term was the reward rate of equation 5.30

$$\Delta_{t+1} = \rho_{t+1} - \sum_{i=1}^{t} \psi(1-\psi)^{t-i}\left(Q_{A(i),t} + b_{A(i),0}\right) \tag{5.33}$$

assuming equal values of learning rate, ψ, for ρ_t and Δ_t. Finally the bias update

rule in equation 5.9, after substituting equation 5.33, became

$$\begin{aligned} b_{\mathrm{a},+} &= b_{\mathrm{a},-} + \xi M_{\mathrm{a}} \Delta_{t_f} \\ &= b_{\mathrm{a},-} + \xi M_{\mathrm{a}} \left[\rho_{t_f} - \sum_{i=1}^{t_f} \psi(1-\psi)^{t_f - i} \left(Q_{A(i),t} + b_{A(i),-} \right) \right] \end{aligned} \tag{5.34}$$

Thus the bias values of the Delta agent are proportional to the integrated value of the reward rate at the end of an episode, ρ_{t_f}. Conversely, the same estimate of reward rate has the opposite relationship with the Q values of the contextual value agent in equation 5.31.

5.2.3 Monte Carlo agent

The simulated tasks, DB_1, DB_2 and DB_3 (Figure 5.1A-C) involved deferred rewards that were proportional to the number of times each action was chosen in the past timesteps of the same episode. These rewards occurred after the final action of each episode. We used an every-visit Monte Carlo (MC) agent, as it was presented by Sutton and Barto (Sutton et al. 2020, section 5.4), for comparisons with the Delta agent. The Monte Carlo agent stored all future rewards after each timestep and used them to learn its Q_{MC} action values offline. This way it learned a recency-weighted average of the episode's reward, G_t, after an action at time t

$$G_t = \sum_{i=t}^{t_f} \gamma^{i-t} R_i \tag{5.35}$$

$$Q_{MC,a,t+1} = Q_{MC,a,t} + \eta(G_t - Q_{MC,a,t}) \tag{5.36}$$

Here we set γ equal to 0, assuming no temporal discounting. Note that G accumulated over each episode and included the deferred reward that was delivered right before the end of it. Rewriting the learned Q_{MC} values in terms of all past accumulated rewards G, revealed that the MC agent assigned accumulated reward to each action in proportion to the number of times that action was taken, $N(a,t)$, until the current timestep

$$\begin{aligned} Q_{MC,a,t+1} &= (1-\eta)^{N(a,t)} \cancelto{0}{Q_{MC,a,0}} + \sum_{i=1}^{N(a,t)} \eta(1-\eta)^{N(a,t)-i} G_{T(a,i)} \\ &= \sum_{i=1}^{N(a,t)} \eta(1-\eta)^{N(a,t)-i} G_{T(a,i)} \end{aligned} \tag{5.37}$$

assuming initial action vales equal to 0. The $T(a, i)$ function indicated the timestep when action a was chosen for the i^{th} time in an episode. We used the MC agent for comparison, as an ideal model that held all actions and rewards in memory for offline learning and updated its action values in a similar manner to the deferred reward generation rule. The similarity of its learning rule to the deferred reward's one became more evident by rewriting the reward of an episode after time t as a function of the deferred reward

$$G_t = R_d + \sum_{i=t}^{t_f} R_i \tag{5.38}$$

since R_d was always delivered at $t = t_f$ and $\gamma = 1$. Then by substituting in (5.37) we got

$$Q_{MC,a,t+1} = \sum_{i=1}^{N(a,t)} \left[\eta(1-\eta)^{N(a,t)-i} \left(R_d + \sum_{j=T(a,i)}^{t_f} R_j \right) \right] \tag{5.39}$$

while at the end of an episode we had $N(a, t_f) = N_a$ by definition and the Q_{MC} values were

$$Q_{MC,a,t_f} \propto R_d \sum_{i=1}^{N_a} \eta(1-\eta)^{N_a-i} \tag{5.40}$$

The MC agent's learning rule, equation 5.40, showed that the agent integrated the deferred reward R_d into all of its action values $Q_{MC,a}$ in proportion to the deferred reward magnitude and a weight term that depended on N_a. Unlike the linear relationship between R_d and N_a for deferred reward generation in equation 5.2, the MC agent's learning had a nonlinear dependence on N_a. However, among classic, tabular reinforcement learning agents, this agent was the most advantageous and thus it was used as an ideal case. Additionally, the agent considered all potential delays on the Delayed Bandit task (Figure 5.1D), as it integrated all future rewards in an episode. Since the agent had no information about the distribution of time delays, considering all future rewards, up to the episode's end, was most beneficial.

5.2.4 Software

All code to define the Delta agent and the DB and TB environments and run the simulations was developed in Julia (v.1.6) (Bezanson et al. 2015) and is freely available online (`https://github.com/harisorgn/deferred_RL`).

5.3 Results

The Delta agent was tested along with the Monte Carlo agent and a variant of the Delta agent without any offline learning, named Delta-online, which had all $b_{a,\cdot}$ values fixed to 0. The latter was included to assess the contribution of the offline component of bias learning. A sensitivity analysis was conducted over the online learning rate, η, the common parameter for all agents, in the range $[0.05, 0.95]$ with a step of 0.1. Simulations were run for 50 steps per episode, 300 episodes, repeated for 1000 runs, for each value of η.

The policy and agents' parameters used in these simulations are summarised in Table 5.1. The η values shown are the ones that led to optimal performance during the sensitivity analysis for each agent. The learning rate for the episode's residue, Δ, was set to 0.02 to match the timescale of the episode (50 timesteps). The offline learning rate for bias values, ξ, the number of offline updates M and the lapse rate ϵ were chosen after a coarse grid search.

Delta				Delta-online	Monte Carlo		ϵ-greedy policy
η	ψ	ξ	M	η	η	γ	ϵ
0.05	0.02	0.1	10	0.05	0.05/0.55	1.0	0.1

Table 5.1: **Agent and policy parameters :** These parameters were used to run the Deferred Bandits and the Delayed Bandits tasks. Online learning rate parameters η were chosen based on a grid search and only the Monte Carlo's η was changed to 0.55 for the Delayed Bandits simulation. η: online learning rate, ψ: learning rate for Δ (episode's residue), ξ: offline learning rate, M: number of offline sampling steps, γ: reward discount rate, ϵ: probability of choosing an action at random, Delta-online: the variant of Delta agent that does not include the offline, bias-learning, component.

Two metrics were chosen for the sensitivity analysis over the online learning rate: the first episode T_0 when the agent's episode reward matched that of the random agent and R_e, the reward of the final episode. This way agents were tested both in terms of their learning speed and of their long-term performance. The complete results of the sensitivity analysis are shown in Figures 5.2 and 5.3 for T_0 and R_e respectively.

In the Deferred Bandits tasks, adding the offline bias learning to the Delta agent gave the agent an apparent advantage over a purely online variant, particularly in the reward of the final episode R_e, where the Delta agent outperforms the Delta-online one (Figure 5.3A-C). The steepness of their learning curves was matched for small learning rate values, as indicated by T_0 (Figure 5.2A-C). The Delta agent's

learning speed deteriorated in DB_1 (Figure 5.2A) with increasing online learning rates, which was the opposite relationship of all other agents and variants. However it still managed to reach comparable reward amounts at episode 300, noticeably higher than Delta-online, as shown in Figure 5.3A.

For small learning rates the Delta agent had similar performance to the MC agent. The Delta's reward at the final episode (Figure 5.3) tended to stay larger in the investigated range of learning rates. Comparing learning speed, the MC agent fell short of Delta for small values on DB_1 (Figure 5.2A), with a quick rise for larger values. In DB_2 (Figure 5.2B), Delta's performance rose more quickly throughout the range, while the two agents' T_0 completely overlap in DB_3 (Figure 5.2C).

The second type of task was the Delayed Bandit task. The average time delay, λ of the Poisson distribution, was set to four timesteps. In these simulations, the Delta agent accumulated more reward (Figure 5.3D) and faster (Figure 5.2D) for online learning rate values between 0.05 and approximately 0.2 compared to the other two models. While its T_0 value remained small for larger values of η, the last episode's reward, R_e, deteriorated. Conversely, the R_e value for the MC agent increased over the first half of the range of η and remained higher in the second half. The benefits of offline learning became apparent when looking at Figure 5.3D, where the Delta agent's curve was at a fixed distance above that of the Delta-online agent.

Finally, Figure 5.4 includes an example run for all agents using the online learning rate values that resulted in the optimal values for T_0 and R_e during the grid search. The plotted metric here was the accumulated reward for each episode, relative to the reward earned by an agent following a random policy, $\tilde{R}$.

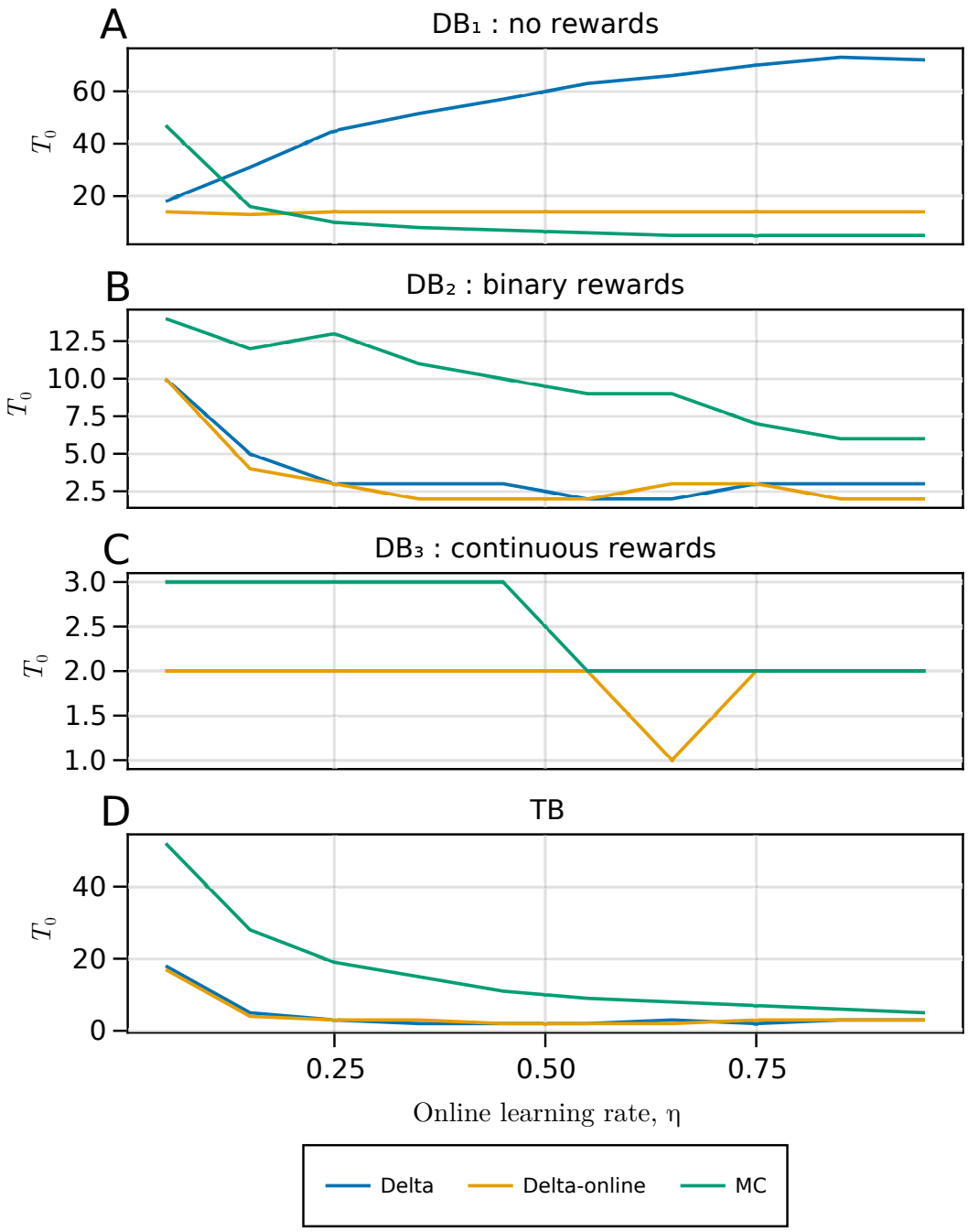


Figure 5.2: **Sensitivity analysis of learning speed :** The learning speed of all agents was assessed by measuring the number of episodes, T_0, it took each one to match the performance of an agent that takes actions at random. The performance metric used was $\hat{R}$, the accumulated reward per episode relative to the random agent. In the case of the DB_3 environment, the Delta and MC agents completely overlap. Lines represent average values over 1000 runs. Delta-online was the variant of Delta agent that did not include the offline, bias-learning, component. DB : Deferred Bandits, TB : Tardy Bandits.

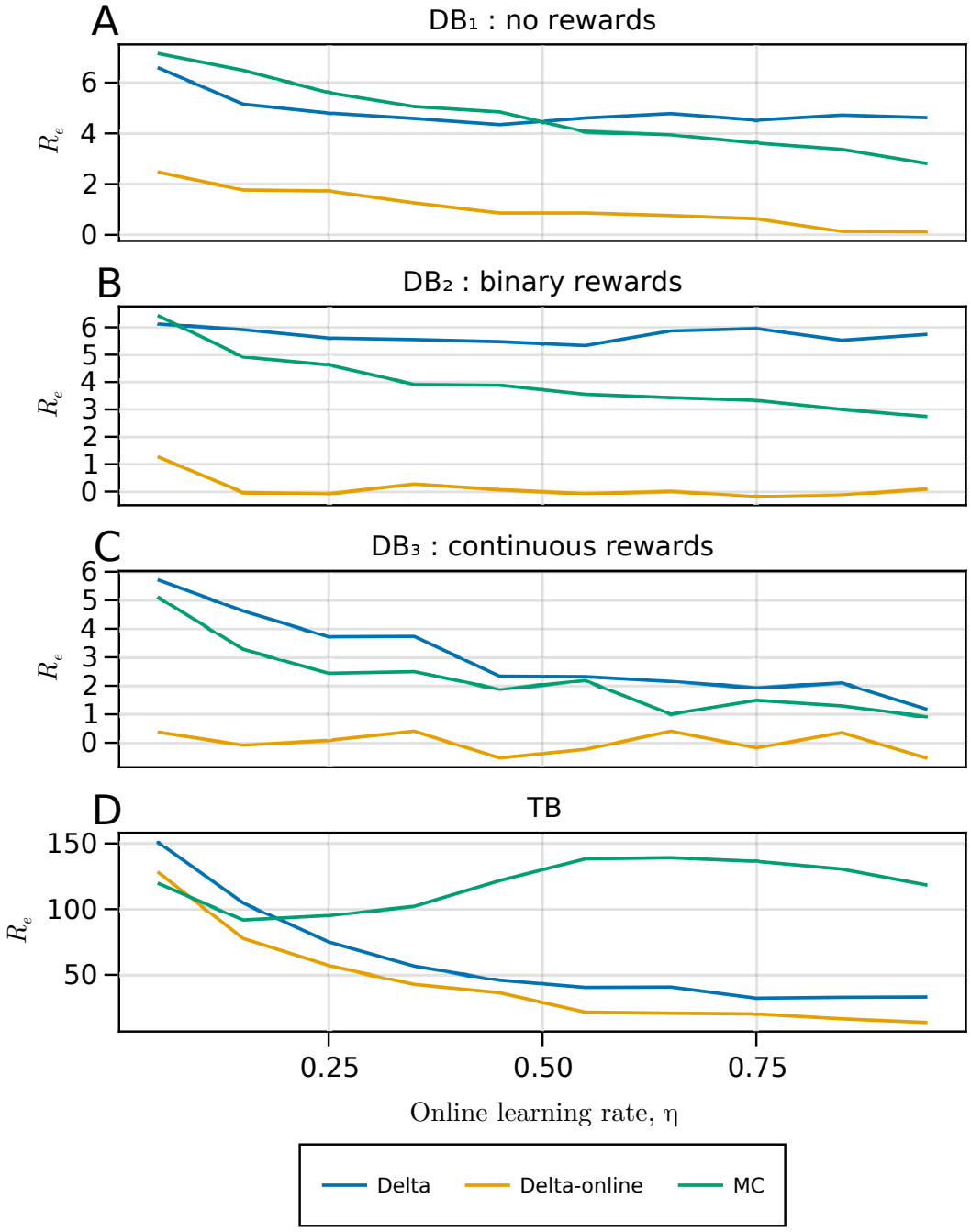


Figure 5.3: **Sensitivity analysis of long-term performance :** The performance of all agents after 300 episodes, R_e, was evaluated by measuring the accumulated reward at the 300$^{\text{th}}$ episode, relative to that of a random agent. Lines represent average values over 1000 runs. Delta-online was the variant of Delta agent that did not include the offline, bias-learning, component. DB : Deferred Bandits, TB : Tardy Bandits.

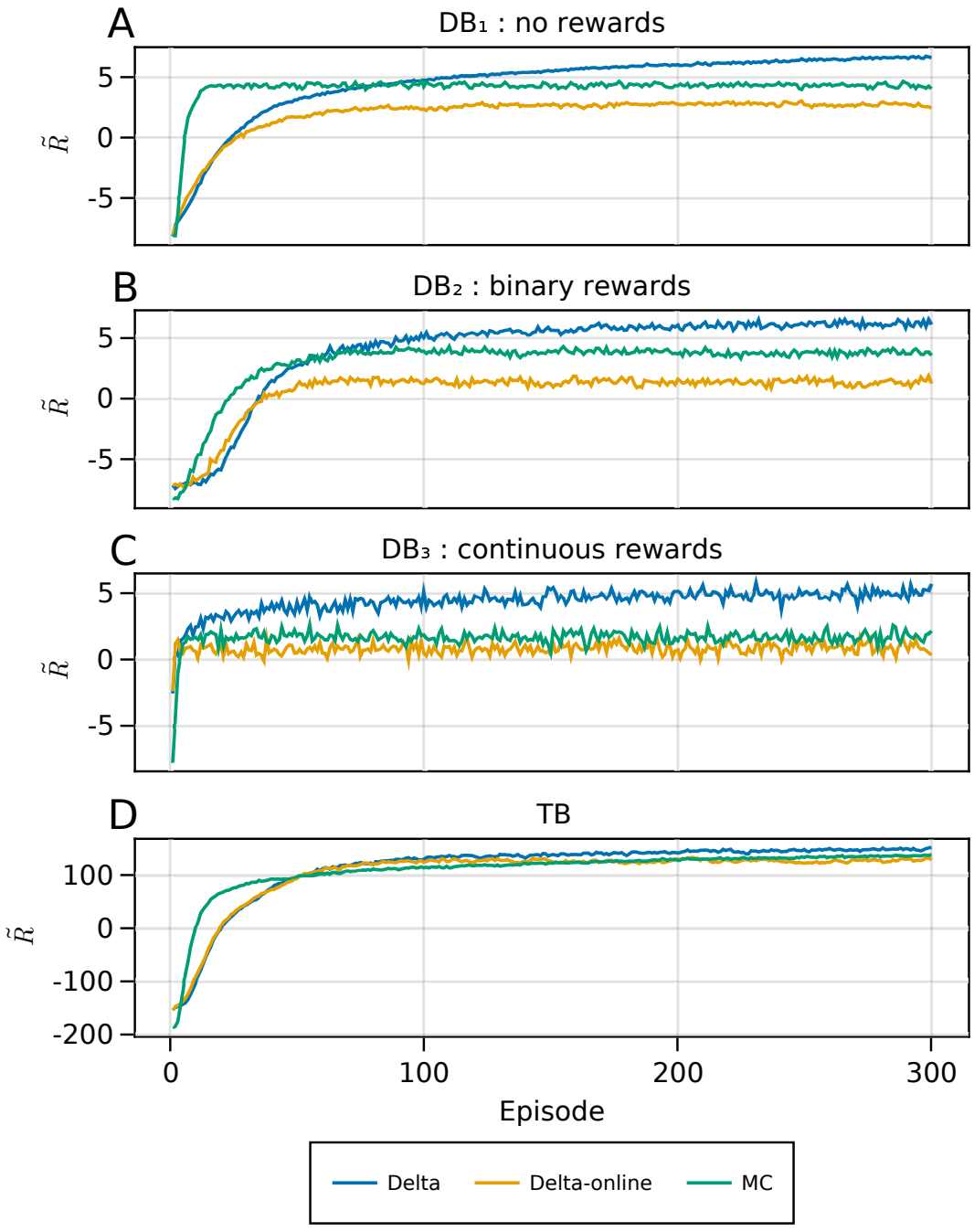


Figure 5.4: **Example run for all agents :** A timeline of the accumulated reward per episode $\hat{R}$, relative to that of an agent implementing a random policy is shown for the three agents, over the four environments. Lines represent average values over 1000 runs. Delta-online was the variant of Delta agent that did not include the offline, bias-learning, component. DB : Deferred Bandits, TB : Tardy Bandits.

5.4 Discussion

One would not have to look further than everyday interactions to come up with situations where an individual's actions would lead to outcomes that do not obey some strict temporal order. Multiple actions could be taken before potential outcomes would be experienced by the individual, with outcome delays not being constant or known in advance in novel environments. Assigning credit to the right action to ensure future proliferation becomes even more challenging in such environments. For instance, an animal that falls sick after consuming various kinds of food within a day needs to figure out which food source led to its poor health. Without substantial prior experience of these food sources, the animal might be forced to avoid more than one of the food sources that were experienced during the same day. Though common in the real world, non-local action-reward contingencies tend to be absent from laboratory experiments investigating learning and decision-making or simulated environments, designed to assess agent models.

In this chapter, we put forward a hypothesis claiming that the non-local associations, observed in human and animal behavioral studies, are the result of learning rules developed to cope with the challenging nature of causal action-reward relations in natural environments. The Delta agent was presented as a mathematical implementation of our hypothesis. The agent held no a priori knowledge of the structure of our simulated tasks, yet it built non-local contingencies indirectly through offline replay. The guiding quantity for learning offline was a leaky integrator of reward prediction errors (RPEs), which could be thought of as a proxy for affective state, as past experimental and theoretical work has argued (Blain et al. 2020; Bennett et al. 2020; Eldar et al. 2015; Rutledge et al. 2014).

Our simulation results indicated that the hypothesised offline learning rule was beneficial most of the time over a simpler, online Q learning model, both in terms of the learning speed and eventual performance. Additionally it matched, and in some cases, surpassed an MC agent, which updated each action according to all outcomes that occurred after it. These results highlight the functional benefits of using a mood-like quantity as the gradient for offline learning. Moreover, the Delta agent is a more plausible model for animals' behaviour than the MC agent. The MC agent required the log-term future gain after each timestep to be stored in memory and subsequently the actions of the episode to be replayed in the order they occurred. On the other hand, the Delta agent updated action values using the immediate rewards and only stored the surprisingness of each action in memory, along with the

value of the day's residue, for the offline learning phase.

Synergy between the online Q values and the offline bias values of the Delta agent is more robustly attained for smaller values of the online learning rate. In this regime, Q values do not closely track the immediate outcomes delivered, leaving room for the RPE integrator, or episode's residue Δ, to accumulate errors and subsequently guide bias learning offline.

The timescale of the RPE integrator was set equal to an episode's length and was not included in the sensitivity analysis, as the only relevant information from the rodent ABT experiments was that all experiences within a training day influenced the observed choice bias during testing. This timescale was conceptually related to the frequency of offline replay events. As the model calculated an average of errors and during an offline phase tried to account for these errors by looking back to the actions taken, the timescale of the RPE integrator, Δ, should match the time between two consecutive offline phases. How replay events are triggered remains an open question. This work assumed that they only occur after an episode, when the agent is not engaged with a task.

Our Delta agent suggests a process of sampling from past experiences in order to account for the agent's current affective state. The quantity that is being integrated to approximate affective state and the algorithm used to sample past experiences were parsimonious choices that could be easily extended in more sophisticated environments. For example, the leaky integrator over RPEs could be replaced by an integrator of advantage. Advantage is an estimate of how valuable an action is over its alternatives in a certain state. It could be approximated by an RPE integrator or, more faithfully, by using counterfactual information about alternative actions. It has recently been shown to account for mood-related effects on learning tasks, while resulting in reduced variance during learning in simulations (Bennett et al. 2020). Integrating advantage instead of RPE values, as the episode's residue Δ, could extend the efficacy of the Delta agent to environments that offer counterfactual information.

Our prioritised sampling process during replay could also be extended to include utility metrics for sampling each action-state pair in environments with multiple states and where the agent has some knowledge of the possible transitions between states through its actions. Examples of such metrics, supported by experimental evidence, are the the estimated frequency of state visits and the long-term gain after a state visit (Mattar et al. 2018), or the prioritised replay of unrewarded action that

were later avoided (Eldar et al. 2020).

Animal studies, based on the Affective Bias Test, have indicated that actions are being associated with temporally distant internal outcomes, in the form of affective state changes of either valence. These observations along with the fact that affective state is intertwined with most psychiatric disorders, has motivated the design of the DB and TB tasks, based on more naturalistic action-outcome contingencies, along with the Delta agent as a reinforcement learning model, augmented with affective-based rules for replay and offline learning.

The work of this chapter was a simulation-based proof of concept, which showed how biases due to affective state might develop. Future human experiments employing our DB and TB tasks, or similar ones that deviate from the traditional trial-based structure, along with affective state assessments at various times around the learning sessions could provide more naturalistic data. Such tasks could aid in the characterisation of deficits related to learning and decision-making, which are only present in the interaction of psychiatric patients with natural environments (Scholl et al. 2018). Our DB and TB tasks were designed with the specific goal of capturing the non-local action-reward contingencies of natural environments, where multiple actions could be taken by an agent before an outcome is delivered. Additionally, computational models, such as the Delta agent, that are able to account for such features of real environments, are beneficial in linking the results of studies in patient groups with experiments involving animal models, such as the ABT.

Inference of model parameters and comparison between the Delta agent and alternative designs, for example the contextual value agent, would be critical in elucidating how affective state is implicated in the consolidation of recent experiences. An open question of translational value is how sudden changes in affective state due to a treatment, like ketamine, could cause reconsolidation of past experiences without re-exposure to them, as has been shown in rats using the ABT (Stuart et al. 2015b).

Chapter 6

General discussion

The present thesis incorporated different approaches for measuring the affective state of rats by using observations of their actions in two behavioural tasks, involving both rewards and avoidance of punishments. An initial inference by the summary statistics of actions in the Judgement Bias Task (JBT) of chapter 2 was complemented by a hierarchical statistical model in chapter 4, in an effort to dissect the animals' behaviour into factors that were deemed important. A similar approach was employed in chapter 3 where a hierarchical model was used to capture the variability of individual animals within and across each condition of the Conditioned Suppression Task (CST). Finally, a novel theoretical model was proposed in 5, inspired by observations in a different assay where biases due to affective state have been measured, the Affective Bias Test (ABT). The current chapter has brought together common themes that emerged throughout the work of my thesis.

6.1 Affective state as part of a closed loop

Multiple lines of experimental work have investigated affective state as a causal factor behind an animal's actions. For instance, the JBT, the CST, and most of the literature that was discussed in the introductory chapter 1 were concerned with this directional relationship between affective state and multiple types of actions, such as responding after the interpretation of ambiguous information or suppressing a foraging behaviour because of an anticipated threat. However, a causal relationship of the opposite direction seems most likely to exist as well, with actions and their outcomes changing one's affective state.

Recently, there has been a plethora of studies that tried to address this bidirectional relationship between affective state and actions (Rutledge et al. 2014; Eldar et al. 2015; Eldar et al. 2016; Rutledge et al. 2017; Neville et al. 2021). In order to account for unobservable changes in affective state, these studies employed computational models, which incorporated hypotheses about how affective state changes given recent experiences. My present work on a large-scale analysis of the JBT found evidence suggesting that the animals were not interpreting the ambiguous stimuli purely based on their prior expectations of a small or large reward, but they also were influenced by past feedback. This influence could be interpreted as a result of learning new action-outcome contingencies for the ambiguous cue. However, an alternative explanation is that it is also a result of changes to the animals' affective state during the session. The probabilistic feedback after each response to the ambiguous cue could result in a mismatch between the acquired outcome and the animals' expectation. This mismatch has been known to modulate humans' affective state (Rutledge et al. 2014; Eldar et al. 2015; Rutledge et al. 2017; Villano et al. 2020; Bennett et al. 2020). Assuming that a similar modulation of affective state by prediction errors is present in non-human animals', it is possible that the prediction errors in ambiguous trials in the JBT lead to changes in affective state, which in turn affect the way the ambiguous cue is being interpreted.

This proposal about the modulation of affective state in animals was further corroborated in the theoretical work of chapter 5. In simulated environments, which captured more naturalistic contingencies between actions and their outcomes, affective state was shown to be a beneficial component of a learning agent. There, affective state was situated in a feedback loop, where it could change as a result of prediction errors. It was further used in a predictive manner, as the agent considered the future affective state that it could find itself experiencing, when deciding on how to act in its environment. Thus, affective state is an active part of both feedback-adaptive and feedforward-predictive internal loops that shape an animals' actions. These ideas have been previously expressed as descriptive theories of emotions (Baumeister et al. 2007), whereas now computational tools are available to investigate them in greater detail.

Similar ideas were part of classical theories about mood disorders. People in pathological states are entrenched in a loop of negative expectations leading to inaction or missed opportunities to experience positive outcomes, which further validate the negative expectations (Beck 1967). A more recent theory has extended this statement, claiming that prior to the development of pathology, associations can be created

between affectively charged experiences and an individual's actions (Robinson et al. 2016). Our Delta agent is able to capture such associations. Thus the agent can be used to make long-term predictions of this theoretical proposition by running longitudinal simulation studies to evaluate how making associations between the agent's actions and its affective state may lead to the development of pathological states.

Statistical models, such as the one that was fit to JBT data in chapter 4, could be extended by theoretical models, such as the Delta agent of chapter 5 in order to capture a bidirectional relationship between the animals' affective state and their actions. Particularly when trying to infer the effect of manipulations with translational value, such as ketamine, it is worth considering how malleable affective state is under the influence of the drug. For instance, a successful antidepressant would have to be able to cause a positive affective state that is stable enough against the typical negative surprises of the environment, at least in the short term, before the patient is able to maintain a healthy affective state on their own.

6.2 Rewarding avoidance and punishing omission

The presence or omission of rewards and punishments are distinct types of outcome, which were present in both of the tasks that were implemented in the present thesis. In JBT the outcomes were food rewards, which could have been omitted after an incorrect action or due to the probabilistic feedback of the ambiguous cue. In CST, every trial presented a choice between avoiding a potential punishment or keep exerting effort for a food reward. Since both tasks aimed at inferring an animal's affective state via its actions, the valence of reward omission and punishment avoidance will be evaluated in light of our theoretical model, which situates affective state in-between actions and outcomes.

The case of reward omission is more straightforward according to reinforcement learning theory (Sutton et al. 2020). An action is positively reinforced, and hence more readily repeated, after a reward exceeds an animal's expectation for it. This reinforcement implies that the animal has learned to expect a positive outcome after its action. Even though the JBT was not designed to encourage learning, its premise was that animals act according to their expectation of reward. Naturally then, when this expectation is not met, a negative prediction error is created, which is known to negatively modulate affective state, as discussed in the previous section and in chapter 4. Thus, the omission of reward could be considered an affectively negative

event. Our theoretical model would predict that the negative affective state caused by the negative prediction errors after reward omission would result in a bias away from the action that led to this error. This proposition has implications particularly for the trials where the ambiguous cue is presented, as there is a 50% chance that there will be no reward after an animal's response. The basic premise of JBT is that animals respond to the ambiguous cue according to their expectations of reward, thus a nonzero prediction error must be present during these trials.

Avoidance of punishment can not be explained as easily as acting to acquire reward, since after the punishment has been successfully avoided, there is no outcome to reinforce the avoidance action. Previous theoretical work has assumed that punishment avoidance needs to be rewarding so that the avoidance action can be reinforced (Kim et al. 2006; Moutoussis et al. 2008), although it is also possible to choose an avoidance action of neutral value, because the action to stay has been negatively reinforced (Maia 2010). More recently, a contextual value model of reinforcement learning was used to explain how the avoidance action acquires a positive value (Palminteri et al. 2015). In chapter 5, our Delta agent model was shown to incorporate this definition of contextual value, yet it used it in the opposite way compared to the contextual value model.

The contextual value model predicts that the avoidance action will acquire a positive value, since it is learned by comparing the immediate outcome with the average value of the context, which is negative due to past electric shocks. On the other hand, our Delta agent predicts that the action of not escaping will acquire a negative bias, during the offline phase when the animal tries to account for its negative affective state. The action value of escaping in our model, reflecting the immediate outcome of taking it, would be neutral, since no explicit outcome occurs after an avoidance action.

Even though the action value of our model is not reinforced immediately after an escape, the day's residue is positively updated. The day's residue is the running average of past prediction errors, reflecting the animal's affective state. It would be negative during training, since only negative surprises would take place in the form of electric shocks. After an escape, there is no immediate outcome and hence a prediction error of zero, which will decrease the magnitude of the day's residue, as this error is integrated in it. Thus, our model suggests that avoidance of punishment in our CST leads to an immediate positive shift in affective state, while the same action is chosen not because of a positive value, but because of a much more negative

value of the alternative action.

Overall, the decision on whether to stay or flee depends on the difference between the values of the two choices (Seymour et al. 2008). There are multiple ways to produce the same difference. The contextual value model and our Delta agent could in principle produce similar difference in values, even though they embody distinct learning processes.

Going beyond the JBT and the CST, towards more naturalistic environments, where there are multiple available actions varying by context, delayed outcomes and uncertainty in both rewards and punishments, could help to delineate the difference between the learning processes proposed by our proposed model and alternative ones. Chapter 5 presented several ideas about naturalistic task designs. Future work could implement these tasks as behavioural experiments in order to inform theoretical predictions by empirical evidence. After inferring model parameters from the choice data, the inferred values for each action, in each context it appears, could be inferred for each subject, thus providing an estimate of how rewarding is punishment avoidance and how punishing is the omission of reward. Complementary measures of affect, such as questionnaires in humans, could be utilised intermittently to assess whether punishment avoidance and reward omission influences affective state in a similar way as our model predicts.

6.3 The average rat

It does not exist. Individual differences between animals were observed in both of the behavioural tasks that were implemented in the present thesis, involving both rewarding and aversive experiences. Such differences could hide potential effects and bias the inference process towards an average case. Particularly when studying affective state, which is an unobservable quantity, intertwined with subjective experience, accounting for individual differences in behaviour could aid in inferring the true state of the animal and how that state influences actions.

Individual differences could be innate or they could depend on the past experiences of each animal (Gomez-Marin et al. 2019). Previous work has suggested that all historical information about the living conditions of animals should be reported in studies of affective state, as these factors can greatly confound the experimental results (Tye 2018). Hierarchical statistical models offer a way to account for indi-

vidual differences both within and across conditions, allowing for inference on both a subject level and a population level. This model structure was employed in the current thesis, in conjunction with generative models of actions in JBT and response times in CST.

The model-based inference of JBT data, across cohorts of rats originating in different breeding facilities, revealed differences in the parameter that was primarily linked to the interpretation bias hypothesis, which the task was based on. Moreover individual animals within each breeding facility group, exhibited variations in the weight they assigned to feedback from the most recent and longer-term past. A non-zero influence of past experiences could diminish the effect of an interpretation bias, thus confounding the interpretation of the experimental results as optimistic and pessimistic actions. Similarly in the CST, some animals exhibited the predicted suppression of reward-seeking, yet other animals had a clear opposite effect, as they sped up their responding under the threat of shock.

Looking into the behaviour of individual animals in both tasks, it is evident that different strategies could have been employed. Even animals, whose behaviour did not conform with the original hypothesis -responding according to a prior expectation of reward in JBT or suppressing responses in CST- could behave in ways that are influenced by their affective state. For instance, animals that change their action after a loss could have an increased sensitivity to negative feedback, a symptom that is common in patients with major depressive disorder (Murphy et al. 2003).

Averaging over animals would discard all information about such differences. This is particularly detrimental in animal models of psychiatric disorders. Diagnosis in patients is conducted on an individual level, yet in animal studies inference often follows group averages, thus neglecting potential differences that a treatment could have on individuals (Ardi et al. 2016). This disparity could reduce the translational value of animal models. One argument in favour of averaging is the reduction of noise or uninformative variability in the data (Estes 2002). However, the uninformative variability could be discarded along with useful information about each subject and how a treatment might have affected them. On the other hand, hierarchical statistical models allow for a more structured organisation of variability without loss of important information about individual subjects (Gelman et al. 2015; McElreath 2016; Ahn et al. 2017).

Particularly when paired with Bayesian inference, hierarchical models can be generative models, able to produce fictive data while preserving the heterogeneity between

animals. The posterior distribution over model parameters, which is the result of inference, could work as the prior expectation when designing future experimental studies. Fictive data could be created based on parameter values, which are sampled from the posterior distribution, to simulate cohorts of animals. If the goal of the experimental study is the investigation of particular strategies or biases, thought to be modulated by affective state, then an expectation for the prevalence of these behavioural patterns among the fictive cohort could be calculated. Assuming the inference was conducted properly given all the diagnostic checks, this expectation could work as a prediction for the prevalence of the relevant behaviour in future cohorts of animals. This prediction could aid in the optimisation of the number of animals employed in future studies.

References

Aarts, Emmeke et al. (Apr. 2014). "A Solution to Dependency: Using Multilevel Analysis to Accommodate Nested Data". In: *Nature Neuroscience* 17.4, pp. 491–496. ISSN: 1546-1726. DOI: 10.1038/nn.3648.

Aarts, Emmeke et al. (Dec. 2015). "Multilevel Analysis Quantifies Variation in the Experimental Effect While Optimizing Power and Preventing False Positives". In: *BMC Neuroscience* 16.1, p. 94. ISSN: 1471-2202. DOI: 10.1186/s12868-015-0228-5.

Ahn, Woo-Young, Nathaniel Haines, and Lei Zhang (Oct. 2017). "Revealing Neurocomputational Mechanisms of Reinforcement Learning and Decision-Making With the hBayesDM Package". In: *Computational Psychiatry (Cambridge, Mass.)* 1, pp. 24–57. ISSN: 2379-6227. DOI: 10.1162/CPSY_a_00002.

Akrami, Athena et al. (Feb. 2018). "Posterior Parietal Cortex Represents Sensory History and Mediates Its Effects on Behaviour". In: *Nature* 554.7692, pp. 368–372. ISSN: 1476-4687. DOI: 10.1038/nature25510.

Anagnostaras, Stephan G. et al. (Sept. 2010). "Automated Assessment of Pavlovian Conditioned Freezing and Shock Reactivity in Mice Using the Video Freeze System". In: *Frontiers in Behavioral Neuroscience* 4, p. 158. ISSN: 1662-5153. DOI: 10.3389/fnbeh.2010.00158.

Anderson, David J. and Ralph Adolphs (Mar. 2014). "A Framework for Studying Emotions across Species". In: *Cell* 157.1, pp. 187–200. ISSN: 1097-4172. DOI: 10.1016/j.cell.2014.03.003.

Anderson, Michael H., Marcus R. Munafò, and Emma S. J. Robinson (Apr. 2013). "Investigating the Psychopharmacology of Cognitive Affective Bias in Rats Using an Affective Tone Discrimination Task". In: *Psychopharmacology* 226.3, pp. 601–613. ISSN: 1432-2072. DOI: 10.1007/s00213-012-2932-5.

Anselme, Patrick (Feb. 2015). "Incentive Salience Attribution under Reward Uncertainty: A Pavlovian Model". In: *Behavioural Processes* 111, pp. 6–18. ISSN: 0376-6357. DOI: 10.1016/j.beproc.2014.10.016.

Antonov, Georgy et al. (Apr. 2021). "Optimism and Pessimism in Optimised Replay". In: *bioRxiv*, p. 2021.04.27.441454. DOI: 10.1101/2021.04.27.441454.

Ardi, Ziv et al. (Apr. 2016). "Behavioral Profiling as a Translational Approach in an Animal Model of Posttraumatic Stress Disorder". In: *Neurobiology of Disease* 88, pp. 139–147. ISSN: 0969-9961. DOI: 10.1016/j.nbd.2016.01.012.

Ashwood, Zoe C. et al. (Feb. 2022). "Mice Alternate between Discrete Strategies during Perceptual Decision-Making". In: *Nature Neuroscience*. ISSN: 1546-1726. DOI: 10.1038/s41593-021-01007-z.

Aylward, Jessica et al. (Jan. 2020). "Translating a Rodent Measure of Negative Bias into Humans: The Impact of Induced Anxiety and Unmedicated Mood and Anxiety Disorders". In: *Psychological Medicine* 50.2, pp. 237–246. ISSN: 0033-2917, 1469-8978. DOI: 10.1017/S0033291718004117.

Ayres, John J. B. (1968). "Differentially Conditioned Suppression as a Function of Shock Intensity and Incentive." In: *Journal of Comparative and Physiological Psychology* 66.1, pp. 208–210. ISSN: 0021-9940. DOI: 10.1037/h0025971.

Bach, Dominik R. and Peter Dayan (May 2017). "Algorithms for Survival: A Comparative Perspective on Emotions". In: *Nature Reviews Neuroscience* 18.5, pp. 311–319. ISSN: 1471-0048. DOI: 10.1038/nrn.2017.35.

Baddeley, Alan (2007). *Working Memory, Thought, and Action*. Oxford Psychology Series. Oxford: Oxford University Press. ISBN: 978-0-19-852801-2. DOI: 10.1093/acprof:oso/9780198528012.001.0001.

Bali, Anjana and Amteshwar Singh Jaggi (Dec. 2015). "Electric Foot Shock Stress: A Useful Tool in Neuropsychiatric Studies". In: *Reviews in the Neurosciences* 26.6, pp. 655–677. ISSN: 2191-0200. DOI: 10.1515/revneuro-2015-0015.

Bar-Haim, Yair et al. (Jan. 2007). "Threat-Related Attentional Bias in Anxious and Nonanxious Individuals: A Meta-Analytic Study". In: *Psychological Bulletin* 133.1, pp. 1–24. ISSN: 0033-2909. DOI: 10.1037/0033-2909.133.1.1.

Bardgett, Mark E. et al. (2009). "Dopamine Modulates Effort-Based Decision Making in Rats". In: *Behavioral Neuroscience* 123.2, pp. 242–251. ISSN: 1939-0084. DOI: 10.1037/a0014625.

Bari, Andrea et al. (May 2010). "Serotonin Modulates Sensitivity to Reward and Negative Feedback in a Probabilistic Reversal Learning Task in Rats". In: *Neuropsychopharmacology* 35.6, pp. 1290–1301. ISSN: 1740-634X. DOI: 10.1038/npp.2009.233.

Baribault, Beth and Anne Collins (Dec. 2021). *Troubleshooting Bayesian Cognitive Models: A Tutorial with Matstanlib*. DOI: 10.31234/osf.io/rtgew.

Barrett, Lisa Feldman (Jan. 2017). "The Theory of Constructed Emotion: An Active Inference Account of Interoception and Categorization". In: *Social Cognitive and Affective Neuroscience* 12.1, pp. 1–23. ISSN: 1749-5016. DOI: 10.1093/scan/nsw154.

Bauer, Elizabeth P. (Jan. 2015). "Serotonin in Fear Conditioning Processes". In: *Behavioural Brain Research*. Special Issue: Serotonin 277, pp. 68–77. ISSN: 0166-4328. DOI: 10.1016/j.bbr.2014.07.028.

Baumeister, Roy F. et al. (May 2007). "How Emotion Shapes Behavior: Feedback, Anticipation, and Reflection, Rather Than Direct Causation". In: *Personality and Social Psychology Review* 11.2, pp. 167–203. ISSN: 1088-8683. DOI: 10.1177/1088868307301033.

Bechara, Antoine and Antonio R. Damasio (Aug. 2005). "The Somatic Marker Hypothesis: A Neural Theory of Economic Decision". In: *Games and Economic Behavior*. Special Issue on Neuroeconomics 52.2, pp. 336–372. ISSN: 0899-8256. DOI: 10.1016/j.geb.2004.06.010.

Beck, Aaron T (1967). *Depression: Clinical, Experimental, and Theoretical Aspects*. New York: Hoeber Medical Division, Harper & Row.

Bell, Andrew and Kelvyn Jones (Jan. 2015). "Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data*". In: *Political Science Research and Methods* 3.1, pp. 133–153. ISSN: 2049-8470, 2049-8489. DOI: 10.1017/psrm.2014.7.

Bennett, Daniel, Guy Davidson, and Yael Niv (Aug. 2020). *A Model of Mood as Integrated Advantage*. Preprint. PsyArXiv. DOI: 10.31234/osf.io/dzsme.

Bennett, Daniel, Yael Niv, and Angela J. Langdon (2021). "Value-Free Reinforcement Learning: Policy Optimization as a Minimal Model of Operant Behavior". In: *Current Opinion in Behavioral Sciences* 41, pp. 114–121. ISSN: 2352-1554. DOI: 10.1016/j.cobeha.2021.04.020.

Berghorst, Lisa et al. (2013). "Acute Stress Selectively Reduces Reward Sensitivity". In: *Frontiers in Human Neuroscience* 7. ISSN: 1662-5161.

Berridge, Kent C and Terry E Robinson (Dec. 1998). "What Is the Role of Dopamine in Reward: Hedonic Impact, Reward Learning, or Incentive Salience?" In: *Brain Research Reviews* 28.3, pp. 309–369. ISSN: 0165-0173. DOI: 10.1016/S0165-0173(98)00019-8.

Berwian, Isabel M. et al. (May 2020). "Computational Mechanisms of Effort and Reward Decisions in Patients With Depression and Their Association With Relapse After Antidepressant Discontinuation". In: *JAMA Psychiatry* 77.5, pp. 513–522. ISSN: 2168-622X. DOI: 10.1001/jamapsychiatry.2019.4971.

Betancourt, M. J. and Mark Girolami (Dec. 2013). "Hamiltonian Monte Carlo for Hierarchical Models". In: *arXiv:1312.0906 [stat]*. arXiv: 1312.0906 [stat].

Betancourt, Michael (July 2018). "A Conceptual Introduction to Hamiltonian Monte Carlo". In: *arXiv:1701.02434 [stat]*. arXiv: 1701.02434 [stat].

Bezanson, Jeff et al. (July 2015). "Julia: A Fresh Approach to Numerical Computing". In: *arXiv:1411.1607 [cs]*. arXiv: 1411.1607 [cs].

Bitzer, Sebastian et al. (2014). "Perceptual Decision Making: Drift-Diffusion Model Is Equivalent to a Bayesian Model". In: *Frontiers in Human Neuroscience* 8, p. 102. ISSN: 1662-5161. DOI: 10.3389/fnhum.2014.00102.

Blain, Bastien and Robb B Rutledge (Nov. 2020). "Momentary Subjective Well-Being Depends on Learning and Not Reward". In: *eLife* 9. Ed. by Daeyeol Lee, Joshua I Gold, and Daeyeol Lee, e57977. ISSN: 2050-084X. DOI: 10.7554/eLife.57977.

Blaney, Paul H. (1986). "Affect and Memory: A Review". In: *Psychological Bulletin* 99.2, pp. 229–246. ISSN: 1939-1455. DOI: 10.1037/0033-2909.99.2.229.

Bogdan, Ryan and Diego A. Pizzagalli (Nov. 2006). "Acute Stress Reduces Reward Responsiveness: Implications for Depression". In: *Biological Psychiatry* 60.10, pp. 1147–1154. ISSN: 0006-3223, 1873-2402. DOI: 10.1016/j.biopsych.2006.03.037.

Bornstein, Aaron M. et al. (June 2017). "Reminders of Past Choices Bias Decisions for Reward in Humans". In: *Nature Communications* 8.1, p. 15958. ISSN: 2041-1723. DOI: 10.1038/ncomms15958.

Bouhuys, A. L., E. Geerts, and M. C. Gordijn (Oct. 1999). "Depressed Patients' Perceptions of Facial Emotions in Depressed and Remitted States Are Associated with Relapse: A Longitudinal Study". In: *The Journal of Nervous and Mental Disease* 187.10, pp. 595–602. ISSN: 0022-3018. DOI: 10.1097/00005053-199910000-00002.

Bower, Gordon H. (1981). "Mood and Memory". In: *American Psychologist* 36.2, pp. 129–148. ISSN: 1935-990X. DOI: 10.1037/0003-066X.36.2.129.

Bowman, E. M., T. G. Aigner, and B. J. Richmond (Mar. 1996). "Neural Signals in the Monkey Ventral Striatum Related to Motivation for Juice and Cocaine Rewards". In: *Journal of Neurophysiology* 75.3, pp. 1061–1073. ISSN: 0022-3077. DOI: 10.1152/jn.1996.75.3.1061.

Bradley, Brendan P., Karin Mogg, and Stacey C. Lee (Oct. 1997). "Attentional Biases for Negative Information in Induced and Naturally Occurring Dysphoria". In: *Behaviour Research and Therapy* 35.10, pp. 911–927. ISSN: 0005-7967. DOI: 10.1016/S0005-7967(97)00053-3.

Bravo-Rivera, Christian et al. (July 2014). "Neural Structures Mediating Expression and Extinction of Platform-Mediated Avoidance". In: *Journal of Neuroscience* 34.29, pp. 9736–9742. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.0191-14.2014.

Brown, Timothy A., Bruce F. Chorpita, and David H. Barlow (1998). "Structural Relationships among Dimensions of the DSM-IV Anxiety and Mood Disorders and Dimensions of Negative Affect, Positive Affect, and Autonomic Arousal". In: *Journal of Abnormal Psychology* 107.2, pp. 179–192. ISSN: 1939-1846. DOI: 10.1037/0021-843X.107.2.179.

Browning, Michael et al. (Apr. 2015). "Anxious Individuals Have Difficulty Learning the Causal Statistics of Aversive Environments". In: *Nature Neuroscience* 18.4, pp. 590–596. ISSN: 1546-1726. DOI: 10.1038/nn.3961.

Bryant, Camron D. (Dec. 2011). "The Blessings and Curses of C57BL/6 Substrains in Mouse Genetic Studies". In: *Annals of the New York Academy of Sciences* 1245, pp. 31–33. ISSN: 0077-8923. DOI: 10.1111/j.1749-6632.2011.06325.x.

Bush, David E. A., Francisco Sotres-Bayon, and Joseph E. LeDoux (2007). "Individual Differences in Fear: Isolating Fear Reactivity and Fear Recovery Phenotypes". In: *Journal of Traumatic Stress* 20.4, pp. 413–422. ISSN: 1573-6598. DOI: 10.1002/jts.20261.

Buzsáki, György, Adrien Peyrache, and John Kubie (2014). "Emergence of Cognition from Action". In: *Cold Spring Harbor symposia on quantitative biology* 79, pp. 41–50. ISSN: 0091-7451. DOI: 10.1101/sqb.2014.79.024679.

Caldji, Christian et al. (Mar. 2000). "The Effects of Early Rearing Environment on the Development of GABAA and Central Benzodiazepine Receptor Levels and Novelty-Induced Fearfulness in the Rat". In: *Neuropsychopharmacology* 22.3, pp. 219–229. ISSN: 1740-634X. DOI: 10.1016/S0893-133X(99)00110-4.

Caligiuri, M. P. and J. Ellwanger (2000 Jan-Mar). "Motor and Cognitive Aspects of Motor Retardation in Depression". In: *Journal of Affective Disorders* 57.1-3, pp. 83–93. ISSN: 0165-0327. DOI: 10.1016/s0165-0327(99)00068-3.

Carandini, Matteo and Anne K Churchland (July 2013). "Probing Perceptual Decisions in Rodents". In: *Nature neuroscience* 16.7, pp. 824–831. ISSN: 1097-6256. DOI: 10.1038/nn.3410.

Carboni, Ezio et al. (2021). "Repurposing Ketamine in Depression and Related Disorders: Can This Enigmatic Drug Achieve Success?" In: *Frontiers in Neuroscience* 15. ISSN: 1662-453X.

Cartoni, Emilio, Stefano Puglisi-Allegra, and Gianluca Baldassarre (Nov. 2013). "The Three Principles of Action: A Pavlovian-instrumental Transfer Hypothesis".

In: *Frontiers in Behavioral Neuroscience* 7, p. 153. ISSN: 1662-5153. DOI: 10.3389/fnbeh.2013.00153.

Charnov, Eric L. (Apr. 1976). "Optimal Foraging, the Marginal Value Theorem". In: *Theoretical Population Biology* 9.2, pp. 129–136. ISSN: 0040-5809. DOI: 10.1016/0040-5809(76)90040-X.

Cisler, Josh M. and Ernst H. W. Koster (Mar. 2010). "Mechanisms of Attentional Biases towards Threat in Anxiety Disorders: An Integrative Review". In: *Clinical Psychology Review* 30.2, pp. 203–216. ISSN: 0272-7358. DOI: 10.1016/j.cpr.2009.11.003.

Damasio, Antonio R. (2006). *Descartes' Error*. Random House. ISBN: 978-0-09-950164-0.

Davis, Hank (1968). "Conditioned Suppression: A Survey of the Literature". In: *Psychonomic Monograph Supplements* 2.14, pp. 283–291.

Davis, Hank and Janet Wright (Sept. 1979). "Procedural and Parametric Variability in Studies of Conditioned Suppression". In: *Bulletin of the Psychonomic Society* 14.3, pp. 149–150. ISSN: 0090-5054. DOI: 10.3758/BF03329429.

Dayan, Peter (Jan. 2009). "Prospective and Retrospective Temporal Difference Learning". In: *Network: Computation in Neural Systems* 20.1, pp. 32–46. ISSN: 0954-898X. DOI: 10.1080/09548980902759086.

Der-Avakian, A. et al. (Aug. 2013). "Assessment of Reward Responsiveness in the Response Bias Probabilistic Reward Task in Rats: Implications for Cross-Species Translational Research". In: *Translational Psychiatry* 3.8, e297–e297. ISSN: 2158-3188. DOI: 10.1038/tp.2013.74.

Der-Avakian, Andre et al. (May 2017). "Social Defeat Disrupts Reward Learning and Potentiates Striatal Nociceptin/Orphanin FQ mRNA in Rats". In: *Psychopharmacology* 234.9, pp. 1603–1614. ISSN: 1432-2072. DOI: 10.1007/s00213-017-4584-y.

Derryberry, Douglas and Marjorie A. Reed (May 2002). "Anxiety-Related Attentional Biases and Their Regulation by Attentional Control". In: *Journal of Abnormal Psychology* 111.2, pp. 225–236. ISSN: 0021-843X. DOI: 10.1037//0021-843x.111.2.225.

Detert, Julia A., Nicholas D. Kampa, and James R. Moyer (Dec. 2008). "Differential Effects of Training Intertrial Interval on Acquisition of Trace and Long-Delay Fear Conditioning in Rats". In: *Behavioral Neuroscience* 122.6, pp. 1318–1327. ISSN: 0735-7044. DOI: 10.1037/a0013512.

"APPETITIVE-AVERSIVE INTERACTIONS AND INHIBITORY PROCESSES" (1980). In: *Mechanisms of Learning and Motivation*. Ed. by A. Dickinson and R. A. Boakes. Psychology Press. ISBN: 978-1-315-80243-5.

Duane, Simon et al. (Sept. 1987). "Hybrid Monte Carlo". In: *Physics Letters B* 195.2, pp. 216–222. ISSN: 0370-2693. DOI: 10.1016/0370-2693(87)91197-X.

Dunning, David, Detlef Fetchenhauer, and Thomas Schlösser (June 2017). "The Varying Roles Played by Emotion in Economic Decision Making". In: *Current Opinion in Behavioral Sciences*. Mixed Emotions 15, pp. 33–38. ISSN: 2352-1546. DOI: 10.1016/j.cobeha.2017.05.006.

Ebert, Bjarke et al. (Aug. 1997). "Norketamine, the Main Metabolite of Ketamine, Is a Non-Competitive NMDA Receptor Antagonist in the Rat Cortex and Spinal Cord". In: *European Journal of Pharmacology* 333.1, pp. 99–104. ISSN: 0014-2999. DOI: 10.1016/S0014-2999(97)01116-3.

Eippert, Falk et al. (2007). "Regulation of Emotional Responses Elicited by Threat-Related Stimuli". In: *Human Brain Mapping* 28.5, pp. 409–423. ISSN: 1097-0193. DOI: 10.1002/hbm.20291.

Eldar, Eran and Yael Niv (Jan. 2015). "Interaction between Emotional State and Learning Underlies Mood Instability". In: *Nature Communications* 6.1, p. 6149. ISSN: 2041-1723. DOI: 10.1038/ncomms7149.

Eldar, Eran et al. (Jan. 2016). "Mood as Representation of Momentum". In: *Trends in Cognitive Sciences* 20.1, pp. 15–24. ISSN: 1364-6613, 1879-307X. DOI: 10.1016/j.tics.2015.07.010.

Eldar, Eran et al. (June 2020). "The Roles of Online and Offline Replay in Planning". In: *eLife* 9. Ed. by Thorsten Kahnt, Kate M Wassum, and Samuel J Gershman, e56911. ISSN: 2050-084X. DOI: 10.7554/eLife.56911.

Enkel, Thomas et al. (Mar. 2010). "Ambiguous-Cue Interpretation Is Biased Under Stress- and Depression-Like States in Rats". In: *Neuropsychopharmacology* 35.4, pp. 1008–1015. ISSN: 1740-634X. DOI: 10.1038/npp.2009.204.

Estes, W. K. (Mar. 2002). "Traps in the Route to Models of Memory and Decision". In: *Psychonomic Bulletin & Review* 9.1, pp. 3–25. ISSN: 1531-5320. DOI: 10.3758/BF03196254.

Fard, Pouyan R. et al. (May 2017). "A Bayesian Reformulation of the Extended Drift-Diffusion Model in Perceptual Decision Making". In: *Frontiers in Computational Neuroscience* 11. ISSN: 1662-5188. DOI: 10.3389/fncom.2017.00029.

Fink, Daniel (1997). "A Compendium of Conjugate Priors". In: *undefined*.

Fontanesi, Laura et al. (Aug. 2019). "A Reinforcement Learning Diffusion Decision Model for Value-Based Decisions". In: *Psychonomic Bulletin & Review* 26.4, pp. 1099–1121. ISSN: 1531-5320. DOI: 10.3758/s13423-018-1554-2.

Friston, Karl J. et al. (2014). "Computational Psychiatry: The Brain as a Phantastic Organ". In: *The Lancet Psychiatry* 1.2, pp. 148–158. ISSN: 2215-0374. DOI: 10.1016/S2215-0366(14)70275-5.

Ge, Hong, Kai Xu, and Zoubin Ghahramani (Mar. 2018). "Turing: A Language for Flexible Probabilistic Inference". In: *International Conference on Artificial Intelligence and Statistics*. PMLR, pp. 1682–1690.

Geller, Irving and Joseph Seifter (Nov. 1960). "The Effects of Meprobamate, Barbiturates, d-Amphetamine and Promazine on Experimentally Induced Conflict in the Rat". In: *Psychopharmacologia* 1.6, pp. 482–492. ISSN: 1432-2072. DOI: 10.1007/BF00429273.

Gelman, Andrew and Donald B. Rubin (Nov. 1992). "Inference from Iterative Simulation Using Multiple Sequences". In: *Statistical Science* 7.4, pp. 457–472. ISSN: 0883-4237, 2168-8745. DOI: 10.1214/ss/1177011136.

Gelman, Andrew et al. (July 2015). *Bayesian Data Analysis*. Third. New York: Chapman and Hall/CRC. ISBN: 978-0-429-11307-9. DOI: 10.1201/b16018.

Gelman, Andrew et al. (Nov. 2020). "Bayesian Workflow". In: *arXiv:2011.01808 [stat]*. arXiv: 2011.01808 [stat].

Gilboa-Schechtman, Eva et al. (Feb. 2004). "Depression Impairs the Ability to Ignore the Emotional Aspects of Facial Expressions: Evidence from the Garner Task". In: *Cognition & Emotion* 18.2, pp. 209–231. ISSN: 1464-0600. DOI: 10.1080/02699930341000176a.

Gomez-Marin, Alex and Asif A. Ghazanfar (Oct. 2019). "The Life of Behavior". In: *Neuron* 104.1, pp. 25–36. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2019.09.017.

Griesius, Simonas, Jack R. Mellor, and Emma Sj Robinson (Aug. 2020). "Comparison of Acute Treatment with Delayed-Onset versus Rapid-Acting Antidepressants on Effort-Related Choice Behaviour". In: *Psychopharmacology* 237.8, pp. 2381–2394. ISSN: 1432-2072. DOI: 10.1007/s00213-020-05541-9.

Gygax, Lorenz (2017). "Wanting, Liking and Welfare: The Role of Affective States in Proximate Control of Behaviour in Vertebrates". In: *Ethology* 123.10, pp. 689–704. ISSN: 1439-0310. DOI: 10.1111/eth.12655.

Hales, Claire A., Conor J. Houghton, and Emma S. J. Robinson (Dec. 2017). "Behavioural and Computational Methods Reveal Differential Effects for How Delayed and Rapid Onset Antidepressants Effect Decision Making in Rats". In: *European Neuropsychopharmacology: The Journal of the European College of Neu-*

ropsychopharmacology 27.12, pp. 1268–1280. ISSN: 1873-7862. DOI: 10.1016/j.euroneuro.2017.09.008.

Hales, Claire A., Emma S. J. Robinson, and Conor J. Houghton (2016). "Diffusion Modelling Reveals the Decision Making Processes Underlying Negative Judgement Bias in Rats". In: *PloS One* 11.3, e0152592. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0152592.

Hales, Claire Amanda (2016). "The Impact of Emotions on Decision Making". PhD thesis. University of Bristol.

Harding, Emma J., Elizabeth S. Paul, and Michael Mendl (Jan. 2004). "Cognitive Bias and Affective State". In: *Nature* 427.6972, pp. 312–312. ISSN: 1476-4687. DOI: 10.1038/427312a.

Harmer, C. J. et al. (Jan. 2003). "Acute SSRI Administration Affects the Processing of Social Cues in Healthy Volunteers". In: *Neuropsychopharmacology* 28.1, pp. 148–152. ISSN: 1740-634X. DOI: 10.1038/sj.npp.1300004.

Harmer, Catherine J. et al. (Oct. 2009). "Effect of Acute Antidepressant Administration on Negative Affective Bias in Depressed Patients". In: *The American Journal of Psychiatry* 166.10, pp. 1178–1184. ISSN: 1535-7228. DOI: 10.1176/appi.ajp.2009.09020149.

Haselton, Martie G. and Daniel Nettle (Feb. 2006). "The Paranoid Optimist: An Integrative Evolutionary Model of Cognitive Biases". In: *Personality and Social Psychology Review* 10.1, pp. 47–66. ISSN: 1088-8683. DOI: 10.1207/s15327957pspr1001_3.

Heal, David J et al. (June 2013). "Amphetamine, Past and Present – a Pharmacological and Clinical Perspective". In: *Journal of Psychopharmacology (Oxford, England)* 27.6, pp. 479–496. ISSN: 0269-8811. DOI: 10.1177/0269881113482532.

Heffner, Henry E. et al. (Mar. 1994). "Audiogram of the Hooded Norway Rat". In: *Hearing Research* 73.2, pp. 244–247. ISSN: 0378-5955. DOI: 10.1016/0378-5955(94)90240-2.

Hinchcliffe, Justyna K (Aug. 2019). "Investigating the Mechanisms Contributing to Affective Biases in Major Depressive Disorder". PhD thesis. University of Bristol.

Hinchcliffe, Justyna K, Megan G Jackson, and Emma SJ Robinson (Jan. 2022). "The Use of Ball Pits and Playpens in Laboratory Lister Hooded Male Rats Induces Ultrasonic Vocalisations Indicating a More Positive Affective State and Can Reduce the Welfare Impacts of Aversive Procedures". In: *Laboratory Animals*, p. 00236772211065920. ISSN: 0023-6772. DOI: 10.1177/00236772211065920.

Hirsch, C. R. and A. Mathews (Nov. 2000). "Impaired Positive Inferential Bias in Social Phobia". In: *Journal of Abnormal Psychology* 109.4, pp. 705–712. ISSN: 0021-843X. DOI: 10.1037//0021-843x.109.4.705.

Hoffman, Matthew D. and Andrew Gelman (2014). "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo". In: *Journal of Machine Learning Research* 15.47, pp. 1593–1623. ISSN: 1533-7928.

Huang, He, Wesley Thompson, and Martin P. Paulus (Sept. 2017). "Computational Dysfunctions in Anxiety: Failure to Differentiate Signal From Noise". In: *Biological Psychiatry*. Computational Psychiatry 82.6, pp. 440–446. ISSN: 0006-3223. DOI: 10.1016/j.biopsych.2017.07.007.

Hull, C. L. (1943). *Principles of Behavior: An Introduction to Behavior Theory*. Principles of Behavior: An Introduction to Behavior Theory. Oxford, England: Appleton-Century, pp. x, 422.

Huys, Quentin J. M., Tiago V. Maia, and Michael J. Frank (Mar. 2016). "Computational Psychiatry as a Bridge from Neuroscience to Clinical Applications". In: *Nature Neuroscience* 19.3, pp. 404–413. ISSN: 1546-1726. DOI: 10.1038/nn.4238.

Huys, Quentin J. M. et al. (July 2020). "Advances in the Computational Understanding of Mental Illness". In: *Neuropsychopharmacology*, pp. 1–17. ISSN: 1740-634X. DOI: 10.1038/s41386-020-0746-4.

Huys, Quentin Jm et al. (June 2013). "Mapping Anhedonia onto Reinforcement Learning: A Behavioural Meta-Analysis". In: *Biology of Mood & Anxiety Disorders* 3.1, p. 12. ISSN: 2045-5380. DOI: 10.1186/2045-5380-3-12.

Iigaya, Kiyohito et al. (Nov. 2016). "Cognitive Bias in Ambiguity Judgements: Using Computational Models to Dissect the Effects of Mild Mood Manipulation in Humans". In: *PLOS ONE* 11.11, e0165840. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0165840.

Ivan, Victorita E. et al. (2018). "Lose-Shift Responding in Humans Is Promoted by Increased Cognitive Load". In: *Frontiers in Integrative Neuroscience* 12. ISSN: 1662-5145.

Jaric, Ivana et al. (Feb. 2022). *Rearing Environment Persistently Modulates the Phenotype of Mice*. DOI: 10.1101/2022.02.11.480070.

Ji, Guangchen et al. (2018 Jan-Dec). "Fear Extinction Learning Ability Predicts Neuropathic Pain Behaviors and Amygdala Activity in Male Rats". In: *Molecular Pain* 14, p. 1744806918804441. ISSN: 1744-8069. DOI: 10.1177/1744806918804441.

Jocham, Gerhard et al. (Apr. 2016). "Reward-Guided Learning with and without Causal Attribution". In: *Neuron* 90.1, pp. 177–190. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2016.02.018.

Jones, Samantha et al. (Aug. 2018). "Assessing Animal Affect: An Automated and Self-Initiated Judgement Bias Task Based on Natural Investigative Behaviour". In: *Scientific Reports* 8.1, p. 12400. ISSN: 2045-2322. DOI: 10.1038/s41598-018-30571-x.

Kahneman, Daniel and Dale T. Miller (1986). "Norm Theory: Comparing Reality to Its Alternatives". In: *Psychological Review* 93.2, pp. 136–153. ISSN: 1939-1471. DOI: 10.1037/0033-295X.93.2.136.

Kamin, L. J., C. J. Brimer, and A. H. Black (1963). "Conditioned Suppression as a Monitor of Fear of the CS in the Course of Avoidance Training". In: *Journal of Comparative and Physiological Psychology* 56.3, pp. 497–501. ISSN: 0021-9940. DOI: 10.1037/h0047966.

Kamiyama, H. et al. (Mar. 2011). "Mechanisms Underlying Ketamine-Induced Synaptic Depression in Rat Hippocampus-Medial Prefrontal Cortex Pathway". In: *Neuroscience* 177, pp. 159–169. ISSN: 0306-4522. DOI: 10.1016/j.neuroscience.2010.12.012.

Kelter, Riko (Apr. 2020). "Analysis of Bayesian Posterior Significance and Effect Size Indices for the Two-Sample t-Test to Support Reproducible Medical Research". In: *BMC Medical Research Methodology* 20.1, p. 88. ISSN: 1471-2288. DOI: 10.1186/s12874-020-00968-2.

Kim, Hackjin, Shinsuke Shimojo, and John P. O'Doherty (July 2006). "Is Avoiding an Aversive Outcome Rewarding? Neural Substrates of Avoidance Learning in the Human Brain". In: *PLOS Biology* 4.8, e233. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.0040233.

Kruschke, John K. (June 2018). "Rejecting or Accepting Parameter Values in Bayesian Estimation". In: *Advances in Methods and Practices in Psychological Science* 1.2, pp. 270–280. ISSN: 2515-2459. DOI: 10.1177/2515245918771304.

Kumar, Ravin et al. (Jan. 2019). "ArviZ a Unified Library for Exploratory Analysis of Bayesian Models in Python". In: *Journal of Open Source Software* 4.33, p. 1143. ISSN: 2475-9066. DOI: 10.21105/joss.01143.

Kunisato, Yoshihiko et al. (Dec. 2012). "Effects of Depression on Reward-Based Decision Making and Variability of Action in Probabilistic Learning". In: *Journal of Behavior Therapy and Experimental Psychiatry* 43.4, pp. 1088–1094. ISSN: 1873-7943. DOI: 10.1016/j.jbtep.2012.05.007.

Laughlin, Lindsay C. et al. (July 2020). "Reducing Shock Imminence Eliminates Poor Avoidance in Rats". In: *Learning & Memory* 27.7, pp. 270–274. ISSN: 1072-0502, 1549-5485. DOI: 10.1101/lm.051557.120.

Lazarus, Richard S. (1982). "Thoughts on the Relations between Emotion and Cognition". In: *American Psychologist* 37.9, pp. 1019–1024. ISSN: 1935-990X. DOI: 10.1037/0003-066X.37.9.1019.

LeDoux, J. E. et al. (Jan. 2017). "The Birth, Death and Resurrection of Avoidance: A Reconceptualization of a Troubled Paradigm". In: *Molecular Psychiatry* 22.1, pp. 24–36. ISSN: 1476-5578. DOI: 10.1038/mp.2016.166.

Leeuw, Jan de and Erik Meijer (2008). "Introduction to Multilevel Analysis". In: *Handbook of Multilevel Analysis*. Ed. by Jan de Leeuw and Erik Meijer. New York, NY: Springer, pp. 1–75. ISBN: 978-0-387-73186-5. DOI: 10.1007/978-0-387-73186-5_1.

Leppänen, Jukka M. et al. (Sept. 2004). "Depression Biases the Recognition of Emotionally Neutral Faces". In: *Psychiatry Research* 128.2, pp. 123–133. ISSN: 0165-1781. DOI: 10.1016/j.psychres.2004.05.020.

Ma, Liya et al. (Aug. 2015). "Ketamine-Induced Changes in the Signal and Noise of Rule Representation in Working Memory by Lateral Prefrontal Neurons". In: *Journal of Neuroscience* 35.33, pp. 11612–11622. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.1839-15.2015.

MacKay, David J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Copyright Cambridge University Press.

MacLeod, C. and E. M. Rutherford (Sept. 1992). "Anxiety and the Selective Processing of Emotional Information: Mediating Roles of Awareness, Trait and State Variables, and Personal Relevance of Stimulus Materials". In: *Behaviour Research and Therapy* 30.5, pp. 479–491. ISSN: 0005-7967. DOI: 10.1016/0005-7967(92)90032-c.

MacLeod, Colin and Andrew Mathews (2004). "Selective Memory Effects in Anxiety Disorders: An Overview of Research Findings and Their Implications". In: *Memory and Emotion*. Series in Affective Science. New York, NY, US: Oxford University Press, pp. 155–185. ISBN: 978-0-19-515856-4. DOI: 10.1093/acprof:oso/9780195158564.003.0005.

Maia, Tiago V. (Feb. 2010). "Two-Factor Theory, the Actor-Critic Model, and Conditioned Avoidance". In: *Learning & Behavior* 38.1, pp. 50–67. ISSN: 1532-5830. DOI: 10.3758/LB.38.1.50.

Mathews, Andrew and Colin MacLeod (2005). "Cognitive Vulnerability to Emotional Disorders". In: *Annual Review of Clinical Psychology* 1.1, pp. 167–195. DOI: 10.1146/annurev.clinpsy.1.102803.143916.

Mathews, Andrew, Valerie Ridgeway, and Donald A. Williamson (Sept. 1996). "Evidence for Attention to Threatening Stimuli in Depression". In: *Behaviour Re-*

search and Therapy 34.9, pp. 695–705. ISSN: 0005-7967. DOI: 10.1016/0005-7967(96)00046-0.

Mattar, Marcelo G. and Nathaniel D. Daw (Nov. 2018). "Prioritized Memory Access Explains Planning and Hippocampal Replay". In: *Nature Neuroscience* 21.11, pp. 1609–1617. ISSN: 1546-1726. DOI: 10.1038/s41593-018-0232-z.

McCullagh, P. and J. A. Nelder (Jan. 2019). *Generalized Linear Models*. Second. New York: Routledge. ISBN: 978-0-203-75373-6. DOI: 10.1201/9780203753736.

McElreath, Richard (Dec. 2016). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. New York: Chapman and Hall/CRC. ISBN: 978-1-315-37249-5. DOI: 10.1201/9781315372495.

Medin, Douglas L. (1972). "Role of Reinforcement in Discrimination Learning Set in Monkeys". In: *Psychological Bulletin* 77.5, pp. 305–318. ISSN: 1939-1455. DOI: 10.1037/h0032548.

Mendl, Michael and Elizabeth S. Paul (May 2020). "Animal Affect and Decision-Making". In: *Neuroscience & Biobehavioral Reviews* 112, pp. 144–163. ISSN: 0149-7634. DOI: 10.1016/j.neubiorev.2020.01.025.

Mendl, Michael et al. (May 2009). "Cognitive Bias as an Indicator of Animal Emotion and Welfare: Emerging Evidence and Underlying Mechanisms". In: *Applied Animal Behaviour Science*. Special Issue: Animal Suffering and Welfare 118.3, pp. 161–181. ISSN: 0168-1591. DOI: 10.1016/j.applanim.2009.02.023.

Mion, Georges and Thierry Villevieille (2013). "Ketamine Pharmacology: An Update (Pharmacodynamics and Molecular Aspects, Recent Findings)". In: *CNS Neuroscience & Therapeutics* 19.6, pp. 370–380. ISSN: 1755-5949. DOI: 10.1111/cns.12099.

Montague, P. R., P. Dayan, and T. J. Sejnowski (Mar. 1996). "A Framework for Mesencephalic Dopamine Systems Based on Predictive Hebbian Learning". In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 16.5, pp. 1936–1947. ISSN: 0270-6474.

Moutoussis, Michael et al. (Jan. 2008). "A Temporal Difference Account of Avoidance Learning". In: *Network: Computation in Neural Systems* 19.2, pp. 137–160. ISSN: 0954-898X. DOI: 10.1080/09548980802192784.

Murphy, F. C. et al. (Apr. 2003). "Neuropsychological Impairment in Patients with Major Depressive Disorder: The Effects of Feedback on Task Performance". In: *Psychological Medicine* 33.3, pp. 455–467. ISSN: 1469-8978, 0033-2917. DOI: 10.1017/S0033291702007018.

Neville, Vikki et al. (Jan. 2020a). "Pharmacological Manipulations of Judgement Bias: A Systematic Review and Meta-Analysis". In: *Neuroscience and Biobehav-*

ioral Reviews 108, pp. 269–286. ISSN: 1873-7528. DOI: 10.1016/j.neubiorev.2019.11.008.

Neville, Vikki et al. (July 2020b). "Reward and Punisher Experience Alter Rodent Decision-Making in a Judgement Bias Task". In: *Scientific Reports* 10.1, p. 11839. ISSN: 2045-2322. DOI: 10.1038/s41598-020-68737-1.

Neville, Vikki et al. (Jan. 2021). "Dissecting the Links between Reward and Loss, Decision-Making, and Self-Reported Affect Using a Computational Approach". In: *PLOS Computational Biology* 17.1, e1008555. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1008555.

Nguyen, Ho A. T., Chao Guo, and Judith R. Homberg (2020). "Cognitive Bias Under Adverse and Rewarding Conditions: A Systematic Review of Rodent Studies". In: *Frontiers in Behavioral Neuroscience* 14. ISSN: 1662-5153.

Niv, Yael (June 2009). "Reinforcement Learning in the Brain". In: *Journal of Mathematical Psychology*. Special Issue: Dynamic Decision Making 53.3, pp. 139–154. ISSN: 0022-2496. DOI: 10.1016/j.jmp.2008.12.005.

Noworyta-Sokolowska, Karolina et al. (Aug. 2019). "Sensitivity to Negative and Positive Feedback as a Stable and Enduring Behavioural Trait in Rats". In: *Psychopharmacology* 236.8, pp. 2389–2403. ISSN: 1432-2072. DOI: 10.1007/s00213-019-05333-w.

Nusbaum, Amy T. et al. (July 2018). "Induced Positive Mood and Cognitive Flexibility: Evidence from Task Switching and Reversal Learning". In: *Collabra: Psychology* 4.1. Ed. by Rolf Zwaan and Christopher Madan, p. 25. ISSN: 2474-7394. DOI: 10.1525/collabra.150.

Nygren, Thomas E. et al. (Apr. 1996). "The Influence of Positive Affect on the Decision Rule in Risk Situations: Focus on Outcome (and Especially Avoidance of Loss) Rather Than Probability". In: *Organizational Behavior and Human Decision Processes* 66.1, pp. 59–72. ISSN: 0749-5978. DOI: 10.1006/obhd.1996.0038.

Otto, A. Ross, Stephen M. Fleming, and Paul W. Glimcher (Mar. 2016). "Unexpected but Incidental Positive Outcomes Predict Real-World Gambling". In: *Psychological Science* 27.3, pp. 299–311. ISSN: 0956-7976. DOI: 10.1177/0956797615618366.

Palm, Sara et al. (2011). "When Is a Wistar a Wistar? Behavioral Profiling of Outbred Wistar Rats from Five Different Suppliers Using the MCSF Test". In: *Applied Animal Behaviour Science* 1–2.135, pp. 128–137. ISSN: 0168-1591. DOI: 10.1016/j.applanim.2011.08.010.

Palminteri, Stefano et al. (Aug. 2015). "Contextual Modulation of Value Signals in Reward and Punishment Learning". In: *Nature Communications* 6.1, p. 8096. ISSN: 2041-1723. DOI: 10.1038/ncomms9096.

Papciak, Justyna et al. (Nov. 2013). "Chronic Psychosocial Stress Makes Rats More 'pessimistic' in the Ambiguous-Cue Interpretation Paradigm". In: *Behavioural Brain Research* 256, pp. 305–310. ISSN: 1872-7549. DOI: 10.1016/j.bbr.2013.08.036.

Pecoraro, N. C., W. D. Timberlake, and M. Tinsley (Apr. 1999). "Incentive Downshifts Evoke Search Repertoires in Rats". In: *Journal of Experimental Psychology. Animal Behavior Processes* 25.2, pp. 153–167. ISSN: 0097-7403.

Pflugshaupt, Tobias et al. (Jan. 2005). "Hypervigilance–Avoidance Pattern in Spider Phobia". In: *Journal of Anxiety Disorders* 19.1, pp. 105–116. ISSN: 0887-6185. DOI: 10.1016/j.janxdis.2003.12.002.

Pietersen, C. Y. et al. (Apr. 2006). "Fear Conditioning and Shock Intensity: The Choice between Minimizing the Stress Induced and Reducing the Number of Animals Used". In: *Laboratory Animals* 40.2, pp. 180–185. ISSN: 0023-6772. DOI: 10.1258/002367706776319006.

Pike, Alexandra C. and Oliver J. Robinson (Mar. 2022). "Reinforcement Learning in Patients With Mood and Anxiety Disorders vs Control Individuals: A Systematic Review and Meta-analysis". In: *JAMA Psychiatry*. ISSN: 2168-622X. DOI: 10.1001/jamapsychiatry.2022.0051.

Pisupati, Sashank et al. (Jan. 2021). "Lapses in Perceptual Decisions Reflect Exploration". In: *eLife* 10. Ed. by Daeyeol Lee et al., e55490. ISSN: 2050-084X. DOI: 10.7554/eLife.55490.

Pizzagalli, Diego A., Allison L. Jahn, and James P. O'Shea (Feb. 2005). "Toward an Objective Characterization of an Anhedonic Phenotype: A Signal-Detection Approach". In: *Biological Psychiatry* 57.4, pp. 319–327. ISSN: 0006-3223. DOI: 10.1016/j.biopsych.2004.11.026.

Pizzagalli, Diego A. et al. (Nov. 2008). "Reduced Hedonic Capacity in Major Depressive Disorder: Evidence from a Probabilistic Reward Task". In: *Journal of Psychiatric Research* 43.1, pp. 76–87. ISSN: 0022-3956. DOI: 10.1016/j.jpsychires.2008.03.001.

Pulcu, Erdem and Michael Browning (Oct. 2017). "Affective Bias as a Rational Response to the Statistics of Rewards and Punishments". In: *eLife* 6. Ed. by Michael J Frank, e27879. ISSN: 2050-084X. DOI: 10.7554/eLife.27879.

– (Oct. 2019). "The Misestimation of Uncertainty in Affective Disorders". In: *Trends in Cognitive Sciences* 23.10, pp. 865–875. ISSN: 1364-6613, 1879-307X. DOI: 10.1016/j.tics.2019.07.007.

Pury, Cynthia L. S. and Susan Mineka (Sept. 2001). "Differential Encoding of Affective and Nonaffective Content Information in Trait Anxiety". In: *Cognition and Emotion* 15.5, pp. 659–693. ISSN: 0269-9931. DOI: 10.1080/02699930126123.

Quintana, Daniel S. and Donald R. Williams (June 2018). "Bayesian Alternatives for Common Null-Hypothesis Significance Tests in Psychiatry: A Non-Technical Guide Using JASP". In: *BMC Psychiatry* 18.1, p. 178. ISSN: 1471-244X. DOI: 10.1186/s12888-018-1761-4.

Raedt, R. and E. Koster (2010). "Understanding Vulnerability for Depression from a Cognitive Neuroscience Perspective: A Reappraisal of Attentional Factors and a New Conceptual Framework". In: *Cognitive, affective & behavioral neuroscience*. DOI: 10.3758/CABN.10.1.50.

Ratcliff, Roger and Gail McKoon (Apr. 2008). "The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks". In: *Neural computation* 20.4, pp. 873–922. ISSN: 0899-7667. DOI: 10.1162/neco.2008.12-06-420.

Rayburn-Reeves, Rebecca M. et al. (May 2013). "Reversal Learning in Rats (Rattus Norvegicus) and Pigeons (Columba Livia): Qualitative Differences in Behavioral Flexibility". In: *Journal of Comparative Psychology (Washington, D.C.: 1983)* 127.2, pp. 202–211. ISSN: 1939-2087. DOI: 10.1037/a0026311.

Reed, Phil (2016). "Win-Stay and Win-Shift Lever-Press Strategies in an Appetitively Reinforced Task for Rats". In: *Learning & Behavior* 44.4, pp. 340–346. ISSN: 1543-4494. DOI: 10.3758/s13420-016-0225-2.

Richards, A. et al. (Nov. 1992). "Effects of Mood Manipulation and Anxiety on Performance of an Emotional Stroop Task". In: *British Journal of Psychology (London, England: 1953)* 83 (Pt 4), pp. 479–491. ISSN: 0007-1269. DOI: 10.1111/j.2044-8295.1992.tb02454.x.

Ridout, Nathan et al. (Jan. 2003). "Memory Bias for Emotional Facial Expressions in Major Depression". In: *Cognition and Emotion* 17.1, pp. 101–122. ISSN: 0269-9931. DOI: 10.1080/02699930302272.

Robbins, Herbert (1952). "Some Aspects of the Sequential Design of Experiments". In: *Bulletin of the American Mathematical Society* 58.5, pp. 527–535. ISSN: 0002-9904, 1936-881X. DOI: 10.1090/S0002-9904-1952-09620-8.

Robinson, E.S.J. and J.P. Roiser (2016). "Affective Biases in Humans and Animals". In: *Translational Neuropsychopharmacology*. Ed. by Trevor W. Robbins and Barbara J. Sahakian. Current Topics in Behavioral Neurosciences. Cham: Springer International Publishing, pp. 263–286. ISBN: 978-3-319-33913-9. DOI: 10.1007/7854_2015_5011.

Roelofs, Sanne et al. (June 2016). "Making Decisions under Ambiguity: Judgment Bias Tasks for Assessing Emotional State in Animals". In: *Frontiers in Behavioral Neuroscience* 10. ISSN: 1662-5153. DOI: 10.3389/fnbeh.2016.00119.

Roy, Nicholas A. et al. (Feb. 2021). "Extracting the Dynamics of Behavior in Sensory Decision-Making Experiments". In: *Neuron* 109.4, 597–610.e6. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2020.12.004.

Russell, James A. (Jan. 2003). "Core Affect and the Psychological Construction of Emotion". In: *Psychological Review* 110.1, pp. 145–172. ISSN: 0033-295X. DOI: 10.1037/0033-295x.110.1.145.

Rutledge, Robb B. et al. (Aug. 2014). "A Computational and Neural Model of Momentary Subjective Well-Being". In: *Proceedings of the National Academy of Sciences* 111.33, pp. 12252–12257. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1407535111.

Rutledge, Robb B. et al. (Aug. 2017). "Association of Neural and Emotional Impacts of Reward Prediction Errors With Major Depression". In: *JAMA psychiatry* 74.8, pp. 790–797. ISSN: 2168-6238. DOI: 10.1001/jamapsychiatry.2017.1713.

Rychlik, Michal, Eva Bollen, and Rafal Rygula (Feb. 2017). "Ketamine Decreases Sensitivity of Male Rats to Misleading Negative Feedback in a Probabilistic Reversal-Learning Task". In: *Psychopharmacology* 234.4, pp. 613–620. ISSN: 1432-2072. DOI: 10.1007/s00213-016-4497-1.

Rygula, Rafal, Justyna Papciak, and Piotr Popik (July 2014). "The Effects of Acute Pharmacological Stimulation of the 5-HT, NA and DA Systems on the Cognitive Judgement Bias of Rats in the Ambiguous-Cue Interpretation Paradigm". In: *European Neuropsychopharmacology: The Journal of the European College of Neuropsychopharmacology* 24.7, pp. 1103–1111. ISSN: 1873-7862. DOI: 10.1016/j.euroneuro.2014.01.012.

Salamone, John D. et al. (May 2016a). "Activational and Effort-Related Aspects of Motivation: Neural Mechanisms and Implications for Psychopathology". In: *Brain* 139.5, pp. 1325–1347. ISSN: 0006-8950. DOI: 10.1093/brain/aww050.

Salamone, John D. et al. (June 2016b). "The Pharmacology of Effort-Related Choice Behavior: Dopamine, Depression, and Individual Differences". In: *Behavioural Processes*. SQAB 2015: Choice and Consequences 127, pp. 3–17. ISSN: 0376-6357. DOI: 10.1016/j.beproc.2016.02.008.

Saland, Samantha K. and Mohamed Kabbaj (Dec. 2018). "Sex Differences in the Pharmacokinetics of Low-dose Ketamine in Plasma and Brain of Male and Female Rats". In: *Journal of Pharmacology and Experimental Therapeutics* 367.3, pp. 393–404. ISSN: 0022-3565, 1521-0103. DOI: 10.1124/jpet.118.251652.

Salvador, Alexandre et al. (Jan. 2022). "Premature Commitment to Uncertain Decisions during Human NMDA Receptor Hypofunction". In: *Nature Communications* 13.1, p. 338. ISSN: 2041-1723. DOI: 10.1038/s41467-021-27876-3.

Saravanan, Varun, Gordon J. Berman, and Samuel J. Sober (July 2020). "Application of the Hierarchical Bootstrap to Multi-Level Data in Neuroscience". In: *arXiv:2007.07797 [q-bio]*. arXiv: 2007.07797 [q-bio].

Scholl, Jacqueline and Miriam Klein-Flügge (Dec. 2018). "Understanding Psychiatric Disorder by Capturing Ecologically Relevant Features of Learning and Decision-Making". In: *Behavioural Brain Research* 355, pp. 56–75. ISSN: 0166-4328. DOI: 10.1016/j.bbr.2017.09.050.

Scholl, Jacqueline et al. (Aug. 2015). "The Good, the Bad, and the Irrelevant: Neural Mechanisms of Learning Real and Hypothetical Rewards and Effort". In: *Journal of Neuroscience* 35.32, pp. 11233–11251. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.0396-15.2015.

Seymour, Ben and Samuel M McClure (Apr. 2008). "Anchors, Scales and the Relative Coding of Value in the Brain". In: *Current Opinion in Neurobiology*. Cognitive Neuroscience 18.2, pp. 173–178. ISSN: 0959-4388. DOI: 10.1016/j.conb.2008.07.010.

Shadmehr, Reza and Alaa A. Ahmed (July 2020). *Vigor: Neuroeconomics of Movement Control*. Cambridge, MA, USA: MIT Press. ISBN: 978-0-262-04405-9.

Shafiei, Naghmeh et al. (Sept. 2012). "Acute Stress Induces Selective Alterations in Cost/Benefit Decision-Making". In: *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology* 37.10, pp. 2194–2209. ISSN: 1740-634X. DOI: 10.1038/npp.2012.69.

Sharot, Tali et al. (Aug. 2012). "How Dopamine Enhances an Optimism Bias in Humans". In: *Current Biology* 22.16, pp. 1477–1481. ISSN: 0960-9822. DOI: 10.1016/j.cub.2012.05.053.

Slaney, Chloe, Justyna K. Hinchcliffe, and Emma S. J. Robinson (2018a). "Translational Shifts in Preclinical Models of Depression: Implications for Biomarkers for Improved Treatments". In: *Biomarkers in Psychiatry*. Ed. by Judith Pratt and Jeremy Hall. Current Topics in Behavioral Neurosciences. Cham: Springer International Publishing, pp. 169–193. ISBN: 978-3-319-99642-4. DOI: 10.1007/7854_2018_44.

Slaney, Chloe L., Claire A. Hales, and Emma S. J. Robinson (Aug. 2018b). "Rat Models of Reward Deficits in Psychiatric Disorders". In: *Current Opinion in Behavioral Sciences* 22, pp. 136–142. ISSN: 2352-1546. DOI: 10.1016/j.cobeha.2018.05.001.

Strickland, JA et al. (Feb. 2021). "Foot Shock Facilitates Reward Seeking in an Experience-Dependent Manner". In: *Behavioural Brain Research* 399, p. 112974. ISSN: 0166-4328. DOI: 10.1016/j.bbr.2020.112974.

Stuart, Sarah A., Justyna K. Hinchcliffe, and Emma S. J. Robinson (Aug. 2019). "Evidence That Neuropsychological Deficits Following Early Life Adversity May Underlie Vulnerability to Depression". In: *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology* 44.9, pp. 1623–1630. ISSN: 1740-634X. DOI: 10.1038/s41386-019-0388-6.

Stuart, Sarah A. and Emma S. J. Robinson (Sept. 2015a). "Reducing the Stress of Drug Administration: Implications for the 3Rs". In: *Scientific Reports* 5.1, p. 14288. ISSN: 2045-2322. DOI: 10.1038/srep14288.

Stuart, Sarah A. et al. (Aug. 2013). "A Translational Rodent Assay of Affective Biases in Depression and Antidepressant Therapy". In: *Neuropsychopharmacology* 38.9, pp. 1625–1635. ISSN: 1740-634X. DOI: 10.1038/npp.2013.69.

Stuart, Sarah A. et al. (Aug. 2015b). "Distinct Neuropsychological Mechanisms May Explain Delayed- Versus Rapid-Onset Antidepressant Efficacy". In: *Neuropsychopharmacology* 40.9, pp. 2165–2174. ISSN: 1740-634X. DOI: 10.1038/npp.2015.59.

Sutton, S. Richard and G. Andrew Barto (2020). *Reinforcement Learning: An Introduction*.

Syka, J. et al. (Oct. 1996). "Auditory Frequency and Intensity Discrimination in Pigmented Rats". In: *Hearing Research* 100.1-2, pp. 107–113. ISSN: 0378-5955. DOI: 10.1016/0378-5955(96)00101-3.

Taylor Tavares, Joana V. et al. (Sept. 2008). "Neural Basis of Abnormal Response to Negative Feedback in Unmedicated Mood Disorders". In: *NeuroImage* 42.3, pp. 1118–1126. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2008.05.049.

Thoern, Hanna A. et al. (Mar. 2016). "Attentional Bias towards Positive Emotion Predicts Stress Resilience". In: *PLOS ONE* 11.3, e0148368. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0148368.

Thomas, Brian L, Niccole Larsen, and John J. B Ayres (Nov. 2003). "Role of Context Similarity in ABA, ABC, and AAB Renewal Paradigms: Implications for Theories of Renewal and for Treating Human Phobias". In: *Learning and Motivation* 34.4, pp. 410–436. ISSN: 0023-9690. DOI: 10.1016/S0023-9690(03)00037-7.

Thomas, Brian L., Craig L. Longo, and John J. B. Ayres (Nov. 2005). "Thwarting the Renewal (Relapse) of Conditioned Fear with the Explicitly Unpaired Procedure: Possible Interpretations and Implications for Treating Human Fears and Phobias".

In: *Learning and Motivation* 36.4, pp. 374–407. ISSN: 0023-9690. DOI: 10.1016/j.lmot.2004.11.005.

Thorndike, Edward L. (1933). "A Proof of the Law of Effect". In: *Science* 77.1989, pp. 173–175. ISSN: 0036-8075.

Todder, Doron, Serdal Caliskan, and Bernhard T. Baune (2009). "Longitudinal Changes of Day-Time and Night-Time Gross Motor Activity in Clinical Responders and Non-Responders of Major Depression". In: *The World Journal of Biological Psychiatry: The Official Journal of the World Federation of Societies of Biological Psychiatry* 10.4, pp. 276–284. ISSN: 1814-1412. DOI: 10.3109/15622970701403081.

Treadway, Michael T. et al. (Aug. 2009). "Worth the 'EEfRT'? The Effort Expenditure for Rewards Task as an Objective Measure of Motivation and Anhedonia". In: *PLOS ONE* 4.8, e6598. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0006598.

Treadway, Michael T. et al. (2012). "Effort-Based Decision-Making in Major Depressive Disorder: A Translational Model of Motivational Anhedonia". In: *Journal of Abnormal Psychology* 121.3, pp. 553–558. ISSN: 1939-1846. DOI: 10.1037/a0028813.

Trimmer, Pete C. et al. (Sept. 2013). "On the Evolution and Optimality of Mood States". In: *Behavioral Sciences* 3.3, pp. 501–521. DOI: 10.3390/bs3030501.

Tye, Kay M. (Oct. 2018). "Neural Circuit Motifs in Valence Processing". In: *Neuron* 100.2, pp. 436–452. ISSN: 08966273. DOI: 10.1016/j.neuron.2018.10.001.

Valton, Vincent, Toby Wise, and Oliver J. Robinson (Nov. 2020). "Recommendations for Bayesian Hierarchical Model Specifications for Case-Control Studies in Mental Health". In: *arXiv:2011.01725 [cs, stat]*. arXiv: 2011.01725 [cs, stat].

van de Schoot, Rens et al. (Jan. 2021). "Bayesian Statistics and Modelling". In: *Nature Reviews Methods Primers* 1.1, pp. 1–26. ISSN: 2662-8449. DOI: 10.1038/s43586-020-00001-2.

van Dooren, Roel et al. (June 2021). "The Exploration-Exploitation Trade-off in a Foraging Task Is Affected by Mood-Related Arousal and Valence". In: *Cognitive, Affective, & Behavioral Neuroscience* 21.3, pp. 549–560. ISSN: 1531-135X. DOI: 10.3758/s13415-021-00917-6.

van Rooij, Iris and Mark Blokpoel (Sept. 2020). "Formalizing Verbal Theories". In: *Social Psychology* 51.5, pp. 285–298. ISSN: 1864-9335. DOI: 10.1027/1864-9335/a000428.

Vehtari, Aki, Andrew Gelman, and Jonah Gabry (Sept. 2017). "Practical Bayesian Model Evaluation Using Leave-One-out Cross-Validation and WAIC". In: *Statis-*

tics and Computing 27.5, pp. 1413–1432. ISSN: 0960-3174, 1573-1375. DOI: 10.1007/s11222-016-9696-4. arXiv: 1507.04544.

Vehtari, Aki et al. (Feb. 2021a). "Pareto Smoothed Importance Sampling". In: *arXiv:1507.02646 [stat]*. arXiv: 1507.02646 [stat].

Vehtari, Aki et al. (June 2021b). "Rank-Normalization, Folding, and Localization: An Improved R-hat for Assessing Convergence of MCMC". In: *Bayesian Analysis* 16.2. ISSN: 1936-0975. DOI: 10.1214/20-BA1221. arXiv: 1903.08008.

Villano, William J. et al. (Sept. 2020). "Temporal Dynamics of Real-World Emotion Are More Strongly Linked to Prediction Error than Outcome". In: *Journal of Experimental Psychology. General* 149.9, pp. 1755–1766. ISSN: 1939-2222. DOI: 10.1037/xge0000740.

Vogel, J. R., B. Beer, and D. E. Clody (1971). "A Simple and Reliable Conflict Procedure for Testing Anti-Anxiety Agents". In: *Psychopharmacologia* 21.1, pp. 1–7. DOI: 10.1007/BF00403989.

Vrieze, Elske et al. (Apr. 2013). "Reduced Reward Learning Predicts Outcome in Major Depressive Disorder". In: *Biological Psychiatry* 73.7, pp. 639–645. ISSN: 1873-2402. DOI: 10.1016/j.biopsych.2012.10.014.

Wagenmakers, Eric-Jan and Scott Brown (2007). "On the Linear Relation between the Mean and the Standard Deviation of a Response Time Distribution". In: *Psychological Review* 114.3, pp. 830–841. ISSN: 1939-1471. DOI: 10.1037/0033-295X.114.3.830.

Wardle, Margaret C. et al. (Nov. 2011). "Amping Up Effort: Effects of d-Amphetamine on Human Effort-Based Decision-Making". In: *Journal of Neuroscience* 31.46, pp. 16597–16602. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.4387-11.2011.

Watkins, Christopher J. C. H. and Peter Dayan (May 1992). "Q-Learning". In: *Machine Learning* 8.3, pp. 279–292. ISSN: 1573-0565. DOI: 10.1007/BF00992698.

Watkins, Philip C. (May 2002). "Implicit Memory Bias in Depression". In: *Cognition and Emotion* 16.3, pp. 381–402. ISSN: 0269-9931. DOI: 10.1080/02699930143000536.

Whiteley, Louise and Maneesh Sahani (Mar. 2008). "Implicit Knowledge of Visual Uncertainty Guides Decisions with Asymmetric Outcomes". In: *Journal of Vision* 8.3, p. 2. ISSN: 1534-7362. DOI: 10.1167/8.3.2.

Wilkinson, Matthew P. et al. (Jan. 2020). "Ketamine Decreases Sensitivity of Male Rats to Misleading Negative Feedback in a Probabilistic Reversal-Learning Task". In: *Brain and Neuroscience Advances* 4, p. 2398212820907177. ISSN: 2398-2128. DOI: 10.1177/2398212820907177.

Wilson, Robert C and Anne GE Collins (Nov. 2019). "Ten Simple Rules for the Computational Modeling of Behavioral Data". In: *eLife* 8. Ed. by Timothy E Behrens, e49547. ISSN: 2050-084X. DOI: 10.7554/eLife.49547.

Winkielman, Piotr and Kent C. Berridge (June 2004). "Unconscious Emotion". In: *Current Directions in Psychological Science* 13.3, pp. 120–123. ISSN: 0963-7214. DOI: 10.1111/j.0963-7214.2004.00288.x.

Winkielman, Piotr, Kent C. Berridge, and Julia L. Wilbarger (Jan. 2005). "Unconscious Affective Reactions to Masked Happy Versus Angry Faces Influence Consumption Behavior and Judgments of Value". In: *Personality and Social Psychology Bulletin* 31.1, pp. 121–135. ISSN: 0146-1672. DOI: 10.1177/0146167204271309.

Wittmann, Marco K. et al. (Aug. 2016). "Predictive Decision Making Driven by Multiple Time-Linked Reward Representations in the Anterior Cingulate Cortex". In: *Nature Communications* 7.1, p. 12327. ISSN: 2041-1723. DOI: 10.1038/ncomms12327.

Wittmann, Marco K. et al. (July 2020). "Global Reward State Affects Learning and Activity in Raphe Nucleus and Anterior Insula in Monkeys". In: *Nature Communications* 11.1, p. 3771. ISSN: 2041-1723. DOI: 10.1038/s41467-020-17343-w.

Worthy, Darrell A. and W. Todd Maddox (2012). "Age-Based Differences in Strategy Use in Choice Tasks". In: *Frontiers in Neuroscience* 5, p. 145. ISSN: 1662-453X. DOI: 10.3389/fnins.2011.00145.

Wright, Kristina M and Michael A McDannald (Mar. 2019). "Ventrolateral Periaqueductal Gray Neurons Prioritize Threat Probability over Fear Output". In: *eLife* 8. Ed. by Geoffrey Schoenbaum et al., e45013. ISSN: 2050-084X. DOI: 10.7554/eLife.45013.

Wright, William F and Gordon H Bower (July 1992). "Mood Effects on Subjective Probability Assessment". In: *Organizational Behavior and Human Decision Processes* 52.2, pp. 276–291. ISSN: 0749-5978. DOI: 10.1016/0749-5978(92)90039-A.

Yang, Xin-Hua et al. (Dec. 2014). "Motivational Deficits in Effort-Based Decision Making in Individuals with Subsyndromal Depression, First-Episode and Remitted Depression Patients". In: *Psychiatry Research* 220.3, pp. 874–882. ISSN: 1872-7123. DOI: 10.1016/j.psychres.2014.08.056.

Zajonc, R. (1984). "On the Primacy of Affect." In: DOI: 10.1037/0003-066X.39.2.117.