Universidade de Aveiro
2020

Departamento de Psicologia e Educação

CRISTINA JOY DOS SANTOS THIBODEAU

O EFEITO DAS CONSEQUÊNCIAS E DO TEMPO NA ESCOLHA

THE EFFECT OF REINFORCEMENT AND TIME ON CHOICE

Universidade de Aveiro
2020

Departamento de Psicologia e Educação

CRISTINA JOY DOS SANTOS THIBODEAU

O EFEITO DAS CONSEQUÊNCIAS E DO TEMPO NA ESCOLHA

Tese apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Psicologia, realizada sob a orientação científica do Doutor Armando Machado, Professor Catedrático do Departamento de Psicologia e Educação da Universidade de Aveiro e do Doutor Marco Vasconcelos, Professor Auxiliar do Departamento de Psicologia e Educação da Universidade de Aveiro.

THE EFFECT OF REINFORCEMENT AND TIME ON CHOICE

Thesis presented to the University of Aveiro to fulfill the requirements for the degree of Doctor in Psychology, elaborated under the scientific supervision of Doctor Armando Machado, Full Professor of the Department of Psychology and Education of the University of Aveiro and of Doctor Marco Vasconcelos, Auxiliar Professor of the Department of Psychology and Education of the University of Aveiro.

o júri

presidente

Prof. Doutor Rui Luís Andrade Aguiar
professor catedrático da Universidade de Aveiro

Prof. Doutor Thomas Zentall
full professor, College of Arts & Sciences, University of Kentucky

Prof. Doutor Wolfram Erlhagen
professor associado da Universidade do Minho

Prof. Doutor Jeremie Jozefowiez
maître de conférences, Unité de Formation Et de Recherch de Psychologie, Université de Lille

Prof. Doutor Nuno Alexandre de Sá Teixeira
professor auxiliar convidado da Universidade de Aveiro

Prof. Doutor Armando Domingos Batista Machado
professor catedrático da Universidade de Aveiro

To my sisters.

agradecimentos

First, I would like to express my gratitude to my advisors, for their guidance and immense knowledge, for allowing me to grow as a research scientist, and for pointing me in the right direction when the way was not clear. To Marco, for his insightful comments, the humorous spirit, and always encouraging words. To Armando, for his patience, wise advice, and for the dancing lesson. Thank you both for navigating this endeavor with me.

My most sincere thanks go to Federico, for welcoming me in his lab as one of his team, as one of his students. I am grateful for his time, his unconditional disposition to teach and disinterested advice. Thank you for being an example, a mentor, and a good friend.

Thank you to my two sets of labmates, for the friendship and for being part of this journey. To Renata, Alex, Valeria, and Catarina, from the Animal Learning and Behavior Lab: for the support, the coffee breaks and our little parties. To Carter and Tanya, from the Basic Behavior Processes Lab: for the stimulating discussions, diner breakfasts, and the beers after work.

Again, thank you to Alex. For not only being my labmate, but my roommate, my dearest friend, and a brother. I sincerely can not imagine how this would have been —or if I could have even done it— without you.

To the institutions where, in one way or another, this work came together: Universidade do Minho, Arizona State University, and Universidade de Aveiro.

To the funding agencies. The Mexican Council of Science and Technology (CONACyT) for the Doctoral Fellowship 438355. The Portuguese Foundation for Science and Technology (FCT, PD/BD/128458/2017) and the Portuguese Ministry of Science, Technology, and Higher Education through national funds and cofinanced by FEDER through COMPETE 2020 under the PT2020 partnership agreement (POCI-01-0145-FEDER-007653).

palavras-chave

Comportamento animal, controle comportamental, percepção temporal, taxa de reforço, inversão ao meio da sessão, pombo, rato.

resumo

Em seu ambiente natural, a maioria dos animais consegue sobreviver porque aprende a responder de maneira adequada a dicas que sinalizam a disponibilidade de comida, a possibilidade de acasalamento ou a presença de predadores. Às vezes, mais de uma dica sinaliza a mesma consequência e, frequentemente, essas dicas podem mudar e se tornar menos confiáveis. No laboratório, tarefas de inversão de discriminação são bons testes da adaptabilidade comportamental a ambientes que mudam regularmente. Nesta série de estudos, exploramos os determinantes e as dinâmicas do comportamento quando tempo e a consequência da resposta indicam simultaneamente a disponibilidade de um reforçador em potencial. Assim, analisamos o desempenho de ratos e pombos em diferentes versões da tarefa de inversão ao meio da sessão (*midsession reversal task*). A tarefa tradicional consiste em uma discriminação simples simultânea em que respostas a um estímulo são reforçadas e respostas ao outro estímulo não são e, uma vez durante a sessão, as contingências são invertidas e o estímulo que sinalizava reforço agora sinaliza extinção e vice-versa. Utilizamos essa tarefa porque permite a manipulação independente do tempo e das consequências das respostas como dicas para o reforçamento futuro: Manipulamos a confiabilidade das consequências ao disponibilizar reforçamento contínuo ou parcial para cada alternativa e manipulamos a confiabilidade do tempo mantendo o momento de inversão fixo ou imprevisível. Os resultados sugerem que o controle comportamental varia entre as consequências e o tempo, de acordo com a relativa confiabilidade de cada dica. Simulações simples de modelos matemáticos mostram que as consequências e o tempo podem determinar o comportamento em conjunto, e que a taxa local de reforço pode determinar sua influência relativa. Oferecemos uma descrição geral de como animais se adaptam a ambientes que mudam regulamente.

keywords Animal behavior, behavioral control, timing, response outcomes, reinforcement rate, midsession reversal, pigeon, rat.

abstract In natural environments, most animals survive because they learn to respond appropriately to cues that signal the availability of food, a mate, or a predator. Sometimes there is more than one cue signaling the same outcome, and oftentimes these cues can change and become less reliable. In the laboratory, discrimination reversal tasks are good tests of behavioral adaptability to regularly changing environments. In this series of studies, we explore the determinants and the dynamics of behavior when time and the outcome of the previous response simultaneously signal the availability of a potential reinforcer. Hence, we analyzed the performance of rats and pigeons in different versions of the midsession reversal task. The traditional task consists of a simple simultaneous discrimination where responses to one stimulus are reinforced and responses to the other stimulus are not and, once throughout the session, contingencies reverse and the previously reinforced stimulus is now extinguished and vice versa. We used this task because it allows the independent manipulation of time and response outcomes as cues for future reinforcement: We manipulated the reliability of the outcomes by providing either continuous or partial reinforcement for each response alternative and manipulated the reliability of time by fixing the moment of reversal or making it unpredictable. Results suggest that behavioral control alternates between outcomes and time according to the relative reliability of each cue. Simple mathematical model simulations show that outcomes and time may jointly determine behavior, and that momentary reinforcement rate may determine their relative influence. We offer a general account of how animals may adapt to regularly changing environments.

Table of Contents

Abbreviations, Acronyms, and Symbols

ANOVA	Analysis of Variance
BEM	Behavioral Economic Model
BeT	Behavioral Theory of Timing
CI	Confidence Interval
CV	Coefficient of Variation
FOPP	Free Operant Psychophysical Procedure
ITI	Intertrial Interval
LeT	Learning-to-Time
MSR	Midsession Reversal
S1	Stimulus 1
S2	Stimulus 2
SET	Scalar Expectancy Theory
WSLS	Win-stay/lose-shift
WSLSS	Win-stay/lose-sometimes-shift

Figures

Figure 1.5. Average maximum-likelihood estimates of the location (μ) and scale (σ) parameters of Equation 1 and the coefficient of variation (CV= σ/μ) for the LeT model and the birds in each experimental condition. Error bars represent the 95% confidence interval.

Figure 1.6. Birds' performance and the LeT model predictions in all conditions. Symbols represent the average of the birds' performance and lines represent the average of the individual performances predicted by LeT. The top panel shows the performance in the late part of those conditions with different overall reinforcement rate. The middle panel shows the performance in the late part for those conditions with different relative reinforcement rate on each half of the session. The bottom panel shows the corresponding performance for the early part of the conditions displayed in the middle panel.

Figure 2.1. Average proportion of responses to S1 per trial relative to the proximity of the reversal in the last ten sessions of each condition. The solid lines represent the conditions with continuous reinforcement (q = 1), and the dashed lines the conditions with partial reinforcement (q = .5). Error bars represent 95% CI. Panels A and B show performance in the conditions with fixed and variable reversals, respectively. The grey vertical line indicates the moment between the last trial before the reversal and the first trial after the reversal

Figure 2.2. Proportion of errors in blocks of 5 trials relative to the reversal in each condition. The grey vertical line represents the location of the reversal. Error bars represent the 95% CI.

Figure 2.3. Proportion of errors in blocks of 5 trials relative to the reversal in conditions with variable reversal. Triangles represent performance when the reversal trial occurred early in the session (during the first in half); circles represent performance when the reversal occurred late in the session (during the second half).

Error bars represent the 95% CI. The grey vertical line indicates the location of the reversal.

Figure 2.4. Predicted performance by an alternating timing-WSLS model (top panels), by a WSLSS model (middle panels), and by a mixture timing-WSLSS model (bottom panels). The left panels show performance predicted in every condition of the experiment; the right panels show performance in conditions with variable reversal according to the moment it occurred. The grey vertical line indicates the location of the reversal.

Figure 3.1. Predicted versus observed performance in two different sessions of a MSR task with a variable reversal. The dashed lines represent performance in a session with the reversal on trial 21 and the solid lines represent performance in a session with the reversal on trial 61. The vertical grey lines represent the reversal points. Panel A shows the predicted performance by a win-stay/lose-shift strategy. Panel B presents the predictions of a pure timing model.

Figure 3.2. Proportion of errors in blocks of five trials relative to the reversal with the location of the reversal as a parameter. Each panel represents a different condition. Open circles represent the performance in the sessions where the reversal occurred early (between trials 16 and 32), closed circles represent the performance in the sessions where the reversal occurred late (between trials 50 and 66). Error bars represent the 95% CI.

Figure 3.3. Proportion of S1 responses as a function of blocks of trials relative to the reversal with the condition as a parameter. Each panel represents different sessions according to the location of the reversal. The left panel shows performance in the sessions with an early reversal and the right panel the sessions with a late reversal.

Figure 3.4. Proportion of S1 responses as a function of the reversal location with the condition as a parameter. Each panel represents a different block of trials relative to the

reversal. The left panel shows performance in the block of trials before the reversal and the right panel in the block of trials after the reversal.

Figure 3.5. Proportion of S1 responses by trial with the occurrence of the reversal as a parameter. The grey lines represent performance pre-reversal and the black lines post-reversal. Each panel corresponds to a different condition. The gray shaded area highlights the gap between the pre- and post-reversal curves in the common trials.

Figure 3.6. Predicted performance by a pure timing model (dashed lines) and a flexible win-stay/lose-shift rule (solid lines) in a MSR task with $q_1 = q_2 = 1$ and a variable reversal. The left panel shows the proportion of errors in blocks of five trials relative to the reversal in the sessions with an early (open symbols) and a late (closed symbols) reversal. The right panel shows the proportion of S1 responses in the pre-reversal (grey lines) and post-reversal (black lines) trials.

Figure 3.7. Mixture model I. Left panels show the predicted (lines) versus the observed (symbols) proportion of errors in blocks of five trials relative to the reversal with the location of the reversal as a parameter. Dashed lines and open circles represent the performance in the sessions where the reversal occurred early (between trials 16 and 32), solid lines and closed circles represent the performance in the sessions where the reversal occurred late (between trials 50 and 66). Right panels contrast the predicted (solid lines) versus the observed (dots) proportion of S1 responses in the pre-reversal (gray) and post-reversal (black) trials in each condition.

Figure 3.8. Mixture model II. The left panels show the predicted (lines) versus the observed (symbols) proportion of errors in blocks of five trials relative to the reversal with the location of the reversal as a parameter. The dashed lines and open circles represent the performance in the sessions where the reversal occurred early (between trials 16 and 32); the solid lines and closed circles represent the performance in the sessions where the reversal occurred late (between trials 50 and 66). The right panels contrast the predicted (solid lines) versus the observed (dots) proportion of S1 responses in the pre-reversal (gray) and post-reversal (black) trials in each condition.

Tables

Introduction

The backbone of any form of associative learning is the credit assignment: the process by which learning systems link the outcome of an event to its responsible factors. In the experimental analysis of behavior, the credit assignment is inferred from the study of stimulus control. Traditionally, the main variable of interest has been the control of behavior by reinforcement, while the study of the stimuli that are already present before the response occurs has been relegated to a second place (Dinsmoor, 1995). Yet, the present study is dedicated to understanding behavioral control with special emphasis on the second type of stimuli. We assess the role of two very peculiar discriminative stimuli for choice: the time elapsed since a particular event and the outcome of the previous response.

Despite most living organisms are geared to properly associate biologically relevant stimuli to its preceding events, the assignment of credit becomes challenging when multiple potentially relevant features are concurrently present. Commonly, when two discrete cues (e.g., light and tone, color and shape, etc.) are redundant at signaling reinforcement or its availability, the less salient of the two is overshadowed by the most salient one, acquires less associative strength and produces a weaker conditioned response (Pavlov, 1927). Curiously, timing cues do not seem to compete for associative strength with other discrete cues; temporal conditioning proceeds independently (Williams, Frame, & LoLordo, 1992). Moreover, animals can learn to do very precise estimations of time intervals if they are marked by specific events (Roberts, 2002).

For instance, consider Pavlov's (1927) report of an experiment conducted in his lab by Dr. Feokritova in 1911. In a response-independent (Pavlovian) preparation, a dog was repeatedly fed exactly every thirty minutes and each feeding was preceded a few seconds

by the sound of a metronome. In this situation, both the sound of the metronome and the time elapsed since the last feeding (a 30-min interval) could serve as indicators of the next reinforcer. One would think that the metronome alone would be good enough to predict food and elicit a conditioned response in the dog and that every time the metronome was played the dog would salivate. However, the experimenter claimed to observe absolutely no conditioned response if the metronome was presented at the twenty-ninth minute after the last feeding and, more importantly, that the metronome would only produce a full reaction (salivation) when presented at the thirtieth minute. Despite being hard to believe that there was no generalized conditioned responding to the sound of the metronome —let alone that the dog was able to time so accurately such a long interval— this is probably the first report of joint control of behavior by a discrete cue and time.

Many years later, Roberts (1981) documented evidence of joint behavioral control of time and a discrete cue in a response-dependent (operant) preparation. He trained rats and pigeons in two fixed-interval schedules of different length, and signaled each one of them with either with a light or a tone. Subsequently, when either the tone or the light where presented in a peak procedure (intermixed non-reinforced test trials of twice the duration of the longest fixed-interval); he hound that the rate of responding peaked at the time corresponding to the fixed-interval signaled by the test stimulus. Thus, concluding that joint control was established because reinforcement had operant control over the sequence of behaviors associated with each signal for which the target response had a particular placement.

Overall, it seems settled that when the time elapsed from a distinct event and a discrete cue signal the availability of reinforcement they share behavioral control. However, when the discrete cue, that simultaneously signals the availability of a potential

reinforcer, is not a traditional tone or light but the outcome of the previous response (i.e., reinforcement or non-reinforcement), the credit assignment is no longer so straightforward.

In a reversal learning study conducted by Cook and Rosen (2010), to examine how behavior is organized within a single session when switching between competing tasks, pigeons were required to track two conditional discriminations that used the same colors. On every trial, the subjects were presented with a sample stimulus and two comparison stimuli. During the first half of a session they were reinforced for choosing the comparison stimulus that matched the sample, and during the second half, they were reinforced for choosing the comparison stimulus that did not match the sample. In other words, the first half of each session required a matching-to-sample discrimination, and the second half, an oddity-from-sample discrimination. Despite the pigeons easily learned to perform each discrimination, after some training, performance showed a monotonic decrease in accuracy around the middle of the session. This is, a gradual transition from highly accurate matching-to-sample performance to a later equally strong display of oddity-to-sample behavior. The responding pattern was as if the pigeons were "anticipating" the change in the task requiring an oddity-to-sample discrimination and started switching the response rule head of time, and perseverated their matching behavior beyond the reversal. After putting to the test several explanatory hypothesis, the authors concluded that an interval timing process was the critical modulator of switching between tasks.

Rayburn-Reeves, Molet, and Zentall (2011), replicated these findings in pigeons with a simpler version of this Midsession Reversal (MSR) task requiring only a simultaneous color discrimination, and confirmed that the estimation of the midpoint of the session was the cue determining performance. Thus, the difficulty of the discrimination did not seem to be variable constraining performance to temporal control. Curiously, other

studies found that when the task requires a spatial discrimination, there is less influence of timing processes (Laude, Stagner, Rayburn-Reeves, & Zentall, 2014).

In the present series of experiments we explore the determinants and the dynamics of behavioral control when time and outcomes simultaneously hint on the availability of a potential reinforcement. We systematically manipulate the reliability of each of these cues, assess the effect on choice, and then compare the observed performance with the predictions of different simple mathematical models of behavior describing either temporal control, outcome control, or different combinations of the two. Quantitatively stating and contrasting these hypotheses allows us to discern the plausibility of different behavioral control mechanisms and their contribution in a variety of situations; a mission that otherwise would be speculative if not impossible.

Study 1 examined how biasing time perception affects choice in a MSR task. We trained pigeons in a color-discrimination MSR task and manipulated the reinforcement probability on each half of the session. We compared performance when the overall payoff remained constant but the payoff for each alternative differed, and when the payoff for both alternatives remained equal but the overall varied.

Study 2 assessed how past outcomes and time interact for behavioral control when each cue predicts the availability of reinforcement to a different extent. We trained rats in a MSR task in which we manipulated the reliability of the outcomes by providing either continuous or partial reinforcement and the reliability of time by fixing the moment of reversal or making it unpredictable.

Study 3 explored how time and past outcomes combine to determine performance when time is an uncertain cue but the outcome of the previous response signals the availability of reinforcement in different degrees. We trained pigeons in a MSR task with a

variable and unpredictable reversal and manipulated the payoff for each cue. This study replicated and combined the payoff conditions of Study 1 and the reversal variability of Study2.

All three studies contrast the results with the predictions of simple mathematical models of behavior —describing either temporal control, outcome control, or different combinations of the two—, and identify the features of performance that are consistent with each account and those that are not. Altogether, the present dissertation tests the limits of temporal control of behavior in a small section of all possible choice situations, but attempts to offer a general account of how animals adapt to regularly changing environments.

Study I

The effect of reinforcement probability on time discrimination in the midsession reversal task[1]

Abstract

We examined how biasing time perception affects choice in a midsession reversal task. Given a simultaneous discrimination between stimuli S1 and S2, with choices of S1 reinforced during the first, but not the second, half of the trials, and choices of S2 reinforced during the second, but not the first, half of the trials, pigeons show anticipation errors (premature choices of S2) and perseveration errors (belated choices of S1). This suggests that choice depends on timing processes, on predicting when the contingency reverses based on session duration. We exposed seven pigeons to a midsession reversal task and manipulated the reinforcement rate on each half of the session. Compared to equal reinforcement rates on both halves of the session, when the reinforcement rate on the first half was lower than on the second half, performance showed more anticipation and less perseveration errors, and when the reinforcement rate on the first half was higher than on the second half, performance showed a remarkable reduction of both types of errors. These results suggest that choice depends on both, time into the session and the outcome of previous trials. They also challenge current models of timing to integrate local effects.

Keywords: midsession reversal, reinforcement rate, psychometric function, response bias, timing, key peck, pigeon.

[1] This chapter reproduces the publication: Santos, C., Soares, C., Vasconcelos, M., & Machado, A. (2019). *Journal of the Experimental Analysis of Behavior, 111,* 371-386.

To study how animals behave in frequently changing environments, researchers have used a variety of experimental procedures, including the Free Operant Psychophysical Procedure (FOPP; e.g., Bizo & White, 1994, 1995; Cowie, Bizo, & White, 2016; Machado & Guilhardi, 2000; Stubbs, 1980) and the Midsession Reversal task (MSR; e.g., Cook & Rosen, 2010; Rayburn-Reeves, Molet, & Zentall, 2011). These procedures have a common structure, illustrated in the top panel of Figure 1.1. Subjects choose between two simultaneously available stimuli, one that is reinforced (S+) and another that is not (S-) and, at some point in time, the contingencies associated with the two stimuli reverse so that the initial S+ becomes the S- and the initial S- becomes the S+. The vertical line in the figure marks the moment of the reversal. We refer to the first S+ as S1 and to the second S+ as S2. The two procedures differ mainly in the time scale of the periods before and after the reversal: In the FOPP, the contingencies reverse half way into each trial; in the MSR, they reverse half way through each session. The main question of interest is how animals behave in these changing environments.

To answer the question, researchers examine how the proportion of choices to one alternative varies with time into the trial, or with trials into the session. The bottom panel of Figure 1.1 shows a hypothetical but representative example of this psychometric function. The choice favors S1 during the first period (i.e., when only S1 is reinforced), S2 during the second period (i.e., when only S2 is reinforced), and is indifferent around the moment of reversal. The function also shows the two types of errors that cluster near the reversal (shaded area): anticipation errors, when the animal chooses prematurely S2, and perseveration errors, when the animal chooses belatedly S1.

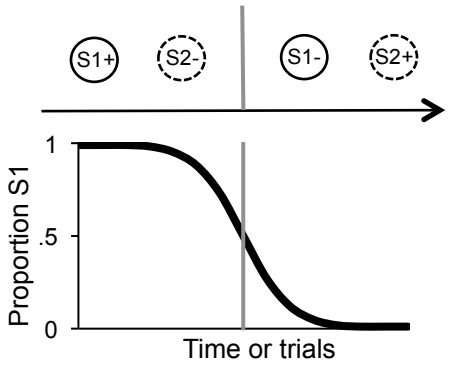

Figure 1.1. Diagram of a contingency reversal task and the typical performance pattern. The top panel shows that the S+ and S- keys are reversed at some point in time, and the bottom panel shows, at steady state, the performance associated to each key as times elapses.

These two error types suggest that the animals are *timing* the moment of the reversal,

using as time marker the beginning of the S1 period. To illustrate, Stubbs (1980) used a FOPP to compare pigeons' discrimination performance across different trial lengths. During the first half of each trial, responses to an orange but not to a green key were reinforced; during the second half, responses to the green but not the orange key were reinforced. He found that the psychometric functions, with both types of errors, remained scale invariant across a wide range of trial durations, a clear expression of Weber's law for time. Similarly, Cook and Rosen (2010) investigated how pigeons behave when the reinforcement rule changes using a MSR task. They found that the frequency of anticipation and perseveration errors increased as a function of proximity to the reversal point. The authors concluded the birds were not using the local cue of presence vs. absence of food to switch rules. Instead, they seemed to rely on the time since the beginning of the session as a cue to switch between rules. Subsequent studies found the same pattern of errors, further supporting the idea that timing processes are involved (e.g., McMillan, Sturdy, Pisklak, & Spetch, 2016; McMillan, Sturdy, & Spetch, 2015; Laude, Stagner, Rayburn-Reeves, & Zentall, 2014; Rayburn-Reeves & Zentall, 2013; Rayburn-Reeves, Laude, & Zentall, 2013; Rayburn-Reeves, et al., 2011). Discriminations between spatial cues seem to hinder the influence of such processes, though (e.g., Rayburn-Reeves, Moore, Smith, Crafton, & Marden, 2018; Laude, et al., 2014; McMillan, Kirk, & Roberts, 2014).

In both the FOPP and the MSR, the contingency change is a critical environmental feature and the evidence suggests that behavior is mostly under control of the time elapsed since a particular time marker (the start of the trial or session) as a cue to reverse choice (Rayburn-Reeves & Cook, 2016). In the FOPP, this time-based strategy is not surprising; given that most choices are unreinforced, the best predictor of the current contingency is time since trial onset. In this task, the errors follow naturally from the animals' limited timing abilities. However, in the MSR task, the use of a time-based strategy is surprising because local cues are readily available. Why animals anticipate and switch to S2 when all its previous choices of S1 have been reinforced remains an open question.

Surprising as it may be, if the MSR task engages the animal's timing processes, then we may use current findings and theories of timing to predict the effect of some variables in this task. Consider first the evidence that differences in the reinforcement rate affect timing. In a FOPP, Bizo and White (1995) manipulated the rate of reinforcement associated with the S1 and S2 choices. In baseline, S1 and S2 were reinforced equally with

variable interval (VI) 60 s schedules. In subsequent conditions, the VI schedules differed, favoring either one or the other choice (e.g., VI 120-s for S1 and VI 40-s for S2, making S2 three times richer than S1). They found that the psychometric function showing the proportion of S1 choices across time into the 50-s trial shifted to the left when S2 was richer (i.e., the pigeons switched to S2 earlier than in baseline), and to the right when the S2 was leaner (the pigeons switched to S2 later than in baseline). They also found that the magnitude of the shift varied with the magnitude of the payoff difference. The psychometric functions conformed always to Weber's law.

Later, Machado and Guilhardi (2000) analyzed performance in the FOPP in light of the Learning-to-Time model (LeT; Machado, 1997; also Machado, Malheiro, & Erlhagen, 2009) and claimed that preference shifted not because the overall reinforcement rates differed between the two trial halves, but because the rates differed around the middle of the trial, the reversal point. The shifts, in other words, did not express a simple biasing effect, but a time-dependent biasing effect. To test the hypothesis, they divided each 60-s FOPP trial into four 15-s periods, each with an independent VI schedule allocating the reinforcers. As usual, during the first half of the trial, only S1 pecks were reinforced, and during the second half only S2 pecks were reinforced. They found that, as LeT predicted, the psychometric functions shifted when the reinforcement rates differed around the reversal point, but not when they differed far from the reversal point. These findings, later replicated with rats by Guilhardi, McInnis, Church, and Machado (2007), are consistent with the model's key idea that both the times of reinforcement and reinforcement rate at those times influence performance.

More recently, Cowie et al. (2016) also examined how relative and absolute reinforcement rates affect response bias and temporal discrimination in the FOPP. Across conditions, the authors varied the absolute reinforcement rate (lean or rich schedules), the relative reinforcement rate on each half of the trial (1:1, 5:1, 1:5), and the trial duration (short or long). They found that the psychometric curves did not change with the overall reinforcement rate. However, they changed with the relative reinforcement rate, shifting towards the richer alternative. Additionally, the slope of the function increased with long trials and with more reinforcers for S2 than S1, a finding that reveals increased sensitivity to time under those conditions, and seems to violate the scalar property of timing (e.g., Gibbon, 1977; Lejeune, & Wearden, 2006). The authors argue that the change in

sensitivity was an indirect effect of the asymmetries in the temporal distribution of the reinforcers. To illustrate, because VI schedules set up the reinforcers, when a pigeon switched prematurely from S1 to S2, any reinforcer for S1 set up subsequently (i.e., between the time of switching and the middle of the trial) was collected at the beginning of the next trial when the pigeon chose S1 again. Cumulated over trials, the net effect on the distribution of reinforcers for S1 could be a distinct mode at trial onset. In any case, Cowie et al.'s results need to be reproduced because previous studies with the FOPP did not report similar violations of the scalar property (e.g., Guilhardi, et al., 2007; Machado and Guilhardi, 2000; Bizo and White 1995; 1994; Stubbs, 1980). Nevertheless, these results together with those presented by Machado and Guilhardi (2000) support the idea that both the times of reinforcement and reinforcement rate determine timing performance on the FOPP.

Consider now the MSR task. We may conceive of it as a scaled up version of the FOPP, a session-wide rather than a trial-wide FOPP (McMillan, Spetch, Roberts, & Sturdy, 2017). If timing plays a central role in both tasks, as the evidence suggests, then we should be able to bias performance in the MSR task by manipulating the same variables that bias performance in the FOPP, and we should be able to account for performance in the MSR task using the same models that accounted for performance in the FOPP. These then were the empirical and theoretical goals of the present study. Empirically, we varied the reinforcement probabilities given S1 and S2 choices and measured the magnitude of the shifts of the psychometric function; theoretically, we extended LeT to the MSR task for the first time and examined how well it accounts for the data.

In the experiment reported below, we used a color-discrimination MSR task for pigeons, as described by Rayburn-Reeves, et al. (2011), and manipulated the probability of reinforcement for a correct response on each half of the session. We compared performance when (1) the scheduled overall reinforcement probability remained constant but the payoff probabilities for S1 and S2 differed, and when (2) the payoff probabilities for S1 and S2 remained equal but the overall payoff probability varied. We predict that when S1 is richer than S2, S1 should remain the preferred option beyond the reversal trial and the psychometric function (showing preference for S1 across trials) should shift to the right. When S1 is leaner than S2, S1 should cease to be the preferred option before the

reversal trial and the psychometric function should shift to the left. Finally, when the probabilities of reinforcement for S1 and S2 are equal, the psychometric function should cross the indifference line at or close to the reversal trial; it should reveal no bias.

By studying how reinforcement and timing interact in the MSR task, we hope to clarify how local (response outcome) and global (time) cues combine to guide choice as the animal adapts to a changing environment. In some respects, the MSR task may be more appropriate than the FOPP to study how reinforcers bias timing. First, the interaction between choice and obtained reinforcers that Cowie et al. (2016) observed in the FOPP cannot occur in the MSR task because, in the latter, scheduled reinforcers follow correct responses immediately. Hence, the distribution of the reinforcers across trials, and a fortiori across time, mirrors the distribution of correct choices across trials/time, and no spurious, procedure-induced asymmetries in the distribution of reinforcers can take place. On the other hand, if the changes in time sensitivity (Cowie et al., 2016) occur on long trials independently of the temporal distribution of reinforcers, then we should observe them in the MSR task as well.

Second, by varying the reinforcement probabilities in the MSR task, we may better understand how a global variable, time, combines with local variables, the response outcomes, in the control of behavior. For though the evidence suggests that animals are timing the moment of reversal, that evidence does not rule out the possibility that choice on the next trial also depends on the food or no-food consequences of previous choices. These local sources of control may become more salient or change their relative influence on choice when the reinforcement probabilities for S1 and S2, are not both equal to 1, as in the standard case, or when they differ. Hence, varying the payoff for S1 and S2 in the MSR task may help understand how local and global cues guide choice in changing environments.

Method

Subjects

Seven pigeons (*Columba livia*) with previous experience (in concurrent chain schedules and bisection tasks) served as subjects. They were housed in individual cages, with water and grit constantly available, in a colony room with controlled temperature

(range: 18 and 26^{o} C), and a 13:11 h light-dark cycle, with lights on at 8:00 a.m. The pigeons received a supplementary feed of mixed grain after the experimental sessions to maintain their body weights between 80% and 85% of their average free-feeding weight. The experimental sessions started approximately at 9:00 a.m. each day, six days a week.

Apparatus

The study used three standard Med Associates chambers for pigeons, each 31.8 x 25.4 x 34.3 cm (length, depth, height) with acrylic walls for the ceiling, left, and right panels, aluminum walls for the front and back panels, and a metal grid for the floor. The front panel had three circular keys, each 2.5 cm in diameter, 20.5 cm above the floor and horizontally centered on the panel (8 cm apart, center-to-center). Each response key had a 12-stimulus IEE (Industrial Electronics Engineers) in-line projector with 28-V, 0.1-A light bulbs. On the same panel, a feeder opening (6 x 6.5 cm) horizontally centered on the wall and 4 cm above the floor, gave access to mixed grain when the feeder was activated and illuminated by a 28-V, 0.04-A light bulb. A 28-V, 0.1-A houselight located in the back panel provided general illumination. A personal computer, using the ABET II® software (Lafayette Instruments), controlled and recorded all experimental events.

Procedure

Each session comprised 80 trials. At the beginning of each trial, the two side keys of the chamber illuminated simultaneously, one with a green hue and the other with a red hue. The position of each color varied randomly with the constraint that each color appeared an equal number of times on each side key per session. A single peck to one of the keys turned both keylights off and led directly to either a 5-s inter-trial interval (ITI) when the response was incorrect, or to 2 s of access to food, with the feeder raised and illuminated, followed by a 3-s ITI when the response was correct. The houselight remained off during the entire session.

During trials 1 to 40, pecks to one key color, S1, were correct while pecks to the other key color, S2, were incorrect; during trials 41 to 80, the previously correct color was now incorrect and vice-versa. The colors associated with S1 and S2 were counterbalanced across birds, but the assignment for each bird remained constant throughout the experiment.

On the first day of training, every correct response was reinforced. For the following two days, only 75% of the trials were reinforced (given a correct response). Then the experiment proper began and lasted for approximately 170 sessions divided into four conditions according to an ABCBCD design. Table 1.1 shows, for each bird, the order of the conditions and the number of sessions per condition.

Table 1.1
Order of conditions and number of sessions completed by each bird.

		Bird					Bird		
Phase	Condition	P458	P917	P935	P730	Condition	P851	PG17	PG23
1	Int-Int	39	40	40	40	Int-Int	40	39	40
2	High-Low	40	40	40	40	Low-High	40	40	40
3	Low-High	43	41	43	43	High-Low	41	40	44
4	High-Low	20	20	20	20	Low-High	21	20	20
5	Low-High	20	20	20	20	High-Low	20	20	19
6	High-High	10	10	10	10	High-High	10	10	10

Note. Condition refers to the proportion of reinforced trials in each half of the session [1st half – 2nd half]: Int = 0.5; High = 1.0; Low = 0.2. In Phases 1 to 3, training continued for approximately 40 sessions, and in all conditions, performance was stable after about 20 sessions. For this reason, the following two phases (4 and 5) consisted of approximately 20 sessions. Phase 6 continued for only 10 sessions because the birds' performance was considered stable by visual inspection.

The payoff probabilities for S1 and S2 defined each condition. In Condition High-High, every correct response was reinforced, allowing a high overall reinforcement rate (1) and equal relative reinforcement rates for each half of the session. In Condition Int-Int, the proportion of reinforced trials was intermediate and the same for both halves of the session. During this condition, at the beginning of a session, 20 trials of each half were "baited", allowing for an intermediate (.5) overall reinforcement rate and no difference between each half of the session. If a correct choice occurred on a baited trial, reinforcement followed; on the non-baited trials, no reinforcement followed even when the choice was correct. In conditions High-Low and Low-High, sessions had a different proportion of reinforced trials on each half and an intermediate (.6) overall reinforcement rate. For the High half, the proportion of reinforced trials was set to 1 (all correct responses were reinforced), while for the Low half, the proportion of reinforced trials was set to .2 by randomly identifying eight trials as "baited", in case of a correct response, and 32 as non-

baited. Again, the maximum number of reinforcers obtainable in that half was eight (20% of the trials) and the distribution of reinforcers varied randomly from session to session.

To avoid any contrast effect when exposed to the first condition with different relative reinforcement rates, in Phase 1 all subjects experienced condition Int-Int. Thus, all birds experienced some change in the relative reinforcement rates when transitioning to Phase 2, and a minimal change in the overall payoff. Only conditions High-Low and Low-High were counterbalanced across subjects in Phases 2 - 5. Lastly, in Phase 6, all subjects were exposed to the High-High condition.

Results

To characterize performance in each condition, we computed from the last 10 sessions of that condition the proportions of S1 choices on trials 1 to 80, the psychometric function. Next, we fit to each individual psychometric function an inverted cumulative Gaussian curve – see Equation 1. The curve relates the probability of choosing S1, P_{S1}, to trial number, x, via four free parameters, γ, λ, μ, and σ, and the standard normal integral, Φ. Parameters γ and λ determine the range of P_{S1}, its minimum ($\min(P_{S1}) = \lambda$) and maximum ($\max(P_{S1}) = 1 - \gamma$); other things equal, the closer λ and γ are to 0, the better the discrimination. The location parameter μ corresponds to the trial on which the probability of choosing S1 is half way between its maximum and minimum (i.e., $P_{S1}(\mu) = (\max + \min)/2$). The scale parameter σ determines (inversely) the slope of the curve at μ. We estimated the best-fitting parameters by the method of maximum likelihood using the quickpsy package (Linares and López-Moliner, 2016) of the R software.

$$P_{S1}(x) = 1 - \left[\gamma + (1 - \gamma - \lambda) * \Phi\left(\frac{x-\mu}{\sigma}\right)\right] \quad (1)$$

Equation 1 fit the data well (all chi square test statistics had $p > .05$). The good fit legitimizes the use of the estimated parameter values to summarize individual performance and compare it across conditions. In addition, because the range parameters γ and λ were always close to 0, we use μ to measure bias (for S1 if $\mu > 40$, and for S2 if $\mu < 40$) and σ to measure sensitivity.

Henceforth, we compare the data and the fits from the High-High, Int-Int, High-Low, and Low-High conditions. Table 1.2 shows the parameter estimates for each pigeon across conditions and the coefficients of variation (CV = σ/μ, a measure of relative timing accuracy). Within conditions, the parameter estimates for different birds were similar. Only the σ of pigeon PG23 in condition Int-Int was more than two standard deviations above the group mean.

Table 1.2
Maximum-likelihood estimates of parameters of Equation 1 and the coefficient of variation (CV) for each pigeon in each experimental condition.

	Conditions									
	Int - Int					**High - High**				
Birds	***μ***	***σ***	***γ***	***λ***	***CV***	***μ***	***σ***	***γ***	***λ***	***CV***
P458	43.77	8.38	0.000	0.005	0.19	38.57	9.97	0.000	0.000	0.26
P917	42.78	13.64	0.000	0.000	0.32	41.58	7.82	0.003	0.037	0.19
P935	40.85	9.17	0.000	0.010	0.22	42.15	6.95	0.000	0.000	0.16
P730	39.23	10.20	0.000	0.008	0.26	40.53	9.41	0.015	0.020	0.23
P851	47.67	9.62	0.021	0.011	0.20	43.03	8.72	0.032	0.029	0.20
PG17	38.91	9.20	0.035	0.073	0.24	38.80	4.13	0.003	0.017	0.11
PG23	32.19	22.16	0.000	0.000	0.69	39.25	9.51	0.008	0.010	0.24
Average	40.77	11.76	0.01	0.02	0.30	40.56	8.07	0.01	0.02	0.20
95% CI	37.19	8.14	0.00	0.00	0.17	39.26	6.57	0.00	0.01	0.16
	44.35	15.39	0.02	0.03	0.43	41.86	9.58	0.02	0.03	0.24
	High - Low					**Low - High**				
Birds	***μ***	***σ***	***γ***	***λ***	***CV***	***μ***	***σ***	***γ***	***λ***	***CV***
P458	42.48	2.68	0.000	0.012	0.06	33.85	11.62	0.040	0.051	0.34
P917	43.95	2.64	0.001	0.030	0.06	31.98	7.13	0.000	0.011	0.22
P935	42.39	1.25	0.020	0.014	0.03	30.87	10.02	0.012	0.004	0.32
P730	43.43	2.80	0.006	0.024	0.06	30.85	11.75	0.000	0.000	0.38
P851	46.89	4.20	0.024	0.085	0.09	37.31	8.53	0.040	0.002	0.23
PG17	43.82	1.79	0.021	0.019	0.04	32.59	9.47	0.024	0.004	0.29
PG23	46.03	7.03	0.005	0.049	0.15	24.41	15.51	0.000	0.006	0.64
Average	44.14	3.20	0.01	0.03	0.07	31.69	10.58	0.02	0.01	0.35
95% CI	42.87	1.77	0.00	0.01	0.04	28.80	8.56	0.00	0.00	0.24
	45.41	4.62	0.02	0.05	0.10	34.59	12.59	0.03	0.02	0.45

Figure 1.2 shows individual performance in the High-High and Int-Int conditions, the conditions with different overall reinforcement rate (1 vs. .5) but equal relative payoff on each half of the session. Each dot represents the proportion of S1 choices on the corresponding trial, a proportion computed from the last 10 sessions, and the smooth

curves represent the best fitting functions with the parameters shown in Table 1.2. The bottom right panel shows the average of the data and of the individual functions.

Performance was similar in the two conditions. With the exception of PG23, the two psychometric functions were close to each other, revealing no systematic differences in their relative position (paired t-test on μs, $t(6) = 0.135$, $p = .897$). In addition, in both conditions, the degree of bias was small and inconsistent across pigeons – the 95%

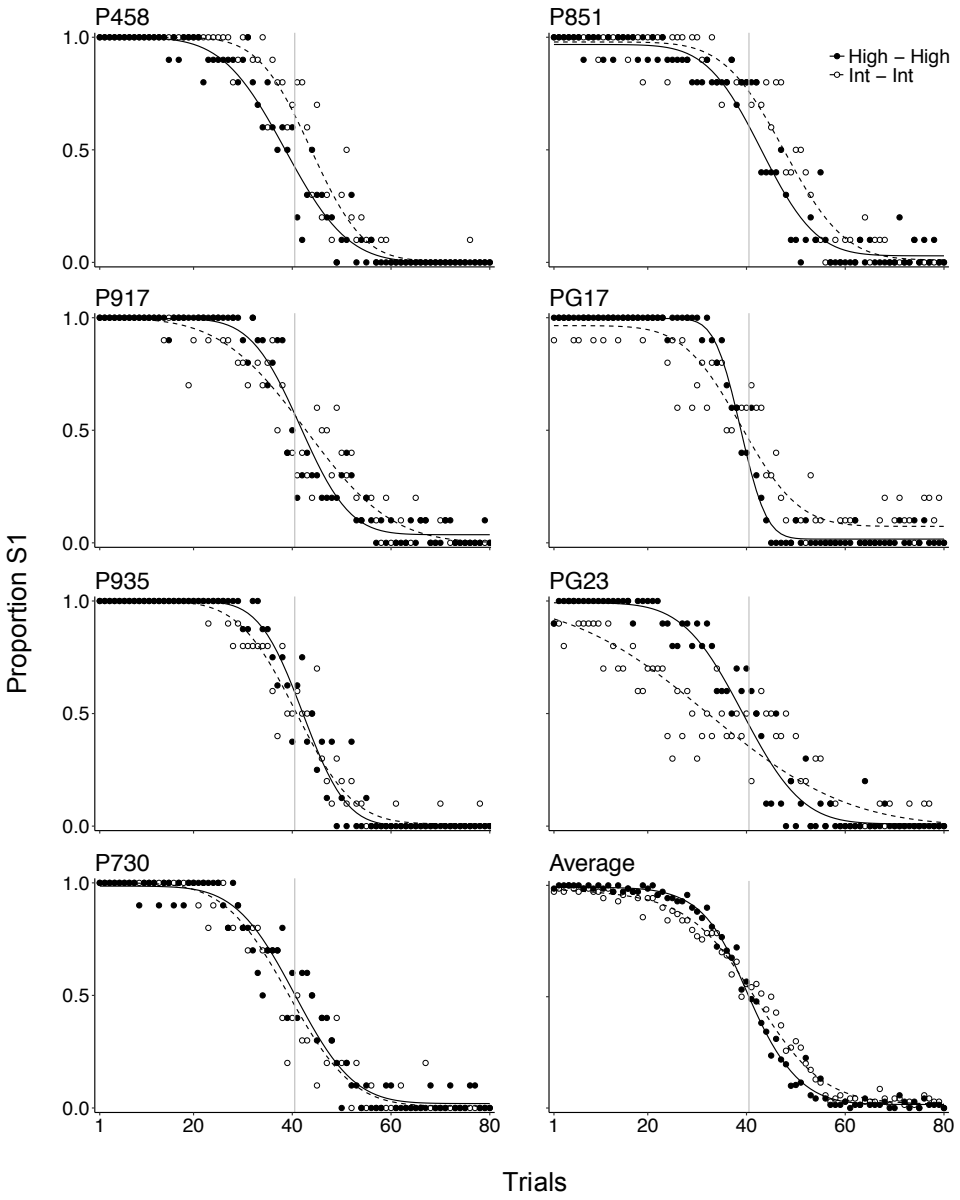

Figure 1.2. Proportion of responses to S1 in the last 10 sessions of conditions Int-Int and High-High for each bird. The symbols show the proportion of S1 choices on each trial in conditions High-High and Int-Int. The lines plot Equation 1 with the parameters from Appendix A. The bottom right panel shows the average of the data and the average of the fitted functions. The gray vertical line represents the reversal point after trial 40.

confidence intervals [95% CI] for the mean of the location parameter, were [39.26 – 41.86] for the High-High condition and [37.19 – 44.35] for the Int-Int condition. The results were similar for the scale parameter, even though the average curve suggests a slightly higher sensitivity during the High-High condition. A paired-samples t-test on the estimated σs revealed no difference between conditions ($t(6) = 2.075$, $p = .083$; the conclusion still holds if we exclude from the test the data from PG23). Finally, the coefficients of variation also suggested similar Weber's fractions between conditions ($t(6) = 1.647$, $p = .151$).

As described before, subjects experienced conditions High-Low and Low-High twice (see Table 1.1), nonetheless their performance appeared similar by visual inspection in each repetition. Paired samples t-tests on the estimated values of the location and scale parameters confirmed that performance in the two High-Low conditions did not differ significantly: For σ, the means (standard deviations) for the two replications were 3.38(2.49) and 7.72(12.37), $t(6) = -0.88$, $p = .411$; for μ, the corresponding values were 44.2(3.00) and 44.0(1.34), $t(6) = 0.21$, $p = .844$). In the two Low-High conditions, the σ did not differ significantly (9.68(2.69) vs. 10.94(2.35), $t(6) = -1.48$, $p = .189$), but the μ did (33.8(4.4) vs. 29.1(2.5), $t(6) = 5.01$, $p = .002$). The difference of about 4.7 trials in the location parameter revealed a slightly stronger bias for S2 during the second exposure to the Low-High condition. However, because the difference was relatively small and there was no difference in the scale parameters, we averaged the psychometric functions of each repetition.

Figure 1.3 shows the data from the High-Low and Low-High conditions, the conditions with the same overall reinforcement rate and different relative payoff per half of the session. Each dot represents the proportion of S1 choices on the corresponding trial, a proportion computed from 20 sessions, the last 10 sessions of the two replications of the condition. The bottom right panel shows the average of the data and of the individual functions.

Performance differed markedly between the conditions. The location parameters were considerable smaller in condition Low-High than in condition High-Low —in Figure 1.3, the curves for the Low-High condition are systematically to the left of the curves for the High-Low condition, indicating that the pigeons started to choose S2 earlier in the

session, anticipating the reversal. The means of the location parameters differed by about 12 trials (M = 32 vs. M = 44; paired-samples t-test, $t(6) = 7.706$, $p < .001$). Moreover, the two conditions yielded biases of different magnitudes and in opposite directions: when S1 was the richer key (High-Low), bias was in the direction of S1 ($\mu > 40$) and of small magnitude (95% CI [42.87 - 45.41]); when S1 was the leaner key (Low-High), bias was in the direction of S2 ($\mu < 40$) and of large magnitude (95% CI [28.80 – 34.59]).

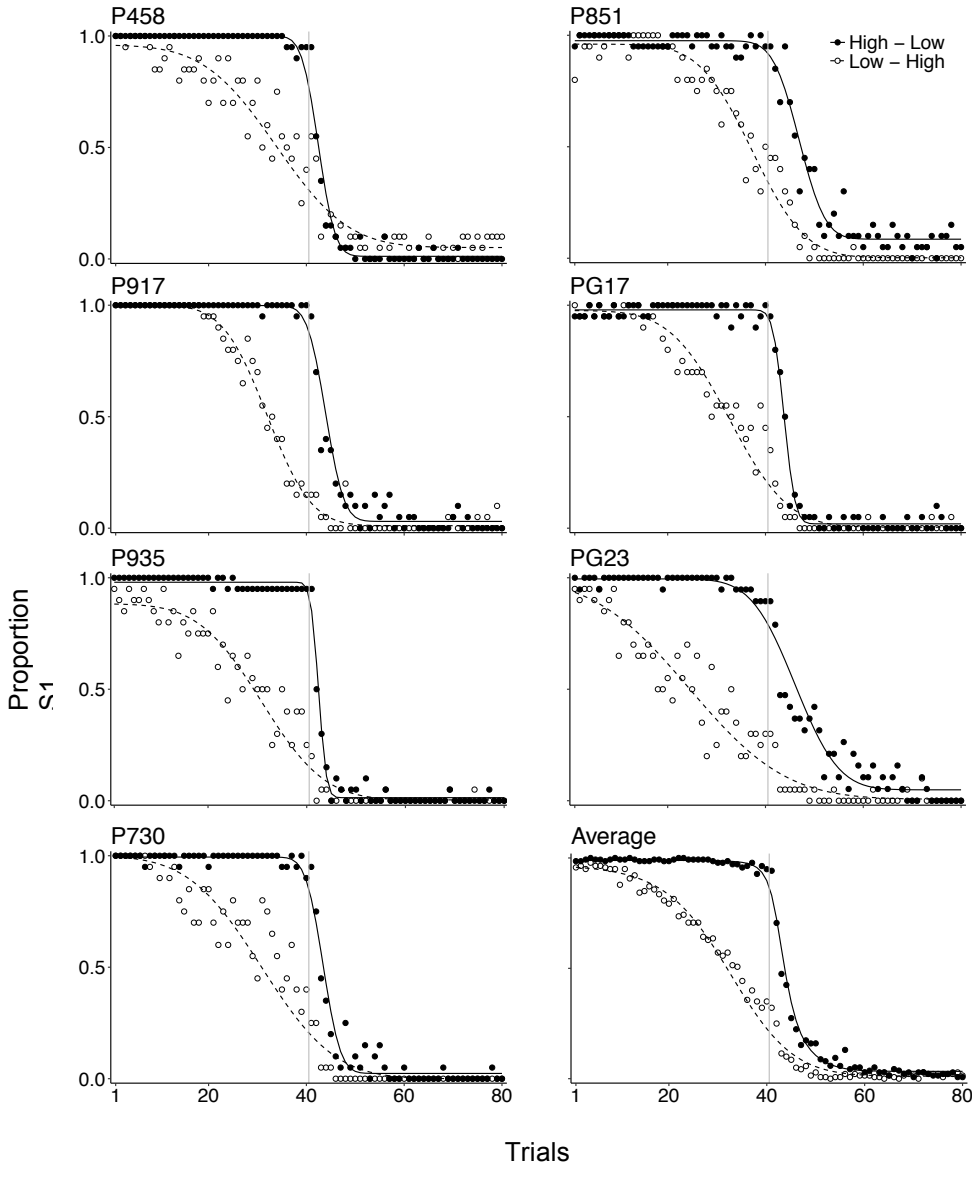

Figure 1.3. Proportion of responses to S1 in the last 10 sessions of each exposure to the High-Low and Low-High conditions for each bird. The symbols show the proportion of S1 responses on each trial in conditions High-Low and Low-High. Each proportion was computed from 20 sessions, the last 10 from each replication of each condition. The lines plot Equation 1 with the parameters from Appendix A. The bottom right panel shows the average of the data and the average of the fitted functions. The gray vertical line represents the reversal point after trial 40.

The scale parameter also differed between conditions. Visually, the filled dots in Figure 1.3 reveal abrupt change in preference, whereas the empty dots reveal gradual changes in preference. Hence, the values of σ were substantially smaller in the High-Low condition (M = 3.20) than in the Low-High condition (M = 10.58; paired-samples t-test, $t(6) = -9.414$, $p < .001$). Because μ increased and σ decreased during the High-Low condition, the coefficients of variation in that condition were much smaller than in any other condition. In fact, the 95% CI for the coefficient of variation in condition High-Low, [0.04 - 0.10], did not overlap the interval for any other condition (Low-High: [0.24 - 0.45], Int-Int: [0.17 - 0.43], and High-High: [0.16 - 0.24]). These results show that Weber's fractions were remarkably small in condition High-Low and hardly consistent with a timing process. We return to this result below.

We compared performance across all four conditions by means of one-way ANOVAs on the estimated parameters. With respect to the location parameter, the ANOVA revealed a significant effect, $F(3, 24) = 17.79$, $MSE = 11.15$, $p < .001$, $\eta^2 = .69$, and Bonferroni's *post hoc* tests showed that only condition Low-High was significantly different from all others (all *p*s < .001). With respect to the scale parameter, the ANOVA revealed also a significant effect, $F(3, 24) = 10.31$, $MSE = 9.79$, $p < .001$, $\eta^2 = .56$, and Bonferroni's *post hoc* tests indicated that only the High-Low condition differed from the others (all *p*s < .05).

In summary, when S1 was richer than S2 (condition High-Low), the pigeons switched their preference from S1 to S2 shortly after the reversal trial and abruptly. Compared to conditions Int-Int and High-High, the indifference point did not change but sensitivity increased substantially. In contrast, when S2 was richer than S1 (condition Low-High), the pigeons switched their preference from S1 to S2 well before the reversal trial and the switch was gradual. Compared to conditions Int-Int and High-High, indifference decreased substantially but sensitivity did not change.

Lastly, to see how preference evolved when the reinforcement probabilities for S1 and S2 changed, we examined the transition between the High-Low and Low-High conditions. Consider Pigeon P917. It experienced the conditions in the following order: High-Low, Low-High, High-Low, and Low-High, each with at least 20 sessions. To examine the transition, we computed first the psychometric functions of the first 10 sessions (early) and the last 10 sessions (late) of each condition, obtaining 8 functions in

total. Then, we averaged the corresponding functions from the two replications (e.g., the Early functions from the two High-Low conditions), obtaining 4 functions in total. Finally, we combined the 80 trials into 16 blocks of five trials each.

Figure 1.4 shows the results. The filled and empty symbols identify the High-Low and Low-High conditions, respectively; the dashed and solid curves identify the early and

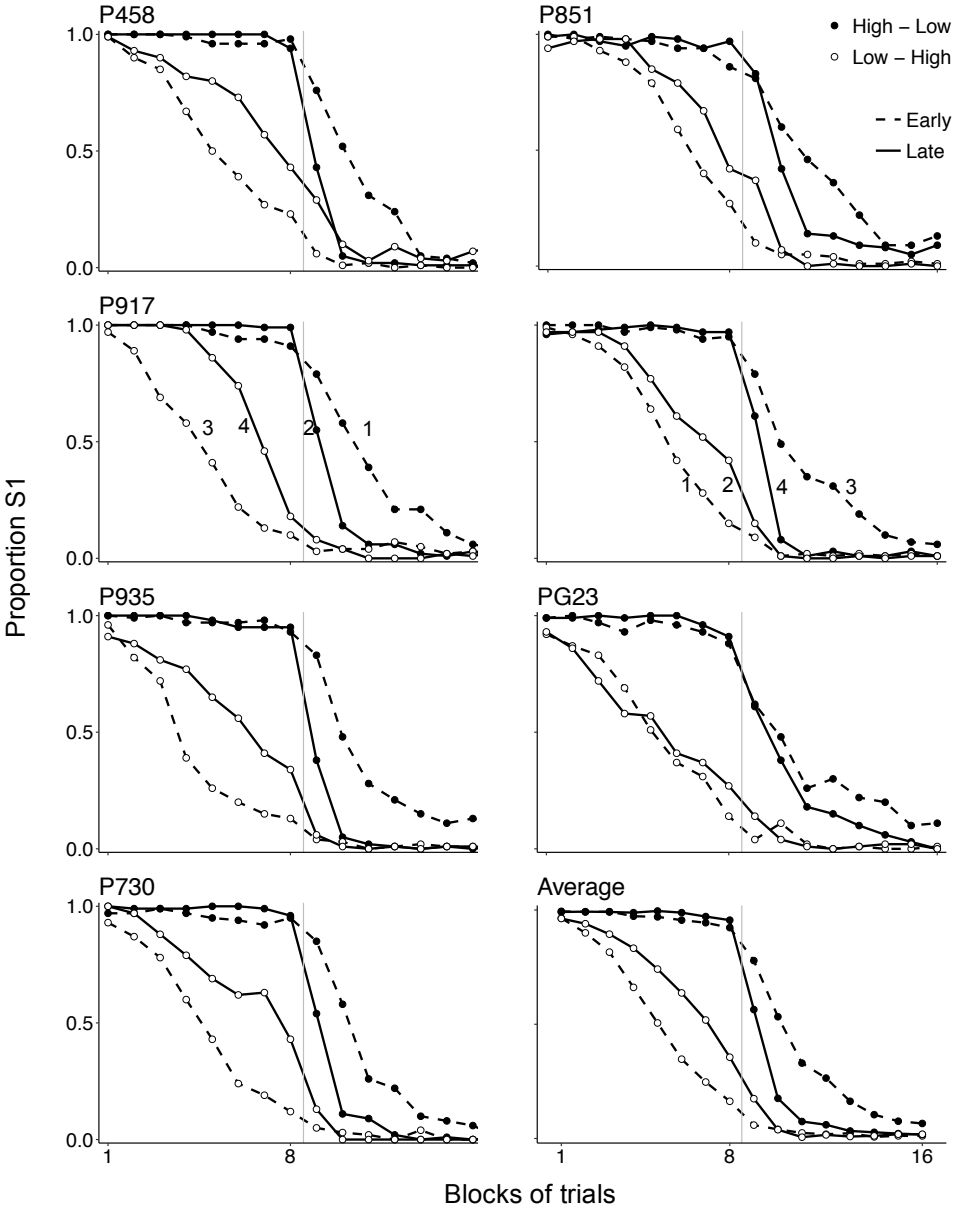

Figure 1.4. Proportion of responses to S1 per blocks of five trials in the Early and Late parts of the High-Low and Low-High conditions for each bird. Proportions were computed from 20 sessions (10 sessions from each replication of each condition), for the Early part, the first 10; and for the Late part, the last 10. The bottom right panel shows the average of the data. The gray vertical line represents the reversal point after trial 40.

late sessions, respectively. The numbers 1 to 4 show the order of the conditions (left panels: same order as P197; right panels: same order as PG17). In the Low-High condition (empty circles), the difference between the early and late sessions occurs mostly during the first half of the session, with the proportion of anticipatory errors decreasing from the early to the late sessions. In contrast, in the High-Low condition (filled circles), the difference between the early and late sessions occurs mostly in the second half of the session, with the proportion of perseverative errors decreasing from the early to the late sessions. Moreover, for all pigeons except PG23, the curves from the early sessions (1 and 3) are further away from the midsession reversal trial than the curves from the late sessions (2 and 4). These dynamic changes occurred in every exposure to the experimental conditions and regardless of the order in which they were experienced. They show that when transitioning from one condition to another, bias changes maximally during the first sessions, but then decreases with additional sessions.

Discussion

The influence of reinforcement on timing behavior has been widely studied, but still little is known about the mechanisms underlying these effects. Our study explored the influence of relative reinforcement rate on timing behavior in the MSR task, a task in which the outcome of each trial could act as a local cue and gain, modulate, or compete for control over behavior.

Based on the results of similar studies with the FOPP, we did not predict any effect on the psychometric functions of the overall reinforcement probability when the proportion of reinforcement for S1 and S2 was the same. Accordingly, in conditions High-High and Int-Int, the average of the psychometric functions overlapped considerably, and the individual psychometric functions did not show consistent changes in location or sensitivity. These results replicate the previous findings of Cowie et al. (2016) and are consistent with several theories of timing such as the Scalar Expectancy Theory (SET; Gibbon, 1977), the Behavioral Economic Model (BEM; Jozefowiez, Staddon, & Cerutti, 2009), and the LeT model (Machado, 1997; Machado et al., 2009).

Nevertheless, one could argue that the High-High and Int-Int conditions are not directly comparable because the pigeons experienced them after different amounts of

training. Given that all pigeons experienced the Int-Int condition first, and the High-High condition last, it is reasonable to ask if performance in the latter condition would have been the same had the birds experienced it in first place and/or for more than ten sessions. To answer this question we examined a set of unpublished data from our lab, data produced by five naïve pigeons exposed for 40 sessions to a MSR task with the exact same parameters as in the present experiment. We fitted the data from these "new" pigeons (NP) in the way described above and compared their performance with that of the High-High (HH) condition of the present experiment. Two independent-samples t-tests on the location and scale parameters indicated that performance did not differ significantly between the two groups (for μ, M_{NP} = 39.6(3.74), and M_{HH} = 40.6(1.75); t(10) = -0.59, p = .567; for σ, M_{NP} = 5.73(2.21), and M_{HH} = 8.07(2.03); t(10) = -1.89, p = .087). That is, performance after only ten sessions in the High-High condition of the present experiment was similar to that of pigeons with no previous experience with the task and exposed to the condition for 40 sessions. Thus, the absence of an effect of the overall reinforcement rate in the present study does not seem to have been due to the peculiarities of the experimental design.

Regarding the conditions with different reinforcement probabilities in each half of the session, we predicted biases towards the richer alternative, biases evinced by shifts in the psychometric function. For every bird, we observed later switching points when the first half of the session had a higher proportion of reinforced trials, and earlier switching points when the second half had the higher proportion of reinforced trials. These results are consistent with hybrid models of timing such as LeT or BeT, but not SET, at least not in its original form (see Machado & Guilhardi, 2000).

Although we did not anticipate any systematic changes in sensitivity as a consequence of our manipulations, there was a robust change in the slope of the psychometric functions when S1 was richer than S2. This result is the opposite of that reported by Cowie et al. (2016) who found an increase in discriminability when the second half of each FOPP trial was richer than the first (S2 richer than S1). The authors argued that (a) the distribution of reinforcers across each trial determines temporal discrimination, and (b) such discrimination improves when the reinforcement rate is higher in the second half of the trial. In the present experiment, the MSR task avoided the asymmetries in the distribution of the reinforcers because the reinforcers were response-, not time-dependent. Thus, the finding that sensitivity did not change in the Low-High condition, as Cowie et

al.'s findings predicted, could be simply because the MSR task maintained the reinforcers more or less evenly distributed during the low-reinforcement half of the session. Yet the significant changes in sensitivity observed in condition High-Low remain unexplained. In fact, the coefficients of variation decreased significantly (average, 0.07) and fell below the typical range (from 0.2 to 0.3) observed in the other conditions and in other timing tasks. Either we assume that in condition High-Low the pigeons continued to time the moment of reversal, and then conclude, from the extremely low coefficients of variation, that they violated the scalar property of timing (Gibbon, 1977), or we assume that local cues, rather than time, controlled switching. Given the generality of the scalar property, the latter hypothesis seems more plausible and parsimonious (but for violations of the scalar property see Grondin, 2014; Bizo, Chu, Sanabria, & Killeen, 2006; and Zeiler & Powell, 1994). For some reason, then, in condition High-Low, switching came under the control of local cues, namely the food/no food outcomes of S1 choices.

Our results suggest that the pigeons use alternative strategies when the difference between the reinforcement probabilities for S1 and S2 is positive (condition High-Low) or negative (condition Low-High). In the latter case, their strategy remains consistent with (biased) timing; in the former case, their strategy seems inconsistent with timing and consistent with reward following. The asymmetry suggests that the local cues –the choice outcomes- may be harder to use as guides to action when changing from a low to a high density of reinforcement within a session than when changing in the opposite direction. With a low proportion of reinforced trials in the first half of the session, the pigeons seem to rely on the passage of time to guide their switching behavior. With a higher proportion of reinforced trials in the first half, the pigeons seem to rely on local cues, continuing to choose S1 until one or more trials without food and then they switch to S2. Surprisingly, perhaps, in condition High-Low, the pigeons finally resort to the win-stay/lose-shift strategy that is conspicuously absent from the condition in which it was most expected, High-High.

We also found a surprising dynamic effect when each half of the session comprised a different proportion of reinforced trials. With every change in condition, the location of the psychometric functions first swung away from the reversal point and then swung back, reducing the distance to the reversal point. This dynamic effect was robust, for it occurred for each pigeon and in each transition between the conditions with different reinforcement

rate (High-Low and Low-High). This suggests that pigeons are able to rapidly adjust their behavior to abrupt changes in reinforcement rate when reliable local cues are available.

In what follows, we ask whether a timing model could account for the dynamics of behavior within each condition of the present experiment. Because LeT provided a reasonable account of performance in the FOPP and other timing tasks (Machado, 1997; Machado et al., 2009), we asked whether it could deal with at least the main properties of responding in the MSR task.

The LeT Model in the MSR task

The LeT model is basically a derivative of Killeen and Fetterman (1988) Behavioral Theory of Timing (BeT). It consists of three major components, a set of sequentially activated *behavioral states*, a *vector of associative links* that change in real time according to a learning rule, and a *response rule* that determines the emission of one of the discriminated responses (Machado, 1997).

Machado and Guilhardi (2000) showed how LeT could account for performance in the FOPP with differential reinforcement rates on each half of the trial. Briefly, the shifts in the psychometric curves are caused by the animal's responsiveness to differences in reinforcement rate at specific times. In LeT's terminology, the strength of the associative links between behavioral states and operant responses reflect the probability of reinforcement for each response at each moment. They also demonstrated that LeT is able to account for the results reported by Bizo and White (1995) and Stubbs (1980).

Although the original formulation of LeT (Machado, 1997) offered a good description of performance in most timing tasks, a common criticism was that it did not comply with the scalar property of time perception. To address this issue, Machado et al. (2009) introduced a hybrid model that combines the scalar inducing features of SET with the learning structure of LeT. One of the main differences between the original and the hybrid LeT model is that in the latter only one state is active at a time. Although the essence of the model remained the same, this and other minor changes broadened the scope of the model (see Carvalho, Machado & Vasconcelos, 2016).

In the MSR task, the model works as follows. At the onset of each session, a random sample from a normal distribution with mean μ and standard deviation σ sets the speed of transition across states. To illustrate, suppose the sample equals 0.1. This means

that the states will be activated serially at the rate of 0.1 states per second or, equivalently, that each state remains active for 10 seconds (i.e. the first state is active from t = 0 to t = 10 s, the second state from t = 10 to t = 20 s, and so on). When a trial starts with the illumination of the two keys, one state, say, n, is active. This state is linked to the instrumental responses S1 and S2, with strengths $W_{(n, S1)}$ and $W_{(n, S2)}$, respectively, each a number between 0 and 1. The animal will choose according to the relative strengths of the links: S1 with probability $W_{(n, S1)}/[W_{(n, S1)}+W_{(n, S2)}]$, and S2 with the complementary probability. If the choice is rewarded, the link of state n with the reinforced response increases and the link of state n with the other response decreases; the magnitude of the changes depends on the reinforcement parameter β. If the choice is not rewarded the link of state n with the non-reinforced response decreases and the link of state n with the other response increases; the magnitude of the changes depends on the extinction parameter α (see Machado et al., 2009, for further details).

To obtain the model predictions, we simulated individual pigeons' performance maintaining the exact same structure of the present experiment regarding number of trials, sessions, and order of the conditions for each one of them, while trying to vary the least number of model parameters across subjects. Two parameters were fixed for all subjects, the initial weights of the associative links, $W_{0 =} .5$ for all states, and the reinforcement parameter β = .95, a much higher value than in previous simulations to allow for a relatively fast acquisition. The other three parameters were slightly different across subjects. The extinction parameter (α) was always low to prevent extinction, particularly on the first and last trials. The average speed of transition across states (μ) had a low value to allow greater residence time (i.e., more trials) in each state compared to similar simulations, because the to-be-timed duration in the MSR task is much longer than that of other timing tasks such as the FOPP or the bisection task. The standard deviation of the rate of transition across the states, σ, also varied and in every case yielded a reasonable coefficient of variation. Appendix A shows the parameters of the model used to simulate individual birds' data.

For each pigeon, we repeated the simulation one hundred times, averaged its output, and then fit Equation 1 to the data. In other words, we analyzed the model's output in the same way as the pigeons' data. Appendix B shows the maximum-likelihood estimated parameters of the fitted psychometric functions

Figure 3.5 compares the average of the location and scale parameters of the functions fitted to the pigeons' data and the model's simulations, as well as the coefficient of variation for each condition. The lines represent the 95% confidence intervals. In general, the two sets of estimated parameters did not differ appreciably. In particular, the model's location estimates were fairly accurate for all conditions. However, there was a major discrepancy in the scale parameter and, consequently, in the coefficient of variation for the High-Low condition, which supports the idea that timing may not have been the main process in this condition. Figure 6 compares the model's predictions with the actual pigeons' data. The curves represent the average of the psychometric functions produced by LeT. The symbols show the pigeons' average performance.

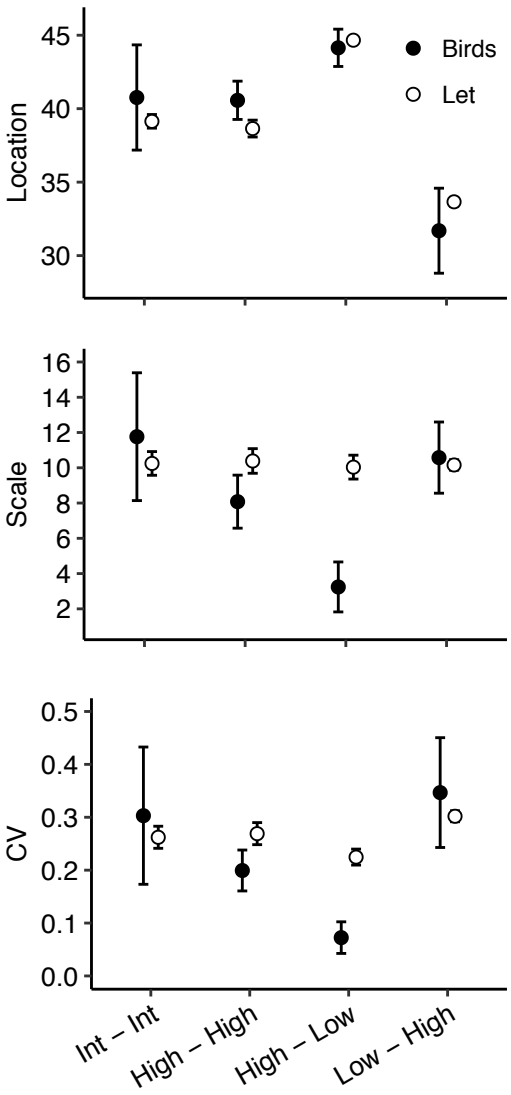

Figure 1.5. Average maximum-likelihood estimates of the location (μ) and scale (σ) parameters of Equation 1 and the coefficient of variation (CV= σ/μ) for the LeT model and the birds in each experimental condition. Error bars represent the 95% confidence interval.

Overall, LeT effectively captured the major performance features in the MSR task. On the one hand, when the proportion of reinforced trials was the same in both halves but the overall reinforcement rate differed (top panel), LeT captured well the smooth transition in preference from S1 to S2. The result strengthens the idea that timing is involved in the standard MSR task and shows one way in which pigeons could learn to choose based on time since the session started. On the other hand, when the overall reinforcement rate remained constant but the proportion of reinforced trials differed between halves (middle panel), LeT captured both the leftward shift in the function and the smooth, time-based, change in preference across trials when

the probability of reinforcement was higher in the second half. However, the model could not account for the change in sensitivity observed when the probability of reinforcement was higher in the first half.

The bottom panel of Figure 1.6 compares the pigeons' average performance in the early parts (first ten sessions) of the Low-High and High-Low conditions to assess the model's predictions on the dynamic effect observed in the transition between conditions. For the Low-High condition, the model's performance is far from accurate in terms of the location of the curve, although its shape is similar to that of the observed performance. For the High-Low condition the model shows a curve that is shifted towards the reversal point with a shape that resembles the observed performance, especially during the second half of the session. Interestingly, this difference in the shape of the observed psychometric functions in the early and late part of the High-Low condition indicates that during the first 10 sessions the birds showed the traditional pattern of time-regulated behavior and, with training, their behavior changed to a win-stay/lose-shift strategy.

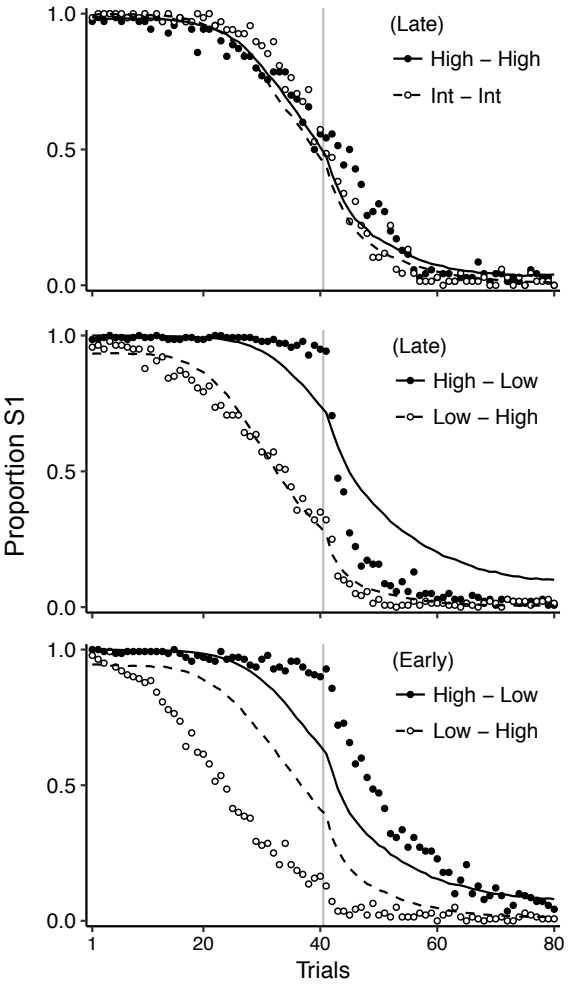

Figure 1.6. Birds' performance and the LeT model predictions in all conditions. Symbols represent the average of the birds' performance and lines represent the average of the individual performances predicted by LeT. The top panel shows the performance in the late part of those conditions with different overall reinforcement rate. The middle panel shows the performance in the late part for those conditions with different relative reinforcement rate on each half of the session. The bottom panel shows the corresponding performance for the early part of the conditions displayed in the middle panel.

The dynamic strengthening and weakening of the associative links is what allows LeT to account for temporally regulated performance in the MSR task. Nevertheless, at least two limitations remain. First, as just alluded, LeT is unable to describe the sharp

transition from responding to S1 to responding to S2 in the High-Low condition. This difficulty could be circumvented if one assumes that, in this condition, pigeons are resorting to a win-stay/lose-shift strategy. Were that the case, the finding that a timing model is unable to capture the main features of behavior not controlled by time is unsurprising. Obviously, this begs the question of what are the conditions under which such strategy is deployed or not. The second difficulty deals with the dynamic effects found in transitions. There is a major discrepancy between the data and the model in the first sessions of the conditions with different probabilities of reinforcement in each half. The data show a large bias in the first ten sessions of each condition, a bias that is reduced in the last ten sessions. The model predicts the opposite trend. In the MSR task, pigeons show a great sensitivity to abrupt changes in reinforcement rate that the model is unable to capture because it relies in gradual changes of the associative links.

Altogether, our findings indicate that, at least under some conditions, timing processes are involved in the MSR as previous research suggests (for a review, see Rayburn-Reeves & Cook, 2016). The changes in psychometric functions that we observed by manipulating the relative rate of reinforcement are at least partially consistent with this hypothesis. Yet the complexities of results in this experiment evidence that very little is known about how response outcomes and time interact to determine behavior. As other researchers have pointed out (e.g., Cowie, Davison, Blumhardt & Elliffe, 2016; Guilhardi, Yi, & Church, 2007; Jozefowiez et al., 2009; Kirkpatrick & Church, 1998), the difficulties and limitations in accounting for the data speak loudly for the need of learning theories that integrate associative and timing processes.

Study II

Past outcomes and time flexibly exert joint control over midsession reversal performance in the rat [2]

Abstract

In a midsession reversal task, subjects choose between two stimuli on every trial; only responses to one stimulus are reinforced. Halfway throughout the session, contingencies are reversed: previously reinforced responses are now extinguished and vice versa. Both, the outcome of the previous trial and the time elapsed since the beginning of the session, may predict the availability of reinforcement and determine choice. Thus, this task has typically been used to study cognitive flexibility and the temporal organization of behavior. This study assessed how past outcomes and time interact for behavioral control when each cue predicts the availability of reinforcement to a different extent. Eight rats were trained in four variations of the midsession reversal task differing in the reliability of outcomes and time as predictors of the reinforced response. We manipulated the reliability of the outcomes by providing either continuous or partial reinforcement, and the reliability of time by fixing the moment of reversal (middle of the session) or making the reversal unpredictable (semi-random trial). Results suggest that behavioral control alternates between outcomes and time according to the relative reliability of each cue. Model simulations show that outcomes and time may jointly determine behavior, and that momentary reinforcement rate may determine their relative influence.

Keywords: midsession reversal, reinforcement, timing, outcomes, win-stay/lose-shift, rat.

[2] This chapter reproduces the publication: Santos, C. & Sanabria, F. (2020). Past outcomes and time flexibly exert joint control over midsession reversal performance in the rat. *Behavioural Processes, 171,* 104028.

A prolific area of research on behavioral flexibility has focused on understanding the nature of the competition between simultaneous discriminative stimuli over control of behavior (for a review, see Rayburn-Reeves and Cook, 2016). Such competition has been examined using the midsession reversal (MSR) task. On each trial of this task, animals have a choice between two stimuli; responses to one of them (S1) are reinforced during the first half of the experimental session; responses to the other one (S2) are reinforced during the second half of the session. Because there is no explicit stimulus indicating the transition from the first to the second half of the session, there are only two discriminative stimuli that can gain control of behavior in this task: the outcome of the previous response, and the time into the session[3]. Thus, the MSR task is a suitable preparation to study how past outcomes and time combine or compete for behavioral control when they are both predictors of a potential reinforcer.

Control by outcomes in the MSR task may be expressed as a win-stay/lose-shift (WSLS) strategy, repeating the previous choice if it was reinforced and switching away if it was not. This strategy yields all available reinforcers minus one. Control by time may be expressed as a sigmoidal function relating the probability of responding on S1 to the number of trials, centered on the middle of the session. Assuming that time estimation involves some degree of error (Gibbon, 1977), such a function entails anticipatory errors (i.e., responding to the S2 before the reversal) and perseverative errors (i.e., responding to S1 after the reversal; McMillan, Spetch, Roberts, & Sturdy, 2017). Unlike the WSLS strategy, timing the reversal trial results in a substantial loss of potential reinforcers.

In the MSR task, when both outcomes and time inform which response will be reinforced, it is unclear how much control over choice may be attributed to each of these cues. There is evidence of behavioral control by previous outcomes (Laude et al., 2014; McMillan, Kirk, & Roberts, 2014; Rayburn-Reeves, Moore, Smith, Crafton, & Marden, 2018; Rayburn-Reeves et al., 2011; Rayburn-Reeves, Stagner, Kirk & Zentall, 2013), but there is also evidence of control by time (Cook and Rosen, 2010; McMillan & Roberts,

[3] Although the passage of time is correlated with the number of trials, there is evidence that, when these two sources of control are uncoupled, time is more likely to gain control over choice in the midsession reversal task in pigeons (Cook and Rosen, 2010; McMillan and Roberts, 2012; McMillan 2015).

2012; McMillan & Spetch, 2019; McMillan, Sturdy, Pisklak, & Spetch, 2016, Laude, Stagner, Rayburn-Reeves, & Zentall, 2014; Rayburn-Reeves, Laude, & Zentall, 2013; Rayburn- Reeves, Molet, & Zentall, 2011, Rayburn-Reeves & Zentall, 2013; Smith, Pattison, & Zentall, 2016).

Although flexible shifts in behavioral control have been shown in rats and pigeons (Laude, et al., 2014; McMillan, et al., 2014; Rayburn-Reeves, et al., 2013), past outcomes appear to exert stronger control over choice in rats, whereas time appears to exert stronger control in pigeons. Rat studies, however, are scarce and some of their results are equivocal (McMillan et al., 2014; Rayburn-Reeves, Moore, Smith, Crafton, & Marden, 2018; Rayburn-Reeves, Stagner, Kirk, and Zentall, 2013; Smith, Pattison, and Zentall, 2016). For instance, Rayburn-Reeves and colleagues (2013) found no evidence of disruption on choice accuracy in rats when the reversal location varied between sessions. In contrast, Smith and colleagues (2016), using the same procedure, equipment, reinforcers, and subjects of the same species, strain, and breeder, showed increased anticipatory and perseverative errors when the reversal trial was unpredictable. These studies provide contradictory evidence regarding the role of time on rat MSR choices when previous outcomes reliably predict the availability of reinforcement.

A potential approach to disentangle the sources of control on rats' MSR performance involves intermittent reinforcement. Rats following a strict WSLS strategy would perform poorly in a MSR task in which reinforcement is intermittent, because non-reinforcement of one choice is not a reliable predictor of future reinforcement of the other choice. Instead, the number of consecutive non-reinforced choices and the passage of time may provide better guidance to minimize errors in an intermittently reinforced MSR task. Therefore, to the extent that time may acquire control over MSR performance in rats, it may be revealed as anticipatory and perseverative errors clustered around the reversal in an intermittently reinforced variant of the task. However, the intermittently-reinforced MSR task has only been tested in pigeons, showing that, as long as correct responses for S1 and S2 are equally reinforced, performance in pigeons remains under strong control of time (Santos, Soares, Vasconcelos, & Machado, 2019; Zentall, Andrews, Case, and Peng, 2019).

Rats transitioning from continuous to intermittent reinforcement in a MSR task may adopt various strategies to maximize reinforcement. They may, for instance, persist

on a WSLS strategy, forgoing a substantial amount of reinforcement. Alternatively, they may switch to a time-based strategy, generating anticipatory and perseverative errors. Another possibility is for rats to use the number of consecutive non-reinforced trials to choose when to switch between S1 and S2, because the likelihood of a contingency reversal increases with this number. Finally, rats may adopt a combination of these strategies. Studies on cognitive flexibility using set-shifting tasks show that rats may perform simple simultaneous discriminations by shifting behavioral control between visual, auditory, olfactory, tactile or spatial cues and response strategies (for a review see Izquierdo et al., 2017). However, to the best of our knowledge, shifts between control by time and by past outcomes have not been explored.

The goal of the present study is to assess how past outcomes and time interact for behavioral control in rats when (1) they are both good predictors of reinforcement, (2) one of them is a better predictor than the other, or (3) they are both poor predictors of reinforcement. More specifically, we exposed a group of rats to four different versions of a MSR task, in which the reliability of past outcomes and time as predictors of reinforced choice was factorially manipulated. The reliability of past outcomes was manipulated by varying the probability of reinforcement (1 or .5) of correct responses. The reliability of time was manipulated by fixing or varying across sessions the trial at which contingencies were reversed.

Methods

Subjects

Eight male Wistar rats (*Rattus norvegicus*) from Charles River Laboratories (Hollister, CA) approximately 200 days old with previous experience in a switch-timing task. The rats had been paired-housed since their arrival on postnatal day 60. Rats were kept on a 12:12 h light cycle, with behavioral training being conducted during the dark phase. Water was always available in their home cage, and food was restricted to 1 h/day 30 min after experimental training, such that at the beginning of the next session their weights were about 85% of their average *ad libitum* weight, estimated from growth charts provided by the breeder. All animal handling procedures followed National Institute for

Health guidelines and were approved by the Arizona State University Institutional Animal Care and Use Committee.

Apparatus

The study used eight standard Med Associates (St. Albans, VT, USA) modular chambers 30 cm x 24 cm x 21 cm (length x width x height), with acrylic walls for the ceiling, front, and back panels, aluminum walls for the side panels, and the floor had a metal grid. On the left panel, a house light centered at the top provided general illumination. On the right panel, access to a dipper (ENV-202 M-S) was provided through a 5 x 5 cm opening centered horizontally along the wall, 1.5 cm above the floor. The dipper was fitted with a 0.01 cc cup (ENV-202C) to hold the liquid reinforcer (a 33% solution of Kroger® sweetened condensed milk diluted in tap water). On the same panel, the opening receptacle was also equipped with a head-entry detector (ENV-254-CB) with a temporal resolution of 10 ms. Two retractable levers (ENV112-CM) were located on each side of the opening receptacle 2.1 cm above the floor. A personal computer controlled and recorded all experimental events with Med-PC IV software.

Procedure

Each session comprised 80 trials in which responses to one lever (S+) were reinforced with probability *q*, and responses to the other lever (S-) were never reinforced. Regardless of whether they were reinforced or not, responses on S+ were deemed *correct,* and those on S-, *errors*. Reinforcement contingencies were reversed once during the session; the originally reinforced lever was no longer reinforced and the originally non-reinforced lever was now reinforced. For half of the rats the left lever was the first S+ (S1) and the right lever the second S+ (S2); the opposite was true for the other half of the rats.

All sessions started with the illumination of the house light, the extension of the levers, and a tone signaling the availability of a free reinforcer. Two seconds after a head entry was recorded in the magazine opening, the house light went off; 5 s later, the first trial began. The onset of the house light signaled the beginning of a trial; the first response to either one of the two levers turned off the house light. If the response was emitted on the S+ lever on a reinforced trial, a 2-s access to the reinforcer was signaled by a tone, followed by a 3-s intertrial interval (ITI) with no programed events. If the trial was not

reinforced or if the response was emitted on the S- lever, a 5-s ITI followed. At the end of each 80-trial session, subjects were removed from the experimental chamber, returned to their home cage for approximately 5 min, and then returned to the experimental chamber to start a new session. Each rat completed five sessions per day.

All subjects experienced four consecutive experimental conditions of 50 sessions each (because the rats completed 5 sessions a day, they experienced every condition for 10 consecutive days). The general structure of the task remained the same across conditions, except that (a) the trial in which contingencies were reversed (the *reversal trial*) could be fixed (at trial 41) or variable (between trials 16 and 66), and (b) the probability of reinforcement of a correct response, q, could be either 1 or .5. Table 2.1 shows how these two variables changed across conditions and the order in which all subjects experienced them. To confirm that our procedure replicated the typical performance of rats in the traditional version of the task, all correct responses were reinforced and the reversal trial was fixed at trial 41 in the first condition (F100). Only one procedural parameter changed between consecutive conditions, starting with the manipulation that motivated this study, a reduction in q (F100 → F50), then adding variability to the reversal trial (F50 → V50), and finally increasing q (V50 → V100).

Table 2.1.
Reinforcement schedule and variability of the reversal trial in each experimental condition.

Condition	Reversal	q
F100	Fixed	1
F50	Fixed	.5
V50	Variable	.5
V100	Variable	1

Note. Experimental conditions are listed in the order in which all rats experienced them.

In the conditions with $q = 1$ (F100 and V100) every correct response was reinforced, whereas in the conditions with $q = .5$ (F50 and V50), at the beginning of every trial, the computer determined with a probability of .5 if a correct response would be reinforced; errors were never reinforced. When the reversal was fixed (F100 and F50), trial 41 was the first trial on which responses on S2 were reinforced; when the reversal was variable (V100 and V50), the first trial on which responses on S2 were reinforced varied between sessions.

In V100 and V50, the reversal trial was determined for each rat at the beginning of each session by randomly sampling without replacement from one of two lists. List A was used in sessions 1 to 40; it consisted of all numbers from 16 to 66 excluding those in list B and number 41, for a total of 40 possible reversal trials. List B was reserved for testing performance in the last 10 sessions; it consisted of numbers 16, 21, 26, 31, 36, 46, 51, 56, 61, and 66. The number selected in a session corresponded to the reversal trial. This procedure was independently repeated for each of the two conditions with variable reversal trial.

Inferential Statistics.

All repeated measures ANOVAs were conducted on logit-transformed data: $\text{logit}(E) = \ln(E + 0.5)/(1 - E + 0.5)$, where E is the probability of making an error in a block of five trials.

The effect of the probability of reinforcement and the variability of the reversal was analyzed using a repeated measures ANOVA with q (1 vs. .5), reversal (fixed vs. variable) and blocks (6 blocks of 5 trials) as factors. To analyze the effect of the location of the reversal on performance in the conditions with variable reversal (V100 and V50), we performed a repeated measures ANOVA for each condition with location (early vs. late) and blocks (6 blocks of 5 trials) as factors. To compare the effect of reinforcement probability on anticipatory and perseverative errors in the sessions with an early reversal, we performed 2 repeated measures ANOVA (q x blocks) one on the last 3 blocks of trials before the reversal in V100 and V50 (anticipatory errors), and the other one on the first 3 blocks of 5 trials after the reversal in V100 and V50 (perseverative errors).

Results

Performance was characterized as the average proportion of responses to S1 per trial in the last ten sessions of each condition. All rats showed the same general pattern: exclusive responding to S1 at the beginning of the session and exclusive responding to S2 towards the end, although there was a distinct pattern of responding around the reversal in each condition. Therefore, for simplicity, all analyses were based on the 30 trials around the reversal (15 before and 15 after contingencies were reversed, including the reversal

trial). Figure 2.1 shows the average performance aligned at the reversal and averaged across trials for all rats.

In both conditions with 100% reinforcement for correct responses (q = 1; solid lines), performance transitioned from exclusive responding to S1 to exclusive responding to S2 abruptly. In these conditions (F100 and V100) there were virtually no anticipatory errors (before the reversal) and very few perseverative errors (after the reversal). Conversely, in the conditions with 50% reinforcement (q = .5; dashed lines), performance transitioned rather slowly, showing modest anticipatory errors and a considerable amount of perseverative errors.

The data in 2.1 are represented in Figure 2.2 as the proportion of errors in blocks of five trials. This figure shows a higher number of anticipatory and perseverative errors in the conditions F50 and V50 relative to F100 and V100 [$F(1, 7) = 178.5$, $p < .001$]. Importantly, it also brings to light a negligible number of errors in the first block of trials after the reversal in conditions F100 and V100. More specifically, a proportion of errors of about .25 in block one of conditions F100 and V100 indicates that, in most sessions, there was only one error in the five trials comprised in the block, possibly in the

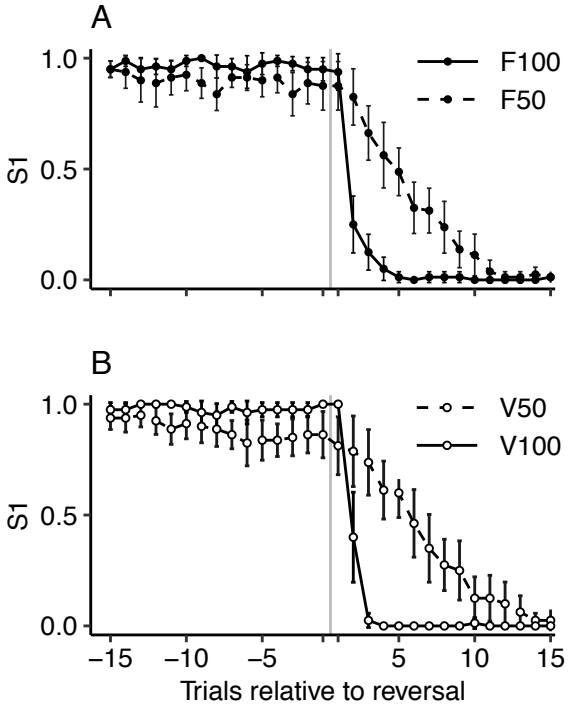

Figure 2.1. Average proportion of responses to S1 per trial relative to the proximity of the reversal in the last ten sessions of each condition. The solid lines represent the conditions with continuous reinforcement (q = 1), and the dashed lines the conditions with partial reinforcement (q = .5). Error bars represent 95% CI. Panels A and B show performance in the conditions with fixed and variable reversals, respectively. The grey vertical line indicates the moment between the last trial before the reversal and the first trial after the reversal.

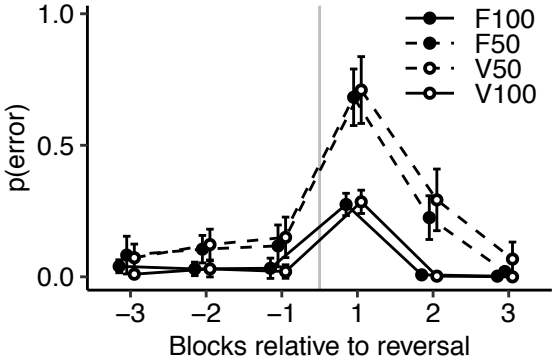

Figure 2.2. Proportion of errors in blocks of 5 trials relative to the reversal in each condition. The grey vertical line represents the location of the reversal. Error bars represent the 95% CI.

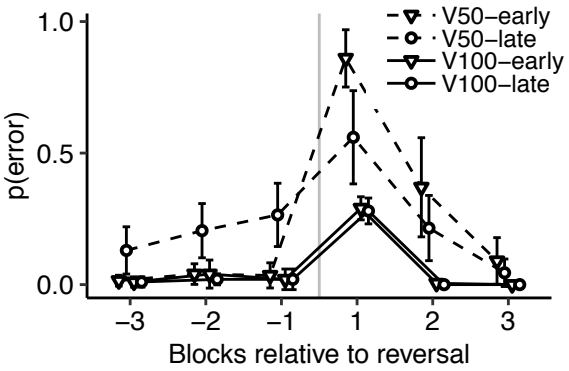

Figure 2.3. Proportion of errors in blocks of 5 trials relative to the reversal in conditions with variable reversal. Triangles represent performance when the reversal trial occurred early in the session (during the first in half); circles represent performance when the reversal occurred late in the session (during the second half). Error bars represent the 95% CI. The grey vertical line indicates the location of the reversal.

reversal trial, because of the absence of feedback at this point. To confirm this hypothesis, we calculated the probability of an error in the five trials following the reversal trial (in F100 and F50, these were trials 42-46), and found that the proportion of errors was M = .09, SEM = .046 for F100, and M = .09, SEM = .079 for V100. In contrast, for the conditions F50 and V50 the proportion of errors was M = .57, SEM = .084 and M = .64, SEM = .057, respectively.

Figure 2.3 further analyzes the data from conditions V50 and V100 showing the proportion of errors in blocks of trials according to whether the reversal occurred in the first (early) or second (late) half of the session. Performance in V100 was very similar regardless of whether the reversal occurred early or late in the session, consistent with a non-significant main effect of reversal location [$F(1, 7) = 0.38$, $p = .56$]. In contrast, performance in V50 showed substantially more anticipatory errors and fewer perseverative errors when the reversal occurred late than when it occurred early in the session, as confirmed by a significant interaction effect of reversal location and blocks [$F(1, 7) = 11.79$, $p < .05$]. Interestingly, in the V50 sessions with early reversal, the anticipatory errors were just as low as in the V100 sessions [$F(1, 7) = 0.13$, $p = .73$], but the perseverative errors were the highest of all [$F(1, 7) = 85.35$, $p < .001$]. These results indicate that the similarity in performance between F50 and V50 conditions shown in Figure 2.2 is an artifact of averaging V50 performance with early and late reversal trials. No such artifact was observed in V100 performance.

Discussion

In the contingency reversal paradigm implemented in this experiment, rats demonstrated near optimal performance when correct responses were always reinforced

(conditions F100 and V100). Under these conditions, performance showed almost no anticipatory errors and only a few perseverative errors (Figure 2.2) circumscribed to the reversal trial. This pattern of errors suggests that rats generally followed a WSLS strategy, as first shown by Rayburn-Reeves et al. (2013) and replicated by Smith, Pattison, and Zentall (2016). In both studies, when the reversal trial was fairly predictable (fixed halfway through the session) independently of the topography of the response, rats showed very high accuracy and virtually no anticipatory errors. Nonetheless, Smith, Pattison, and Zentall (2016) observed that when the reversal trial was variable, anticipatory errors were higher on those sessions with a late reversal (after trial 41) and perseveration errors were higher on those with an early reversal (before trial 41). The present study replicated this finding, but only when reinforcement was intermittent (Figure 2.3, V50 condition). Taken together, these results suggest that, similar to pigeons (Laude et al., 2014; McMillan et al., 2016; McMillan et al., 2015; Rayburn-Reeves et al., 2013; Rayburn-Reeves, et al., 2011; Rayburn-Reeves & Zentall, 2013; Santos et al., 2019), time exerts control over choice in the MSR task in rats, at least in the absence of unequivocal cues of the availability of reinforcement.

Consistent with Rayburn-Reeves et al. (2013), but not with Smith et al. (2016), the present study found no evidence of timing when every correct response was reinforced and the reversal trial was variable. There are three differences between these studies and the present experiment that may account for these inconsistent findings. First, whereas Smith et al. used female rats, Rayburn-Reeves et al. and the present study used male rats. It is important to note, however, that Rayburn-Reeves et al. did not analyze performance according to the moment of the session where the reversal occurred, but only averaged across all sessions, possibly masking differences between sessions with early and late reversals. Hence, although this hypothesis remains plausible, the idea that female rats are more likely to use temporal cues over a WSLS rule in a MSR task with variable reversal location is still tentative. Second, whereas Rayburn-Reeves et al. and Smith et al. selected variable reversal trials from a list of 5 possible trials, the present study selected them from a list of 50 possible trials. It is thus possible that rats were sensitive to the distribution of variable reversal trials, and that a higher dispersion of these trials weakened temporal control over choice. Finally, Rayburn-Reeves et al. and Smith et al. rats transitioned from a condition where time and outcomes were highly reliable and relatively redundant as cues

for the reinforced response, to a condition where the reliability of time was reduced (F100 → V100). In contrast, in the present experiment, rats transitioned from a condition where both time and outcomes were fairly unreliable cues to a condition where outcomes became highly reliable (V50 → V100). It is thus possible that just prior training with a reliable temporal cue yields strong temporal control of choice when other, previously prepotent cues, become unreliable. The extent to which these methodological differences account for divergent levels of temporal control of choice demand further research. Such research would benefit from controlling for potential order effects in experimental conditions, a control that extant research has so far neglected.

A Mixture Model of Midsession Reversal Performance

Considering all four conditions (Figure 2.2), it appears that the probability of reinforcement was the main variable governing changes in performance across conditions. At first glance it seems that, when the probability of reinforcement (q) was 1, choice was solely controlled by the just-preceding outcome, and when the probability of reinforcement was lower (q = .5), choice was controlled by the time elapsed from the beginning of the session. Whereas a WSLS strategy was optimal when every correct response was reinforced, it would have resulted in a large number of errors and a low number of obtained reinforcers when only some correct responses were reinforced. Conversely, relying exclusively on the passage of time when every correct response was reinforced would have resulted in a low number of reinforcers, due to anticipatory and perseverative errors. In this context, it is reasonable to propose that the number of obtained reinforcers was maximized by alternating between WSLS and timing strategies, according to the momentary probability of reinforcement.

We simulated performance in the MSR task with a simple model that assumes that behavioral control alternates between WSLS and timing strategies according to the momentary probability of reinforcement. When q = 1 (conditions F100 and V100), performance was simulated with a strict WSLS mechanism in which reinforced responses were repeated on the next trial, and non-reinforced responses were not. When q = .5 (conditions F50 and V50) performance was simulated with a *pure* timing model. On every trial, this timing model emulates the estimation of the elapsed interval from the beginning of the session (t) by selecting randomly from a normal distribution with mean equal to the

actual elapsed interval[4] (n), and a standard deviation (σ) proportional to n (conforming to Weber's Law). The model assumes knowledge of the average interval from the beginning of the session to the reversal (μ). To determine the response, the timing model follows a simple decision rule: if t is smaller than μ, respond on S1, otherwise respond on S2. There are no further assumptions about the nature of this process as timing or counting, because they can both be understood and explained in the same terms (Davison & Cowie, 2019).

The present experiment was simulated with 1000 replications of this alternating model, with $\mu = 41$ and Weber fraction of 0.25 (a reasonable estimate for several strains of rats in timing tasks; e.g., Orduña, Hong, & Bouzas, 2007). Figure 2.4A shows the predicted performance in every condition of the experiment (cf. Figure 2.2); Figure 2.4B shows the predicted performance in the conditions where the reversal was variable, contrasting earlier versus later reversals (cf. Figure 2.3). Whereas the WSLS model describes very well performance in the conditions where every correct response was reinforced (F100 and V100), the predictions of the timing component of this model for the conditions with intermittent reinforcement (F50 and V50) diverge substantially from observed performance. The timing process yields a symmetrical number of errors relative to the reversal, and predicts a large discrepancy in anticipation and perseveration errors relative to the moment in the session where the reversal took place.

An alternative and more parsimonious account of MSR performance assumes that behavioral control shifts after a fixed number of non-reinforced trials (L). This number is inversely proportional to the probability of reinforcement (q), and may be learned through the consequence of choices. In fact, if q were known and only non-reinforced trials were counted (i.e., there were no counting or timing from the beginning of the session), the strategy that minimizes errors involves always switching between alternatives after L non-reinforced trials, such that $1 - (1 - q)^L > 0.5$; when $q = 0.5$, such optimal L is 2 trials. It is unlikely, however, that animals strictly follow an optimal choice strategy; instead they may approximate optimality with various degrees of error, which demands a more flexible model of choice. Such model may assume L to be proportional to the odds against reinforcement,

$$L = k[(1 - q) / q] + 1. \quad (1)$$

[4] The term interval refers to time into the session or trial number without a distinction. For simplicity, we used the trial number.

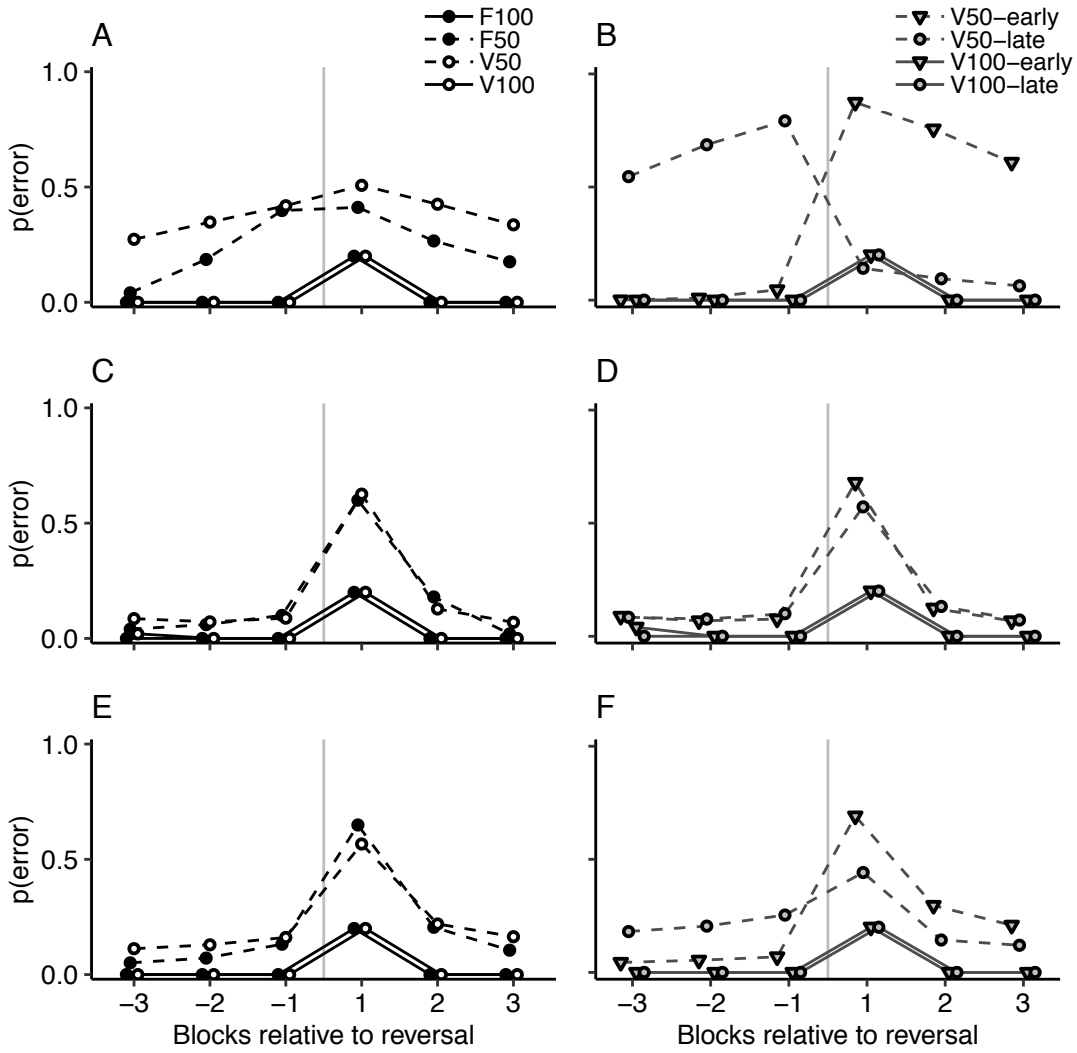

Figure 2.4. Predicted performance by an alternating timing-WSLS model (top panels), by a WSLSS model (middle panels), and by a mixture timing-WSLSS model (bottom panels). The left panels show performance predicted in every condition of the experiment; the right panels show performance in conditions with variable reversal according to the moment it occurred. The grey vertical line indicates the location of the reversal.

Equation 1 can approximate the optimality function that minimizes errors (with $k \approx 1$), and yields the strict WSLS strategy ($L = 1$) typically observed when $q = 1$ in conditions F100 and V100.

A simulation of this win-stay/lose-sometimes-shift (WSLSS) model with $k = 2$ (corresponding to $L = 3$ when $q = 0.5$), and without any timing component, reproduced the average pattern of performance around the reversal trial (Figure 2.4C; cf. Figure 2.2), but failed to reproduce key aspects of V50 and V100 performance (Figure 2.4D; cf. Figure 2.3). In particular, the WSLSS model produced neither the elevated anticipatory errors in late V50 trials nor the elevated perseverative errors in early V100 trials.

Last, we explored the possibility that control of MSR performance alternates between timing and non-timing states in the way originally proposed by Sanabria and Killeen (2008). Based on this idea, we formulated a model in which, on every trial, the subject enters either a timing state with probability p, or a non-timing state with probability $1 - p$. For the timing state we used the timing model described above, and for the non-timing state we used the WSLSS.

The bottom panel of Figure 2.4 shows the predicted performance of this mixture model ($k = 2$, $\mu = 41$, $\sigma = n * .25$, $p = 0$ when $q = 1$, $p = .1$ when $q = .5$). Figure 2.4E shows the predicted performance for each condition (cf. Figure 2.2), and Figure 2.4F the predicted data in the conditions with variable reversal according to the moment it occurred (cf. Figure 2.3). This mixture model of timing and WSLSS was able to account for the data in conditions F100 and V100 as well as the other models, and captured the similarity between the overall performance in F50 and V50. More importantly, it successfully reproduced the performance pattern observed in the rats in V50: higher anticipatory errors when the reversal occurred in the second half of the session compared to when it occurred in the first half, and the opposite pattern for perseverative errors. The similarity between simulated and observed data suggests that when past outcomes are a highly reliable cue, choice follows a WSLS rule, but when they are less reliable, additional cues gain behavioral control, such as the interval from the beginning of the session (timing) or the interval since the last reinforcer (WSLSS).

Altogether, these results suggest that rats learn multiple timing cues simultaneously: they learn the average interval at which the contingency reversal occurs, and the average interval since a particular response lead to the last reinforcer. In other words, MSR performance in rats appears to be a byproduct of tracking the interval from the beginning of the session to the moment of the reversal, remembering the outcome of previous trials, and selectively shifting behavioral control from one cue to another, according to the momentary probability of reinforcement.

In summary, when both outcomes and time are good predictors of the availability of reinforcement, the performance of rats in a MSR task is mainly controlled by the outcome of the previous trial in a WSLS fashion. Also, contrary to what has been observed in pigeons, rats' performance did not show any evidence of timing when past outcomes are better predictors of reinforcement. Furthermore, when time is a more reliable cue,

performance falls under joint control of the time elapsed from the beginning of the session, and of the more local distribution of reinforcers.

Study III

Control of behavior by time and by the outcome of the preceding response: A midsession reversal task reassessment in pigeons

Abstract

The goal of the present study was to explore how time and the outcome of the preceding response combine to determine choice in frequently changing environments. We used a midsession reversal task because it allows the independent manipulation of time and response outcomes as cues for future reinforcement. The task consists of a simple simultaneous discrimination where responses to one stimulus are reinforced and responses to the other stimulus are not; once throughout the session, contingencies reverse and the previously reinforced stimulus is now extinguished and vice versa. We exposed a group of pigeons to four conditions, with an unpredictable reversal point, and differing only in the payoff of the alternatives. The analysis of the effect of the location of the reversal and of the relative payoff of the alternatives on performance revealed dynamic and joint control of time and outcomes over behavior. Moreover, contrasting these results with computational models that combine timing and reinforcement learning algorithms challenges non-associative timing models to account for performance in this task.

Keywords: behavioral flexibility, stimulus control, midsession reversal task, time, past outcome, reinforcement rate, pigeons.

Discrimination reversal tasks are test beds of behavioral flexibility and their underlying processes, because they assess the degree to which an animal is able to learn a new discrimination opposite to one already learned (Bitterman, 1965; Strang & Sherry, 2014; van Horik & Emery, 2018). The midsession reversal (MSR) task is a variant of discrimination reversal tasks in which contingencies reverse only once throughout the session, typically halfway. On every trial of the MSR task, subjects have a choice between two stimuli; during the first half of the session responses to one stimulus (S1) are reinforced while responses to the other stimulus (S2) are not; and during the second half of the session, responses to S1 are extinguished and responses to S2 are reinforced. As in any other discrimination reversal task, to behave adaptively and earn most of the available rewards, it is necessary to attend to the cues that indicate the correct choice and to *ignore* previously learned contingencies once they are no longer active (Staddon, 2010). A heuristic that combines both of these conditions is the win-stay/lose-shift (WSLS) strategy: repeating the previous response if it was followed by reinforcement, and switching to the alternative response if it was not reinforced. This strategy maximizes payoff with minimal effort.

Humans and rats solve this task by following a WSLS strategy (Rayburn-Reeves, Mollet, & Zentall, 2011; Rayburn-Reeves, Stagner, Kirk, & Zentall, 2013; Santos & Sanabria, 2020). However, pigeons behave differently, especially when the task requires a non-spatial (e.g., color) discrimination: They choose accurately towards the beginning and the end of the session, but make a considerable number of anticipatory (responses to S2 before the reversal) and perseverative (responses to S1 after the reversal) errors in the middle of the session. The fact that errors cluster around the reversal suggests that behavior is time-regulated and that choices are not based on the local cues for food or no food, as if they were using a WSLS strategy. (Cook & Rosen, 2010; Laude, Stagner, Rayburn-Reeves & Zentall, 2014; McMillan & Roberts, 2012; McMillan, Sturdy, Pisklak, & Spetch, 2016; Rayburn-Reeves, Laude, & Zentall, 2013; Rayburn-Reeves, Mollet, & Zentall, 2011; Rayburn-Reeves & Zentall, 2013). Even though it is surprising that timing processes control performance in the MSR task, non-programmed time-regularities inevitably introduce a potential cue to signal the reversal: the time elapsed from the

beginning of the session. Because response latencies tend to be remarkably short, trials end up having roughly the same duration and, consequently, contingencies happen to reverse at a relatively predictable moment.

The most robust evidence of temporal control of pigeons' performance in the MSR task is a study by McMillan & Roberts (2012). The authors trained pigeons with a 5-s intertrial interval (ITI) and later tested them with either a 2.5-s or 10-s ITI. Consistent with the timing hypothesis, the pigeons tested with the shorter ITI reversed their preference from S1 to S2 much later than the midsession point, showing no anticipatory errors. Furthermore, the pigeons tested with the 10-s ITI, reversed their preference much earlier than the midsession: starting to show anticipation errors around the time into the session where the reversal occurred in the training phase.

To further examine the role of timing in the MSR task, Santos, Soares, Vasconcelos, and Machado (2019) varied the payoff of each of the alternatives making one leaner than the other by reducing its payoff from 100% to 20%, a manipulation known to bias time perception in temporal discrimination tasks (Bizo & White, 1995; Cambraia, Vasconcelos, Jozefowiez, & Machado, 2019; Machado & Guilhardi, 2000). When S1 was the leaner option, they expected performance to be biased towards S2, increasing anticipatory errors and decreasing perseverative errors. This result was confirmed. When S2 was leaner, they expected a reduction in anticipatory errors and an increase in perseverative errors. This result was only partially confirmed. There were scarcely any anticipatory errors and very few perseverative errors, a pattern more consistent a WSLS strategy. In this condition, behavioral control seemed to have shifted from time to the outcome of the preceding trial.

Zentall, Andrews, Case, and Peng (2019) replicated the results of Santos et al. (2019) in the condition where S1 had a higher payoff than to S2. In addition, they found that another manipulation devaluing S2 produced a similar result: requiring 10 pecks on S2 but only one peck for S1 rendered virtually no anticipatory errors and few perseverative errors. However, the idea that payoff asymmetry in the MSR task can shift behavioral control from one cue to another must be tested by assessing temporal control directly (Zentall, et al., 2019).

Other studies attempt to discourage temporal control of behavior in the MSR task. Rayburn-Reeves et al. (2011) trained pigeons on a variation of the task in which the

reversal was no longer fixed halfway through the session (i.e. the first S2+ always scheduled on trial 41 out of 80), but randomly located at one of 5 possible trials on every session (i.e., either on trial 11, 26, 41, 58, or 71). They found that even when the time elapsed from the beginning of the session no longer hinted on the active contingencies, performance continued to be time-regulated. In particular, choice was more accurate in the sessions where the reversal occurred at the middle of the session (trial 41) than in those where it occurred closer to the beginning or to the end. When the reversal occurred earlier in the session (trials 11 and 26), performance showed significantly more perseverative errors, and fewer anticipatory errors, and the opposite was true when the reversal occurred later in the session (trials 58 and 71). The authors concluded that, although the local history of reinforcement played some role in performance, timing or counting processes maintained some control as well and that pigeons tend to average the reversal locations over sessions.

However, their analysis showed that when characterizing performance as the proportion of S1 responses per trial according to the location of the reversal, all 5 curves followed the same trend from the beginning of the session up to the moment of the reversal, overlapping on the common trials. This is, the pre-reversal segment of the curves followed a concave downward function consistent with a timing process. Yet, the post-reversal segment of each of these curves did not overlap, they dropped (at what seemed to be) the same rate and remained parallel until reaching the lowest values. This result is inconsistent with a timing process and, more importantly, with the idea that performance is determined by the average of the reversals experienced in training, as suggested by the authors. If choice was guided by the average time of the experienced reversals, the pre-reversal segments of the curves with the later reversal (trials 58 and 71) should follow and overlap the trend of the average curve (reversal on trial 41), and this was not the case. In summary, even though the progressively increasing anticipatory errors in the sessions where the reversal occurred later indicates that performance was associated to time, the fact that the pre-reversal segment did not overlap the mid-session reversal curve shows that performance was not guided by the time of the average reversal. The fact that the post-reversal segments of the curves differed according to the location of the reversal shows that behavior was sensitive to the contingency change and was able to adjust accordingly within one session.

Smith, Pattison, and Zentall (2016) adapted Rayburn-Reeves et al. (2011) pigeons' procedure to rats and replicated their findings showing that timing processes also influenced rats' choices. Their results contradict the generalized belief that temporal control in the MSR task is only observed in pigeons. Santos & Sanabria (2020) replicated Smiths, Pattison, and Zentall's (2016) results in a MSR task with partial reinforcement (50%) for both alternatives with rats, and showed that making the reversal trial unpredictable can reveal behavioral control by time, a result that is not evident with a fixed reversal point.

Taken all together, both in rats and pigeons, performance in the MSR task, sometimes appears to be under joint control of the time elapsed form the beginning of the session and the outcome of the previous trial, and other times only determined by either one of these cues. Importantly, even the most popular theories of timing —such as the Scalar Expectancy Theory (Gibbon, 1977), the Behavioral Theory of Timing (Killeen & Fetterman, 1988) and the Learning-to-Time Model (LeT; Machado, 1997; Machado, Malheiro, & Erlhagen, 2009) —fall short accommodating these results for at least two reasons: (1) they assume that animals time events, whether or not they are explicitly reinforced for it (Killeen, Fetterman, & Bizo, 1997), but do not state the boundary conditions in which behavior is to be under temporal control or not (Carvalho, Machado, & Vasconcelos, 2016); and (2) they lack an auxiliary mechanism to adjust behavior within a session when contingency changes are not predictable, as in the study by Rayburn-Reeves, Molet, and Zentall (2011).

The purpose of this study is two-fold. Empirically, we aim to assess the sources of behavioral control in the MSR task when the cues for the availability of reinforcement are differently reliable. We particularly aimed to test the generality of the apparent absence of temporal control and use of a WSLS strategy when S1 had a higher payoff than S2 in the MSR task (Santos et al., 2019). Theoretically, we aim to compare animal behavior data with the prediction of different models to determine how well their timing mechanism integrates with other stimulus control processes to account for MSR performance.

Therefore, we exposed a group of pigeons to a MSR task with changing probabilities of reinforcement and made the reversal unpredictable on every session. By varying the payoff probability of each alternative we manipulated the reliability of the outcome of a response (food or no food) as a predictor of the availability of a reinforcer on

the next trial. By making the reversal variable, time into the session was a highly uncertain cue of the active contingencies. Differences in performance in early- versus late-reversal sessions would expose temporal control of behavior —that might not be evident when the reversal is fixed—, whereas differences in performance before versus after the reversal would reveal the effect of the outcome of previous trials. The additional complexity of the task is compensated by the possibility of separating the effects of time and of the outcome of the preceding trial.

To illustrate this idea, suppose that a subject is trained in a MSR task for several sessions with a different and unpredictable reversal every time and that, after enough experience with the task, we compare performance in a session where the reversal occurs early, say, on trial 21 versus a session where the reversal occurs later, say on trial 61. Figure 3.1 contrasts the performance predicted by two different mechanisms: a WSLS strategy (panel A) and a pure timing model (panel B). Performance is represented as the probability of S1 responses per trial in these two sessions; the dashed lines represent the early reversal session (i.e. trial 21) and the solid line the late reversal session (i.e. trial 61).

Because the moment of the reversal is unpredictable, the optimal strategy would be to follow a WSLS strategy. Panel A shows the expected performance if a subject were to follow this strategy. In both, early and late reversal sessions, performance is the same *around* the moment of the reversal: exclusive responding to S1 (i.e. no anticipatory errors) before the reversal, and one error right after the reversal (which is expected giving the absence of negative feedback up

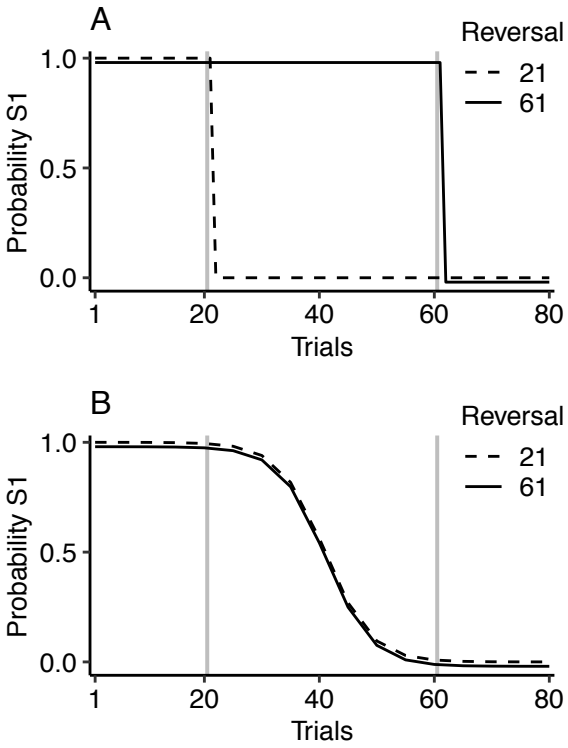

Figure 3.1. Predicted versus observed performance in two different sessions of a MSR task with a variable reversal. The dashed lines represent performance in a session with the reversal on trial 21 and the solid lines represent performance in a session with the reversal on trial 61. The vertical grey lines represent the reversal points. Panel A shows the predicted performance by a win-stay/lose-shift strategy. Panel B presents the predictions of a pure timing model.

to this point) followed by a dramatic drop of S1 responses. However, on the trials *between* the two reversals, performance is drastically different. For instance, on trial 41, in the early reversal session (where the reversal already occurred on trial 21) the probability of responding to S1 is 0, but in the late reversal session (where the reversal is programmed to occur on trial 61), the probability of responding to S1 is 1. In this sense, the pattern of performance produced by the use of a WSLS strategy in early and late reversal session is similar in the trials around the reversal, but different in the trials in between the reversals.

Panel B of Figure 3.1 shows the predictions of a pure timing model. Regardless of the location of the reversal, the pattern of performance is exactly the same for both sessions. Different timing models describe their own mechanism to keep track of the interval elapsed from the beginning of the session to the moment of the reversal and determine choice, but they predict the same outcome. Because behavior is explained by the temporal regularities of the environment, these models describe an organism that behaves as if events were temporarily regular. Thus, in a MSR task with variable reversal, if behavior were solely under temporal control it would show the same pattern as if the reversal was fixed at the average point of all reversals.

Because behavioral control by time and by the outcome of the preceding trial engender distinctive patterns of behavior, making the reversal variable allows us to directly test the joint control of time and past outcomes over choice. On one hand, if behavior were to be under temporal control, the determinant of choice would be the average time at which the reversal has had occurred in the past. Thus, (1) when comparing performance relative to the reversal, we should see different patterns of responding when the reversal occurred early in the session versus when it occurred late, and (2) when comparing performance on each trial across sessions, we should see similar responding regardless of the occurrence of the reversal. On the other hand, if there were no temporal control and performance was only determined by the outcome of the preceding response, the main determinant of choice would be the momentary payoff probability. Hence, (1) when comparing performance on the trials around the reversal, we should see similar performance regardless of the location of the reversal, and (2) when comparing performance on each trial number across sessions, we should see a difference depending if the reversal had already occurred or not.

Method

Subjects

Eight pigeons (*Columba livia*), with previous experience in a matching-to-sample task, participated in this study. All birds were individually housed with free access to water and grit in a temperature and humidity controlled room on a 13:11 light/dark cycle. They were kept at 80 – 85% of their ad libitum weight and supplementary feed with mixed grain after the experimental sessions when necessary.

Apparatus

The study used four identical standard Med Associates chambers for pigeons of 31.8 long, 25.4 cm wide, and 34.3 cm high enclosed in a sound attenuating cubicle equipped with a fan for ventilation and masking noise. The ceiling, back, and front walls of the experimental chamber were acrylic; the left wall was made of metal aluminum panels; metal rods comprised the floor; and the right wall consisted of the response panel with three circular response keys and a food hopper. Response keys were 2.5 cm of diameter, centered 20.5 cm above the floor, separated from each other by 8 cm, and coupled with a 12-stimulus in-line projector (Industrial Electronics Engineers) with a 28-V light bulb. The food hopper was assembled on a horizontally centered 6 x 6.5 cm opening 4 cm above the floor and illuminated with a 28-V light bulb when activated. Experimental events were programed and recorded using ABET II software (Lafayette Instruments®) on a personal computer.

Procedure

To reinstate the pecking response before starting the experimental procedure, all the pigeons completed two 80-trial sessions on a CFR schedule in which one of the two side keys was randomly illuminated with a red or green hue, and pecks were reinforced with a 2-s access to grain. Experimental sessions were scheduled seven days a week and consisted of 80 simultaneous color discrimination trials in which responses to one color key (S+) were reinforced with a certain probability, and responses to the other color key (S-) were never reinforced.

Each trial started with the illumination of the two side keys, one red and one green. The computer semi-randomly assigned colors to the keys so that each color appeared half of the trials on each side. The first peck to a key turned them both off. If the response was emitted on the S+ key on a reinforced trial, the hopper was illuminated and grain was accessible for 2 s, followed by a 3-s ITI. If the trial was not reinforced or the response was emitted on the S- key, a 5-s ITI followed. Reinforcement contingencies reversed once during the session; the originally reinforced color (S1) was no longer reinforced and the originally non-reinforced color was now reinforced (S2). For half of the pigeons S1 was red and S2 was green; the opposite was true for the other half. During the first part of every session, correct responses (S1) were reinforced with a probability q_1; and during the second part, correct responses (S2) were reinforced with a probability q_2.

To facilitate learning the task, all birds received 25 sessions of pre-training in which every correct response was reinforced ($q_1 = q_2 = 1$). The pre-training data was not considered for analysis. Next, all the birds completed four experimental phases of 25 sessions each. Table 3.1 shows the condition in each phase of the experiment with the values of q_1 and q_2. In each phase they experienced a different condition defined by the probability of reinforcement for S1 and S2 (q_1 and q_2, respectively): High = 1; Int = .5 (intermediate); and Low = .2. The first experimental phase introduced condition Int-Int ($q_1 = q_2 = .5$) to avoid any possible contrast effect from exposure to the conditions with different relative reinforcement rates ($q_1 \neq q_2$). The order of conditions High-Low ($q_1 = 1$; $q_2 = .2$) and Low-High ($q_1 = .2$; $q_2 = 1$) was counterbalanced across birds: half experienced High-Low in phase 2, and Low-High in phase 3, and the other half vice versa. Lastly, all birds experienced condition High-High in phase 4 of the experiment.

In each one of the 25 sessions of a phase, the reversal occurred on a different trial. That is, the first trial on which S2 was the S+ (and S1 became the S-) was randomly selected with no replacement from a list of numbers ranging from 16 to 66. To ensure that the reversals were evenly distributed throughout trials, in the pre-training and on phases 2 and 4, the reversal was sampled only from the odd numbers of the list (e.g., 17, 19, 63, 65), and on the remaining phases of the experiment (i.e., phases 1 and 3), the reversal was sampled only from the even numbers of the list (e.g., 16, 18, 64, 66). The even-number list had 26 items but only 25 were sampled for each bird (Appendix A lists the non-sampled trials in each case).

Table 3.1
Condition and values of q_1 and q^2 in each phase of the experiment.

Group	Phase	Condition	q_1	q_2
1	1	Int-Int	.5	.5
	2	Low-High	.2	1
	3	High-Low	1	.2
	4	High-High	1	1
2	1	Int-Int	.5	.5
	2	High-Low	1	.2
	3	Low-High	.2	1
	4	High-High	1	1

Note. The conditions' names refer to the probability of reinforcement for S1 and S2 (q_1 and q_2, respectively): High = 1; Int = .5 (intermediate); and Low = .2. Each group was confirmed by 4 pigeons.

Data Analysis

The data were analyzed in two different ways, one to assess temporal control and another to assess outcome control. To assess temporal control, for each phase of the experiment we focused on the trials around the reversal, and compared performance when the reversal occurred *early* in the session versus when it occurred *late* in the session. We categorized sessions as early when the reversal occurred between trials 16 and 32, and *late* if it occurred between trials 50 and 66; for this analysis, we excluded sessions with the reversal between trials 33 and 49[5]. Next, for each session, we selected the 30 trials around the reversal, the 15 preceding and the 15 succeeding trials, divided them into blocks of five consecutive trials, and computed the average proportion of an error in each block. Before the reversal, all S1 responses were correct and S2 responses were anticipatory errors, whereas after the reversal, all S1 responses were perseverative errors and S2 responses were correct. Differences in the proportion of errors – whether anticipatory or perseverative – between early and late reversals reveal temporal control.

To assess outcome control, we compared the proportion of S1 responses in each trial when it had occurred before the reversal versus when it occurred after the reversal. On every session, each trial was identified *pre-reversal* (trial number ≤ reversal trial) or *post-*

[5] In all cases, performance in these sessions fell in between of that of the sessions with early and late reversals, being similar to both and not significantly different from either one of them.

reversal (trial number > reversal trial). Consider trial 26, we divided the sessions into two groups: one with the sessions in which trial 26 was pre-reversal, and another one with the sessions in which trial 26 was post-reversal. Next, we computed the proportion of S1 on trial 26 separately for each group and obtained two proportions. Finally, we repeated the procedure for the other trials and obtained two series of S1 proportions as a function of the trial number, a pre-reversal series and a post-reversal series.

To ensure minimal reliability, we required at least eight sessions to compute a proportion. For this reason, the two series did not extend across all trials. To illustrate, consider a phase in which the reversals occurred on the odd numbered trials between 16 and 66. In all 25 sessions, trial 1 was pre-reversal; hence, the pre-reversal S1 proportion was estimated by averaging across all 25 sessions; obviously, no post-reversal proportion was computed for trial 1. In the same phase, trial 40 was pre-reversal on 13 sessions (those with the reversal on trials 41 to 65) and post-reversal on 12 sessions (reversal on trials 17 to 39). The pre- and post-reversal proportions were computed from these 13 and 12 sessions, respectively. In contrast, trial 60 was pre-reversal on only 3 sessions (reversal on trials 61, 63, or 65) and post- reversal on 22 sessions (reversal on trials 17 to 59). In this case, only the post-reversal proportion was computed. The net effect of the 8-session requirement was that the number of sessions used to compute the proportions varied with trial number, decreasing in the pre-reversal series and increasing in the post-reversal series. The pre-reversal series extended from trials 1 to 48 and the post-reversal series extended from trials 34 to 80.

The critical feature of this analysis is that performance is averaged over homogeneous contingencies, before the reversal or after the reversal. Moreover, because the pre- and post-reversal proportions on trials 34 to 48 are time-matched (i.e., each trial occurred roughly at the same time into the session), they may be used to isolate any outcome effect: A significant difference between the pre- and post-reversal proportions on the same trial reveals the effect of choice outcomes rather than the effect of time into the session.

For each condition, we analyzed the absolute and relative differences between the pre- and post-reversal proportions on trials 34 to 48. To compute the average absolute difference, we subtracted the post-reversal proportion from the pre-reversal proportion on each of these 15 trials and then averaged them. To compute the relative difference (a

control for potential differences across conditions in pre-reversal performance), we first computed the average absolute difference and then divided it by the average of the pre-reversal proportions. We used the individual average differences to determine whether outcome control varied across conditions.

Results

Temporal control: Early- versus late-reversal sessions

Figure 2 shows the average proportion of errors surrounding the reversal in each condition, contrasting performance in the early- versus late-reversal sessions. The top left panel shows the High-High condition, in which correct choices were always reinforced ($q_1 = q_2 = 1$). Anticipatory errors were substantially more common in *late*- than in *early*-reversal sessions (closed circles systematically above empty circles at the left of the vertical line), whereas perseverative errors showed the opposite pattern, higher in *late*- than *early*-reversal sessions (closed circles below empty circles at the right of the vertical line). The same results occurred during condition Int-Int (top right panel) with intermittent

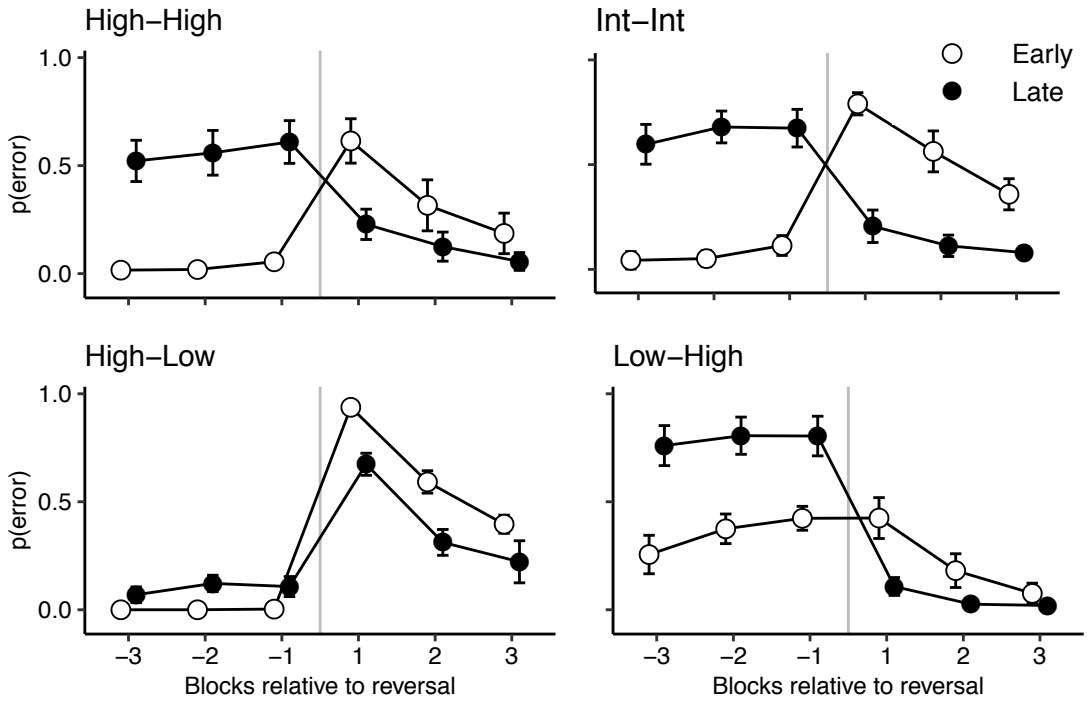

Figure 3.2. Proportion of errors in blocks of five trials relative to the reversal with the location of the reversal as a parameter. Each panel represents a different condition. Open circles represent the performance in the sessions where the reversal occurred early (between trials 16 and 32), closed circles represent the performance in the sessions where the reversal occurred late (between trials 50 and 66). Error bars represent the 95% CI.

but equal payoffs ($q_1 = q_2 = .5$). Together, they reproduce Santos et al.'s (2019) findings.

The bottom left panel shows performance in the High-Low condition ($q_1 = 1$ and $q_2 = .2$). Compared to the $q_1 = q_2$ conditions, anticipatory errors decreased in late-reversal sessions (cf. closed circles before reversal), but perseverative increased in both early- and late-reversal sessions (cf. closed and filled circles after reversal). In the first post-reversal block, for example, perseverative errors were close to 1 during the early-reversal sessions. These results, also consistent with Santos et al. (2019), show that the differential payoff biased performance towards S1, the richer stimulus. However, it was still the case that the relative position of the two curves did not change.

The bottom right panel shows the Low-High condition ($q_1 = .2$ and $q_2 = 1$). The payoff differential biased performance towards S2, increasing anticipatory errors and decreasing perseverative errors. But, as in all other conditions, the two curves retained their relative position.

To summarize, in conditions High-High and Int-Int the two types of errors had similar frequencies, although anticipatory errors predominated in late-reversal sessions while perseverative errors predominated in early-reversal sessions; the curves remained roughly symmetric around the vertical line. In conditions High-Low and Low-High, there were more errors in the part of the session with the lower payoff, perseverative errors in High-Low and anticipatory errors in Low-High; the curves lost their symmetry. These results are consistent with Santos et al. (2019).

A robust finding was that performance seemed consistently determined by the interaction between proximity to the reversal (the blocks of trials in Figure 2) and the moment in the session the reversal occurred (the two curves in each panel of Figure 2): In all experimental conditions, the two curves changed level after the reversal but in opposite directions; they crossed over, the signature of the interaction. The late-reversal sessions showed the highest proportion of anticipatory errors and the early-reversal sessions the highest proportion of perseverative errors. However, the magnitude of this interaction varied across conditions, a result confirmed by a 4 x 2 x 6 repeated-measures ANOVA with condition, location of reversal, and blocks of trials as factors: The three-way interaction was highly significant, $F(15, 336) = 5.95$, $p < .001$.

To further understand the results, we simplified the analysis by reducing the six blocks of trials to two —the average of the first three blocks corresponding to performance

before the reversal, and the average of the last three blocks corresponding to performance after the reversal— and restricted the analysis to a subset of first-order interactions, those with greater theoretical import. First, we considered separately early- and late-reversal sessions and performed two-way repeated measures ANOVAs to examine the effect of condition and block on errors. The ANOVA for the early-reversal sessions (see Figure 3.3, left panel) showed a significant interaction, $F(3, 56) = 78.35$, $p < .001$. Tukey HSD post-hoc tests confirmed that the proportion of errors was significantly lower before than after the reversal in all conditions except for Low-High, which showed a higher proportion of errors before than after the reversal (Low-High: $p = .03$; all other ps $< .001$). The ANOVA for the late-reversal sessions (see Figure 3.3, right panel) also showed significant interaction, $F(3, 56) = 106.27$, $p < .001$, and Tukey's HSD tests showed that the proportion of errors was significantly higher before than after the reversal for all conditions except for the High-Low, which yielded fewer errors before than after the reversal (all ps $< .001$).

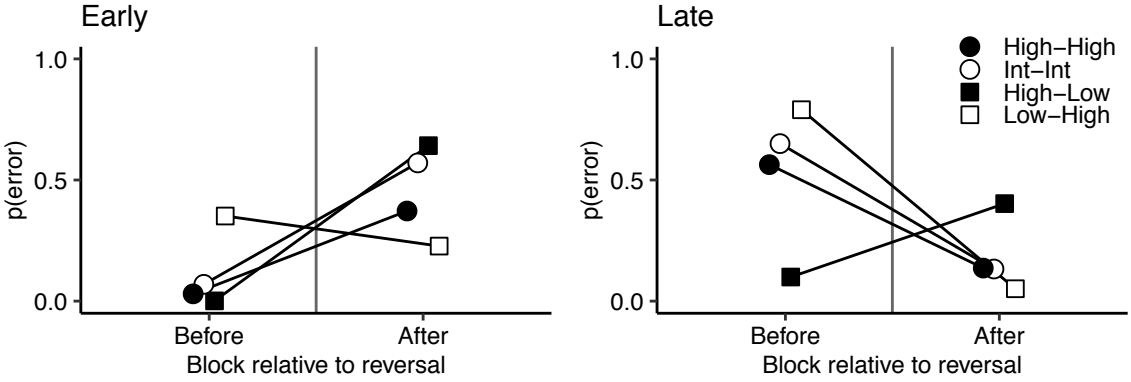

Figure 3.3. Proportion of S1 responses as a function of blocks of trials relative to the reversal with the condition as a parameter. Each panel represents different sessions according to the location of the reversal. The left panel shows performance in the sessions with an early reversal and the right panel the sessions with a late reversal.

Second, we considered separately the errors before and after the reversal and performed two-way repeated measures ANOVAs to examine the effect of condition and location of the reversal. The ANOVA for the errors before the reversal (see Figure 3.4, left panel) indicated a significant interaction between the condition and the location of the reversal, $F(3, 184) = 59.39$, $p < .001$. Tukey HSD post-hoc tests in each condition indicated that the errors before the reversal (anticipatory) were always lower in early- than in late-reversal sessions (High-Low: $p = .014$; all other ps $< .001$). For the errors after the

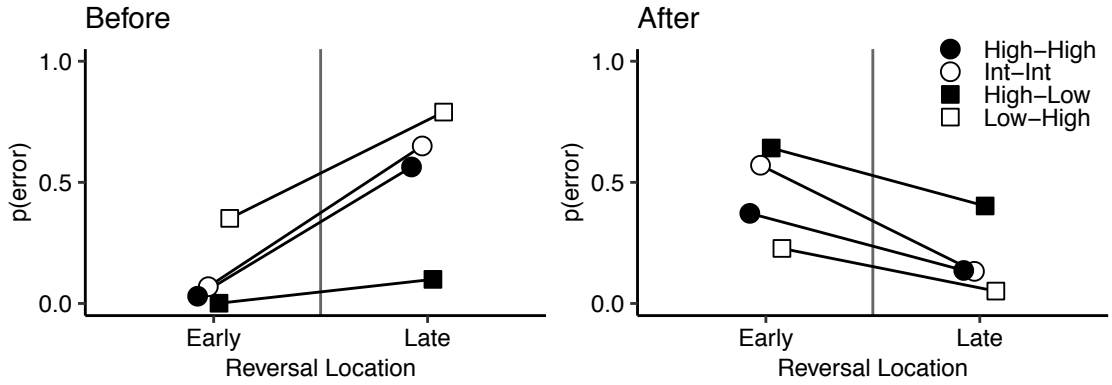

Figure 3.4. Proportion of S1 responses as a function of the reversal location with the condition as a parameter. Each panel represents a different block of trials relative to the reversal. The left panel shows performance in the block of trials before the reversal and the right panel in the block of trials after the reversal.

reversal (see Figure 3.4, right panel), the ANOVA also showed a significant interaction between the condition and the location of the reversal, $F(3, 184) = 4.79$, $p = .003$. Tukey HSD post-hoc tests indicated that perseverative errors were always higher in early- than in late-reversal sessions (Low-High: $p = .019$; all other ps $< .001$). In summary, the condition modulated the differences in errors before and after the reversal according the location (i.e., early vs. late). Compared to the conditions with equal payoffs, the High-Low condition attenuated the differences in anticipatory errors between the early versus the late reversal sessions, whereas the Low-High condition attenuated the differences in perseverative errors between the early versus the late reversal sessions.

Outcome Control: Pre- versus post-reversal trials

Figure 3.5 displays performance in each condition with the gray area highlighting the trials with both pre- and post-reversal performance measures. Seven of the eight curves decreased gradually with trial number. The exception was the pre-reversal curve in the High-Low condition, which remained steady at 1. These results are consistent with time-based control. Moreover, in all conditions, the proportion of S1 responses was always higher before than after the reversal. The clear gaps between the pre- and post-reversal curves on the common trials indicate an abrupt change in choice after the reversal. This result is inconsistent with time-based control but consistent with outcome-based control.

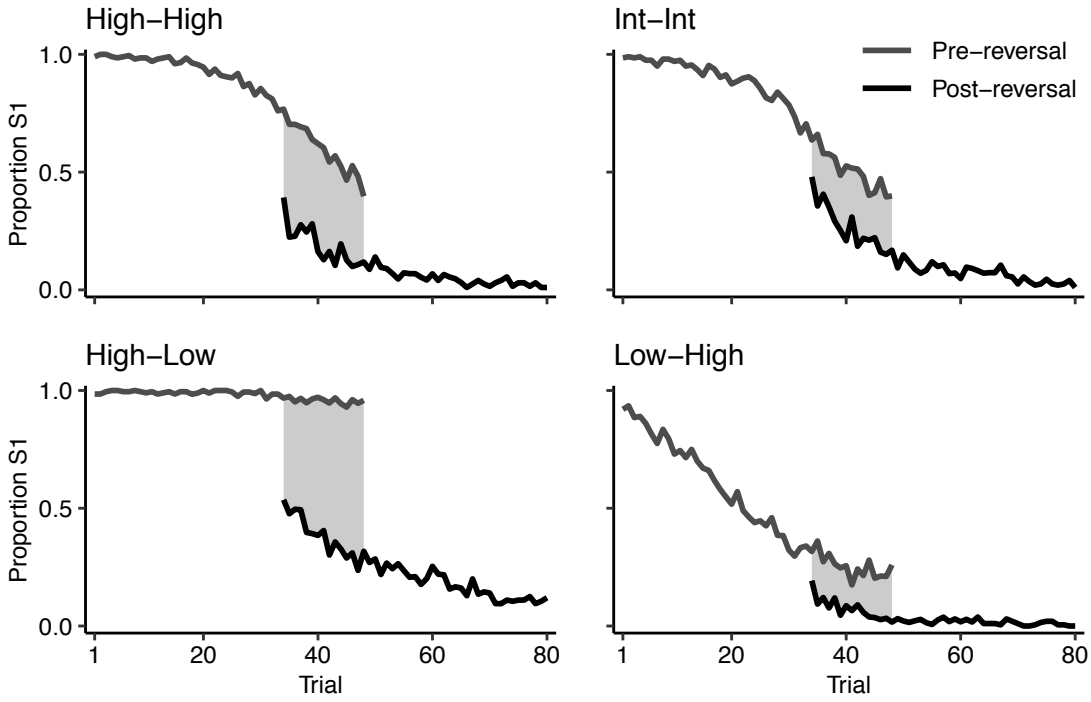

Figure 3.5. Proportion of S1 responses by trial with the occurrence of the reversal as a parameter. The grey lines represent performance pre-reversal and the black lines post-reversal. Each panel corresponds to a different condition. The gray shaded area highlights the gap between the pre- and post-reversal curves in the common trials.

The largest absolute difference in pre- versus post-reversal performance was observed in the High-Low condition (M = .58, SD = .07) followed by the High-High (M= .40, SD = .09), the Int Int (M = .24, SD = .12) and the Low-High (M = .18; SD = .10). A repeated measures one-way ANOVA on the size of these gaps (the average absolute difference between the overlapping trials of the pre- and post-reversal curves), confirmed that the there was a significant effect of the condition on the absolute size of the gap between pre- and post-reversal curves, $F(3, 28) = 24.61$, $p < .001$. Tukey HSD post hoc tests confirmed that all comparisons were statistically significant except for that between conditions Int-Int and Low-High (p = .617, all other ps < .05).

The post-reversal curves started always below the pre-reversal curves. Hence, the magnitude of the maximum gap depended on the absolute level of the pre-reversal curves. If the pre-reversal curve reaches a low value, as in the Low-High condition, the magnitude of the gap also is necessarily small. To control for the differences in pre-reversal performance level, we compared the relative average difference in each condition and found that the reduction in performance after the reversal remained similar (High-High: M

= .69, *SD* = .16; Int-Int: *M* = .47, *SD* = .19; High-Low: *M* = .60, *SD* = .07; Low-High: *M* = .65, *SD* = .20). A repeated measures one-way ANOVA on these relative average differences across conditions yielded no significant differences, $F(3, 28) = 2.68$, $p = .066$. The outcome effect revealed by the gap remained relatively constant across conditions (roughly a 60% decrease in the proportion of S1 responses).

Discussion

In this study, we aimed to analyze the sources of behavioral control in a situation where contingencies changed at a relatively unpredictable time, but the outcomes of previous behaviors could serve —to different extents— as cues for the active contingencies. Moreover, we were particularly interested in assessing if behavior was under temporal control in the MSR task when the S1 had a higher payoff than the S2. In this respect, subjects were exposed to a MSR task with different payoffs, mirroring the procedure described in Santos et al. (2019), but with the addition of the reversal being unpredictable on each session. Performance was different in early- versus late-reversal sessions and in pre- versus post-reversal trials, meaning that, in every condition we found evidence of joint behavioral control by time and by response outcomes.

Above all, we noticed that the overall number of errors before and after the reversal was consistent with our previous study (Santos et al., 2019): In the conditions with asymmetrical payoff (Low-High and High-Low) performance was biased towards the higher payoff alternative, whereas errors were equally distributed in the conditions with symmetrical payoff (High-High and Int-Int). Thus, the finding that unequal payoffs foster response biases in the MSR task seems to be a robust phenomenon not only observed when the reversal is fixed.

Temporal control

Regarding the effect of the location of the reversal on performance, a staple of this experiment was the prevailing pattern of fewer anticipatory and more perseverative errors in the early-reversal sessions, and the opposite trend in the late-reversal sessions (Figure 3.2; open circles always below the closed circles to the left of the vertical line, and open circles always above the closed circles to the right of the vertical line). These results add to

the evidence of time-regulated behavior in the MSR task, and are consistent with those of Rayburn-Reeves, Molet, and Zentall (2011) in pigeons and of Smith, Pattison, and Zentall (2016) in rats. Although both of these studies had only 5 possible reversal trials with continuous reinforcement throughout the session, the present experiment, with 25 possible reversal trials, not only replicated their results in the condition with continuous reinforcement (High-High) but in all others as well (Int-Int, Low-High, and High-Low), suggesting that in every condition behavior was to some extent temporally regulated. Specifically, the evidence of temporal control of behavior in the Int-Int condition is consonant with rats' performance in the MSR task with partial reinforcement and variable reversal (Santos & Sanabria, 2020). Moreover, although no other study has manipulated the payoffs and made the reversal unpredictable simultaneously, it is possible to draw some parallelisms with previous work. On one hand, the performance in the Low-High condition was congruent with the timing evidence presented by Santos et al. (2019): If subjects were to estimate the average time of the reversal and happened to be biased towards S2, performance should show a higher proportion of anticipatory than perseverative errors and differences in the proportion if errors according to the location of the reversal. On the other hand, the remarkably lower anticipatory than perseverative errors (i.e., biased responding towards the richer alternative) in the High-Low condition, replicated the findings of Santos et al. (2019) and Zentall et al. (2019) in the conditions where S1 was richer than S2 but the reversal was fixed. Interestingly, both of these studies agreed that the sharp transitions in choice from S1 to S2 after the reversal evidenced the use of a WSLS rule, and that the elapsed time from the beginning of the session was not the main cue controlling behavior. Thus, encountering evidence of some temporal control in the High-Low condition in the present experiment was unforeseen. Nevertheless, afar from being inconsistent with the aforementioned studies, the difference in performance between the early- and late-reversal sessions might had revealed timing processes that were unobservable when the reversal was fixed (Santos & Sanabria, 2020).

Although the difference in anticipatory errors in the High-Low condition was negligible, it is unlikely that such robust pattern of incurring in more perseverative errors in the sessions with an early reversal than in those with a late reversal was related to resistance to extinction. It is well established that resistance to extinction, as measured by the number of responses to extinction, tends to increase with the number of reinforcers

obtained (Williams, 1938; Zarcone, Branch, Hughes, and Pennypacker, 1997). Thus, if these results were a matter of resistance to extinction, we would expect to see more perseverative errors in the sessions with late reversals than in those with early reversals, because extinction follows more reinforced trials, opposite to what we found in this experiment (in Figure 3.2 the closed circles always below the open circles after the reversal). Instead, this puzzling and surprising result could be an effect of differences in satiation. Few studies have documented the effect but have consistently reported that the number of responses to extinction is directly related to drive level (Sackett, 1939; Crocetti, 1962). Thus, it is possible that when the reversal occurred earlier in the session, the subjects had received relatively few reinforcers, were hungrier, and more resistant to extinction; conversely, when the reversal occurred in later in the session, they had received a considerable number of reinforcers, were less hungry, and less resistant to extinction.

Outcome control

Behavioral control by the outcome of the preceding trial was quantified by the difference between the pre- and post-reversal curves in each condition of the experiment (Figure 3.5). The presence of clear gaps between the two curves suggested strong sensitivity to the change in contingencies. The comparison of the absolute differences between the curves across conditions suggested that the degree of control by the outcomes in the condition with continuous reinforcement for both alternatives (High-High: q_1 and q_2 = 1) was higher than in the conditions with a lower payoff for S1 (Int-Int and Low-High: $q_1 < 1$); which is reasonably expected because, in the conditions Int-Int and Low-High, non-reinforced responses to S1 were ambiguous: they did not necessarily indicate that the reversal had taken place, as they did in the High-High condition where S1 was continuously reinforced. In this context, the most puzzling comparison continues to be between the High-Low and the High-High conditions. It is not clear why there is differential control by the outcomes when, in both conditions, non-reinforced S1 responses were equally reliable or 'informative' of the active contingencies.

However, the analysis of the relative differences in performance showed that, in all conditions, the post-reversal proportion of S1 responses was of about 60% of the pre-reversal, indicating that the relative effect of the outcomes was constant regardless of the alternatives' payoff. Thus, the question to be answered is not why is outcome control

different across condition, but why are there differences in pre-reversal performance across conditions?

To this conundrum at least three explanations have been proposed. The first one alludes to attentional processes, asseverating that in discrimination tasks subjects tend to pay attention to the outcomes of an S+. In the MSR task, when the overall value of both stimuli are similar, the anticipation of the reversal shifts attention to S2 (the anticipated S+) competing with S1 (the current S+), and resulting in errors; but when S1 has a higher payoff than S2, it is preferred and attention is diverted to S1, improving accuracy (Zentall, et al., 2019a).

The second explanation refers to inhibitory processes, proposing that errors are nothing more than a failure to inhibit responses to the S- (McMillan, Sturdy, & Spetch, 2015). The authors have not extended their explanation to the case of the MSR task with different payoff alternatives, but we could try to extend their premise: When both alternatives have a similar payoff subjects have difficulty inhibiting their responses to the S- around the reversal, incurring in both anticipatory and perseverative errors; but when S1 has a higher payoff, S2 is weakly associated with reinforcement making it easier to inhibit responding and resulting in no anticipatory errors.

Both accounts are based on the idea that animals anticipate a contingency reversal, introducing a tacit non-timing component in their explanation. The timing component remains to be elaborated. It is yet to see how these verbal accounts will apply their explanatory principles to account for the systematic finding that the relative effect of the outcomes is independent from the alternatives' payoff, and consistently produces a relative change in performance of about 60%. The rhetoric nature of these accounts limits the scope of their predictions to perhaps ordinal appraisals, but definitely do not allow for any quantitative prediction of this sort.

The third explanation relies on temporal control of behavior. According to the LeT model when S1 and S2 have different payoffs in the MSR task, both the times of reinforcement and the reinforcement rate at those times will foster unequal strengthening of the associative links to each response unit, biasing preference towards the richer alternative. Thus, producing many anticipatory errors when the S2 is richer and very few (or even none) when the S2 is leaner. Despite this account makes accurate quantitative predictions of behavior in some variations of the MSR task, it cannot account for the

performance after the reversal in the High-Low condition (Santos et al., 2019; Zentall et al., 2019) and does not predict any gap between pre- and post-reversal performance as observed in this experiment. Even though, it is reasonable that a pure timing model does not account for non-temporal effects, it begs the question of how compatible it is with other non-temporal explanations of behavioral control.

All three accounts fall short accommodating the variety of phenomena observed in this experiment. Yet, they all seem to agree that when the S1 is richer than the S2 in the MSR task, behavior is no longer under temporal control and it is determined by the outcome of the previous trial. However, there are two key features of performance in High-Low condition in this experiment that must be considered: (1) the fact that the proportion of responses to S1 did not drop 100% (i.e., from 1 to 0) after the first non-reinforced S1 response, but only 60% (see Figure 3.3; the proportion of S1 changed from 1, in the pre-reversal trials, to about .4 in the post-reversal trials) indicates that behavior did not follow a *strict* WSLS strategy, and (2) the slope of the post-reversal curve indicates that the probability of responding to S1 decreased as trials progressed, consistent with temporal control. There are at least two possible interpretations.

On one hand, in the High-Low condition behavior was exclusively under the control of the past outcomes, but extinction of S1 responses after the reversal progressed slowly, completely shifting to S2 only after more than one non-reinforced trials. This is consistent with the use of a more flexible WSLS strategy in which responses do not reverse after the first non-reinforced response, but after a few non-reinforced trials. In other words, this performance is consistent with a subject that is exclusively responding to S1 until the reversal and shifts responding to S2 after a few non-reinforced S1 responses (as if a non-reinforced number of trials would serve as a threshold to shift). Thus, the negative slope in the post-reversal curve could be an artifact produced by the accumulated sum of perseverative errors due to the non-reinforced number of trials required to shift: Meaning that, in the High-Low condition, behavior was exclusively under control of the outcomes of the previous trials. However, if this were the case, there should not be differences in perseverative errors according to the location of the reversal (see Figure 3.2, bottom left panel, open and closed circles to the right of the vertical line are significantly different from each other) as previously shown by Santos & Sanabria (2019) in their simulation of performance in a MSR task using a flexible WSLS strategy.

On the other hand, in the High-Low condition behavior was under joint control of the outcomes and the time elapsed from the beginning of the session, as in the other conditions of the experiment. This would explain not just the slope of the post-reversal curve but also the difference in perseverative errors according to the moment of the reversal. This one being the more parsimonious interpretation of the two, because the difference between the High-Low condition and all others would not be the absolute absence of temporal control but the relative influence of the two sources of control: time and outcomes.

In the next section, we aim to capture the dynamics of behavioral control by time and outcomes altogether. We will contrast behavioral data with the predictions of simple computational models describing timing and WSLS rules likely to be implicated in the decision-making process in the MSR task.

Models of behavioral control by time and outcomes

The previous analyses suggested that pigeons' behavior was under joint control of time and outcomes in every condition, although it is not clear if in the High-Low condition subjects were actually timing the moment of the reversal or simply using a more flexible rule to guide their behavior based on the outcomes of more than one trial. Our first approach to tease apart these two possibilities is to see if a simple model that combines both sources of control and accounted reasonably for similar data of the MSR task (Santos & Sanabria, 2020), can also explain the results of the present experiment.

As a reference, Figure 3.6 shows the predicted performance when $q_1 = q_2 = 1$, the reversal is variable, and behavior is under control either of time or of the outcomes. The dashed lines represent the prediction of a pure timing model, and the solid lines the prediction of a flexible WSLS rule in which the subject would only shift responding after a few (i.e., tree) non-reinforced responses. The left panel shows the expected proportion of errors in blocks of trials around the reversal (c.f. Figure 3.2) and the right panel the expected proportion of S1 responses in the pre and post-reversal trials (c.f. Figure 3.5).

The main difference between the predictions of a pure timing model and the flexible WSLS rule, regarding the probability of making an error around the reversal, is the effect of the location of the reversal (i.e., early vs. late). The left panel of

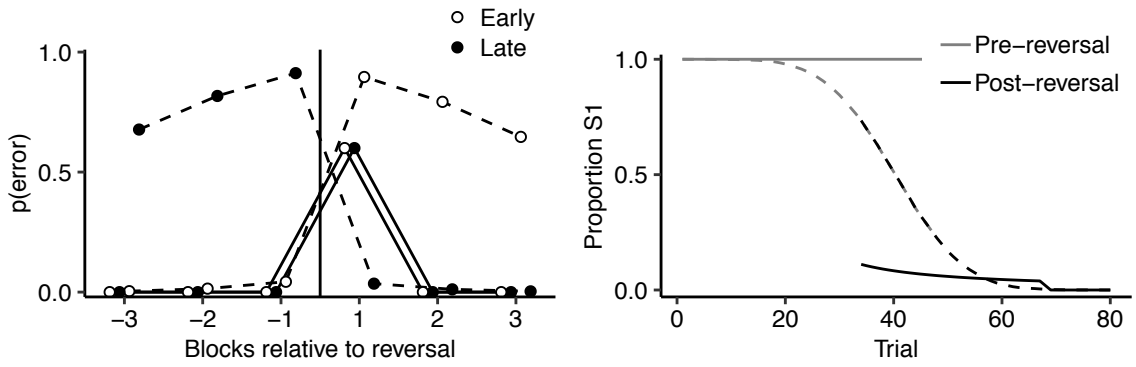

Figure 3.6. Predicted performance by a pure timing model (dashed lines) and a flexible WSLS rule (solid lines) in a MSR task with $q_1 = q_2 = 1$ and a variable reversal. The left panel shows the proportion of errors in blocks of five trials relative to the reversal in the sessions with an early (open symbols) and a late (closed symbols) reversal. The right panel shows the proportion of S1 responses in the pre-reversal (grey lines) and post-reversal (black lines) trials.

Figure 3.6 shows that the timing model predicts no anticipatory errors and large number of perseverative errors for the early reversal sessions and the opposite for the late reversal session; whereas the flexible WSLS rule indistinctly predicts no anticipatory errors followed by only three perseverative errors right after the reversal. Hence, the probability of responding to S1 in block 1 is of .6 because responding on S1 continues for 3 of the first 5 trials after the reversal.

Similarly, the right panel of Figure 3.6 shows that the pre- and post-reversal timing curves overlap on the common trials (i.e. 34 to 48), whereas the WSLS curves do not. In fact, the WSLS rule predicts a gap between the two curves with no anticipatory errors and some perseverative errors. This is, a steady (flat) pre-reversal curve (if $q_1 = q_2 = 1$, p(S1) = 1; if $q < 1$, p(S1) < 1), and a post-reversal curve with a negative slope reflecting the progressive reduction in perseverative errors when averaging across sessions.

Mixture Model I: The main assumption here is that performance alternates between timing and non-timing modes, as originally proposed by Sanabria and Killeen (2008). The model describes a subject that, on every trial, enters either a timing mode with probability p, or a non-timing mode with probability $1 - p$, where p is directly related to the momentary probability of reinforcement.

The timing mode follows the rules of a generic (pacemaker-accumulator) timing model: On every trial (n), the probability of responding to S2 is sampled from a cumulative normal distribution with a mean (μ) equal to the average of the reversal

locations and standard deviation (σ) proportional to the mean. This is, $P_{S2}(n) = \phi\ (n; \mu, \sigma)$ and, correspondingly: $P_{S1}(n) = 1 - P_{S2}(n)$. For simplicity, on trial 1 the subject will always respond to S1 by default. There are no further assumptions about the nature of this process as timing or counting as they can both be understood and explained in similar terms (Davison & Cowie, 2019). The non-timing mode determines that behavior follows a win-stay/lose-sometimes-shift (WSLSS) rule: a response will be repeated until it has been non-reinforced more than L trials in a row, in which case responding will switch to the other alternative. Here, $L = k[(1 - q) / q] + 1$; where k is a free parameter and q corresponds to the probability of reinforcement of that alternative; L is proportional to the odds against reinforcement (Santos & Sanabria, 2020). Notice that the threshold for non-reinforcement on S2 (L_2) would only be active after receiving the first reinforcer on S2, until that moment, the threshold for S1 (L_1) would be active.

We simulated performance of each bird in the present experiment with 100 replications of this model using the parameters estimated by maximum likelihood shown in Table 3.2. The coefficient of variation between the parameters of the timing component were fixed so that $\sigma = \mu * .25$; and according to the average of the programmed reversals, $\mu = 41$. For all birds, the best fitting value of parameter p was always higher than 0, meaning that in all conditions, including the High-Low, the WSLSS component of the model alone was not the best mechanism to explain the subjects' behavior. The alternation between the timing and non-timing components offered the best description of the data. Figure 3.7 shows the average of the outputs of the simulation, treated and analyzed in the same way as the birds' data. The left panels illustrate the predicted (lines) and observed (symbols) proportion of errors in blocks of five trials around the reversal (cf. Figure 3.2), and the right panels of Figure 3.7 show the predicted (lines) and observed (dots) proportion of S1

Table 2
Estimated parameters of the mixture model I

	Mixture model I				
	k	*p*			
Subject		HH	II	HL	LH
P816	2.04	0.90	0.90	0.14	0.90
P389	2.00	0.90	0.90	0.10	0.90
PG16	2.56	0.96	0.96	0.04	0.96
PG18	2.00	0.90	0.90	0.10	0.90
P053	2.20	0.92	0.92	0.08	0.71
P435	2.12	1.00	0.82	0.13	0.78
P795	2.19	0.73	0.91	0.08	0.91
P762	2.00	0.89	0.89	0.09	0.90

Note. Parameter *p* was allowed to vary between conditions and k could only take one value for all conditions.

responses on the pre- and post-reversal trials (cf. Figure 3.5).

The two top-left panels of Figure 3.7 show the similarity between the data and the predictions of the model in the conditions High-High and Int-Int: fewer anticipatory errors in the sessions with an early reversal compared to the sessions with a late reversal and the opposite pattern for the perseverative errors. Consistent with Santos and Sanabria (2019), the model replicated the observed proportion of errors around the reversal in the $q_1 = q_2$ conditions regarding the general trend and order of the data. The two top-right panels also post-reversal curves in the Int-Int condition. Because the difference between the curves indicates outcome control, one could think that increasing the probability of entering the non-timing component would easily solve this issue. However, we explored the effect of reducing the value of parameter *p* and the gap between the curves increased but in detriment of capturing the shape of the pre- and post-reversal curves and of the errors around the reversal. Thus, impairing the overall fit to the data.

The two bottom-left panels of Figure 3.7 show a clear difference between observed and predicted proportion of errors around the reversal in the High-Low and Low-High. Because the timing component of this model is a generic pacemaker-accumulator, it was unable to capture the effect of differential reinforcement. In the Low-High condition, the model showed great difficulty replicating performance in the sessions with an early reversal where —because of the bias towards S2 induced by its higher payoff— the subjects showed more anticipatory errors and fewer perseverative errors than predicted. This shortcoming of the model was evident even in High-Low condition where, although parameter *p* was very low and performance was mainly determined by the WSLSS component, the failure of the timing component to capture the bias towards S1 did not aid replicating the difference in perseverative errors according to the location of the reversal observed in the birds data (Figure 3.7, left High-Low panel: open circles on the right side of the vertical line are consistently above the closed circles).

Similarly, the outstanding difficulty to replicate performance the Low High condition (simulated with a high a *p*), confirms the deficiency of the timing component of the model to capture the gap between the curves and bias performance shifting the curves towards the richer alternative (bottom right panel of Figure 3.7). This limitation of the model was not so evident in the High-Low condition as behavior was mainly under

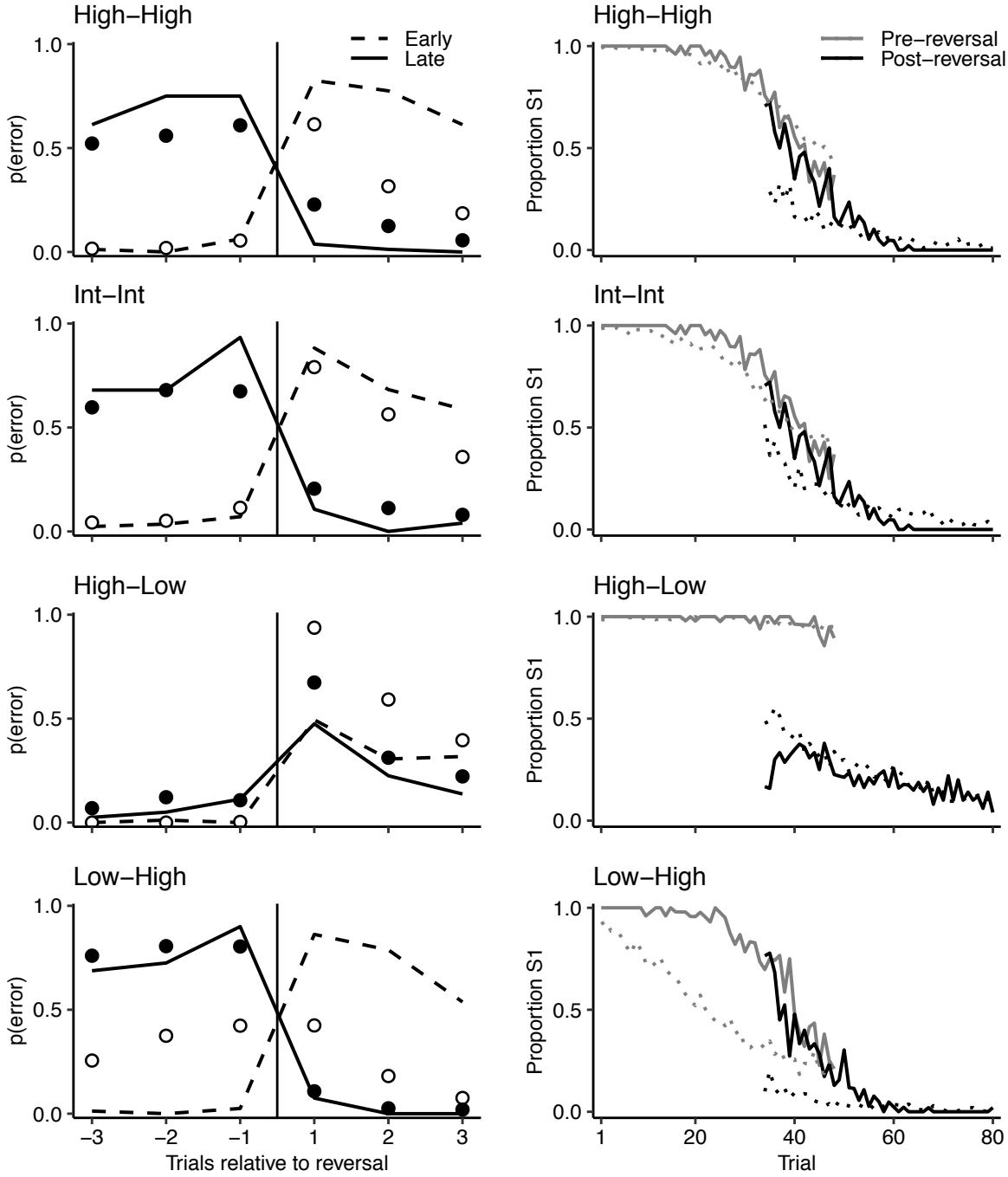

Figure 3.7. Mixture model I. Left panels show the predicted (lines) versus the observed (symbols) proportion of errors in blocks of five trials relative to the reversal with the location of the reversal as a parameter. Dashed lines and open circles represent the performance in the sessions where the reversal occurred early (between trials 16 and 32), solid lines and closed circles represent the performance in the sessions where the reversal occurred late (between trials 50 and 66). Right panels contrast the predicted (solid lines) versus the observed (dots) proportion of S1 responses in the pre-reversal (gray) and post-reversal (black) trials in each condition.

outcome-control and the system rarely entered the timing mode (simulated with a low *p*). Thus, our next step is to modify this model by replacing the generic timing component with one that adopts an associative decision rule that allows capturing the time-perception biases product of the differential payoffs of S1 and S2.

Mixture Model II: To capture the biasing effects of differential payoff on time perception in the MSR task, we modified the previous model by replacing the timing component with the Learning-to-Time model (LeT; Machado, 1997; Machado, Malheiro, & Earlhagen, 2009): a set of sequentially activated *behavioral states*, a *vector of associative links* that change in real time according to a learning rule, and a *response rule* that determines what response will be emitted. The compound system of LeT and the WSLSS rule works as follows: At the onset of each session, a random sample from a normal distribution with mean μ and standard deviation σ ($\sigma = \mu * s$) sets the speed of transition across states. On every trial the subject will enter either a timing mode (with probability *p*) or a non-timing mode (with probability $1 - p$). If the subject enters the timing mode, the LeT system will indicate a response according to the relative strengths of the links between the current state and each instrumental response available: S1 with probability $W_{(n, S1)}/[W_{(n, S1)}+W_{(n, S2)}]$, and S2 with the complementary probability. If the subject enters the non-timing mode, the WSLSS system will determine the response according to the outcome of previous trials (repeat the previous response if it has been non-reinforced less than *L* trials in a row). Regardless of the mode that determined the response, if the choice is rewarded, the link of state *n* with the reinforced response increases and the link of state *n* with the other response decreases; the magnitude of the changes depends on the reinforcement parameter β. If the choice is not rewarded the link of state n with the non-reinforced response decreases and the link of state n with the other response increases; the magnitude of the changes depends on the extinction parameter α (for further details see Machado et al., 2009).

We simulated the performance of each individual subject in the present experiment with 100 replications of this model using the parameters estimated by maximum likelihood that best fitted the data. For the timing component we used the architecture and initial weight of the associate links ($W_0 = .5$) described in Santos et al. (2019), and the coefficient of variation between parameters *s* and μ was constrained to .25 ($s = \mu * .25$). Table 3.3

shows the values of the estimated parameters for each subject and, consistent with the estimations of the Mixture Model I, in all conditions the values of parameter *p* were higher than 0 and smaller than 1, confirming that neither one of the two components of the models sufficient to describe the data, and the alternation between the two components of the model offered the best prediction of the observed performance.

Table 3
Maximum likelihood estimated parameters of the mixture model II

	Mixture model II							
	μ	α	β	k	p			
Subject					HH	II	HL	LH
P816	0.03	0.03	0.95	2.26	0.84	0.84	0.09	0.91
P389	0.02	0.03	0.95	2.00	0.90	0.90	0.10	0.90
PG16	0.02	0.03	0.95	2.00	0.90	0.90	0.10	0.90
PG18	0.02	0.03	0.95	2.00	0.90	0.90	0.10	0.90
P053	0.03	0.03	0.87	2.33	0.90	0.82	0.09	0.82
P435	0.02	0.03	0.95	2.01	0.90	0.90	0.10	0.90
P795	0.03	0.05	0.93	1.96	0.90	0.90	0.10	0.89
P762	0.03	0.05	0.95	2.07	0.88	0.87	0.12	0.88

Note. Parameter *p* was allowed to vary freely between conditions, all other parameters remained constant.

Figure 3.8 shows the average of the model's outputs compared to the subjects' average data. The left panels illustrate the proportion of errors around the reversal in blocks of five trials (cf. Figure 3.2), and the right panels, the proportion of S1 responses in the pre- and post-reversal trials (cf. Figure 3.5). Overall, the model was able to capture the main features of performance in every condition and, although it did not reproduce the same levels of performance showed by the subjects, the general trend was very similar.

In the $q_1 = q_2$ conditions the fit to the data was good both in terms of the proportion of errors (two top left panels of Figure 3.8) and in the proportion of S1 responses in the pre- and post-reversal trials (two top-right panels of Figure 3.8). Moreover, compared to the previous model, it offered a much better description of the data as it closely reproduced the size of the gaps between the pre- and post-reversal curves. The reason behind this difference relies on the interaction between the WSLSS component and speed of the transitions between the behavioral states in the LeT model. In this simulation, the timing component is always active in the background, updating the weight of the associative links

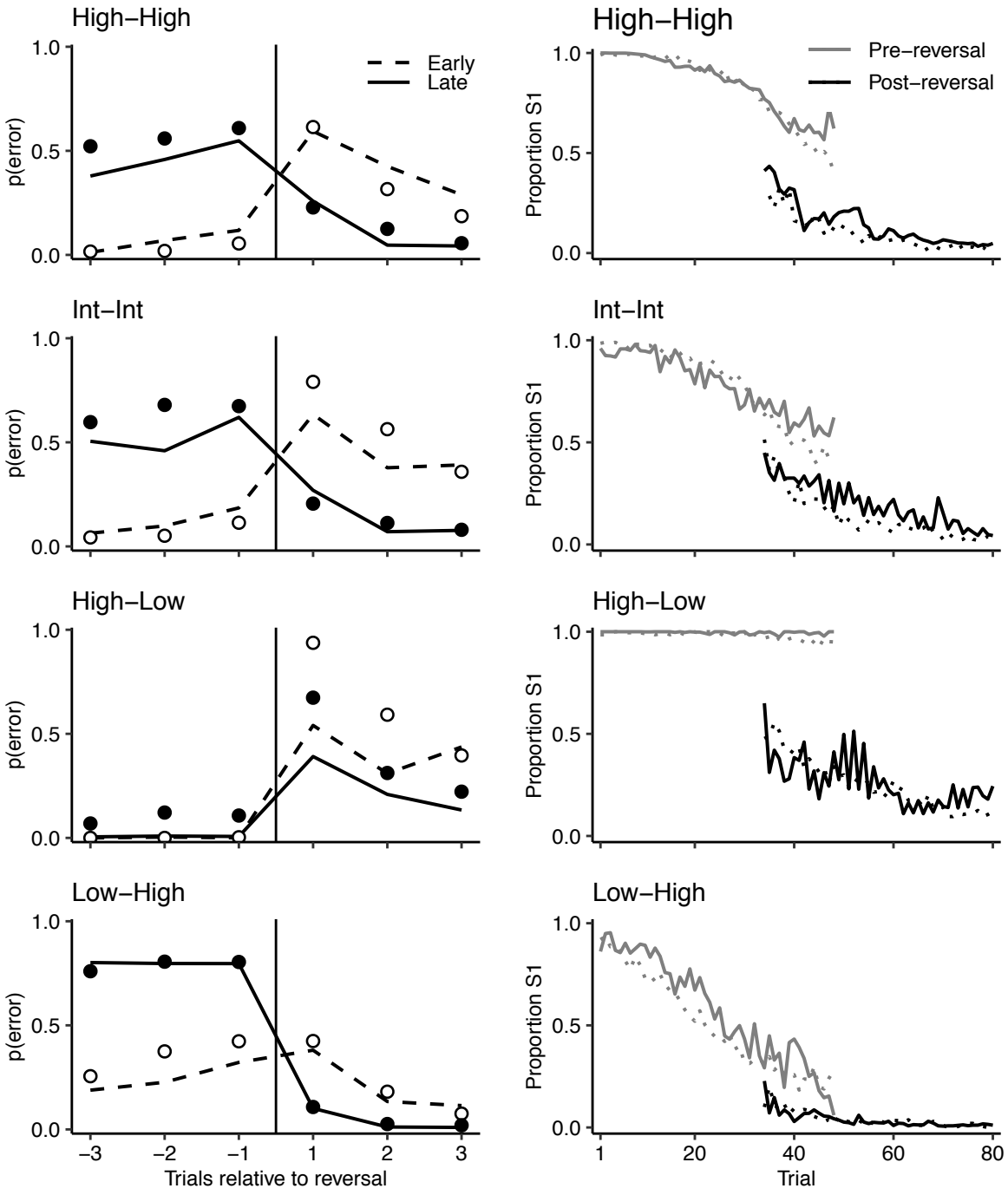

Figure 3.8. Mixture model II. The left panels show the predicted (lines) versus the observed (symbols) proportion of errors in blocks of five trials relative to the reversal with the location of the reversal as a parameter. The dashed lines and open circles represent the performance in the sessions where the reversal occurred early (between trials 16 and 32); the solid lines and closed circles represent the performance in the sessions where the reversal occurred late (between trials 50 and 66). The right panels contrast the predicted (solid lines) versus the observed (dots) proportion of S1 responses in the pre-reversal (gray) and post-reversal (black) trials in each condition.

regardless of the active mode (timing or non-timing), and only changing decision rules between trials according to *p.* Thus, when the speed of the transition between states allows for a residence time of a few trials (in this case, each state was active for about five trials on average), it allows the model to update the strength of the associative link to a state and adjust performance to the active contingencies within a single session. Increasing this effect when timing-mode trials are intermixed with non-timing-mode (WSLSS) trials. The Mixture Model I did not allow for interaction and the effect of each model always remained independent.

This version of the model, did not only reproduce the temporal biases produced by the asymmetric payoffs observed in conditions High-Low and Low-High (two bottom-left panels of Figure 3.8), but also accounted for the differences in performance produced by the outcomes (two bottom-right panels of Figure 3.8). The dynamic strengthening of the associative links as a consequence of both, reinforcement and non-reinforcement, accounts for the temporally regulated responses observed in this experiment, and the WSLSS strategy aided the explanation of the drastic shifts in performance when the S1 had a higher payoff than S2.

Summary and final comments

The main goal of this study was to assess the sources of behavioral control in the MSR task when, compared to the traditional task, the time elapsed from the beginning of the session was no longer an indicator of the active contingencies, and the reliability of the outcome of previous responses was variable. We were particularly interested in whether or not performance was under temporal control when the S1 had a higher payoff than the S2. The data analysis suggested that behavior was always under joint control of time and the past outcomes, although these two cues shared behavioral control to different degrees depending on the differential payoff of the alternatives. Moreover, when S1 had a higher payoff that S2, behavior did show some evidence of temporal control but to a much smaller degree than under any other condition.

As a first attempt to understand how do these two cues combine to determine behavior, we contrasted the experimental data with the predictions of two different mixture models that combined a heuristic strategy with a timing model. The heuristic was a modified WSLS rule (the win-stay/lose-*sometimes*-shift; WSLSS), and its role was to aid

learning form one trial to the next, allowing behavior to adapt to the change in contingencies within a session. In the first mixture model, the timing component was analogous to a generic pacemaker-accumulator model. Although this type of model has previously been sufficient to explain performance in the MSR task at a steady state when the alternatives have equal payoff (Santos & Sanabria, 2020), in this case it did not describe performance when the response alternatives had different payoffs. Hence, in the second mixture model, to capture the biasing effects of different reinforcement rates, the timing component was replaced with a variation of the LeT model. Overall, the mixture model II offered a good description of the birds' observed performance, including the biases produced by asymmetrical payoffs.

Another way of interpreting the mechanism by which the mixture model II was able to describe the birds performance, is by assuming that, rather than switching between timing and non-timing modes, what oscillates is the active decision rule. In other words, the model represents a subject that is always timing and updating its timing system according to the experienced events in real time, but behavior is not always determined by what the timing system dictates, behavior determination switches between global (timing) and local (response outcomes) decision rules depending on biological and environmental constraints.

Mixture models are a versatile way of describing behavior in the MSR task because they capture the idea that subjects could be accurately timing but not show stimulus control by time. One could argue that because this flexible WSLS rule requires counting the number or non-reinforced trials or estimating the average inter-reinforcer interval, it could also be considered a timing process. Here, we have distinguished these two processes by operationalizing them differently: the timing process emulates the estimation of the interval from the beginning of the session to the moment of the reversal, and the heuristic rule emulates tracking the interval without reinforcers. However, a modified version of LeT that is able to keep track of two different intervals —one from the beginning of the session to the reversal and another one between reinforcers— might be able to explain performance and learn to adapt behavior to the change in contingencies within a single session, implementing the heuristic decision rule in an associative form. The architecture of the full model and its dynamic rules remain to be elucidated. Only

further research will continue to piece apart the elements determining behavior when both, time and response outcomes, simultaneously hint on reinforcement availability.

Conclusion

In the present dissertation we assessed the role of the time elapsed since a particular event and the outcome of the previous response as discriminative stimuli for choice. We systematically manipulated the reliability of each of these cues, assessed the effect on choice, and compared the observed performance with the predictions of different simple mathematical models of behavior describing, either temporal control, outcome control, or different combinations of the two. The goal was to explore the necessity of these mechanisms to accurately explain behavior, or the insufficiency of either one of them to account for the results of the present experiments, and shed light on how time and outcomes combine to determine performance in regularly changing environments.

Study I was partially consistent with the idea that performance in the MSR task can be biased in the same way as in other timing tasks. When the reversal trial was fixed, but the reinforcement rate changed, performance continued to be mainly under temporal control behavior, as in the traditional version of the task. Behavior was susceptible of temporal bias induced by differential payoff only when the S2 was richer than the S1. In the case where we tried to bias temporal performance towards the S1 by making it the richer cue, we observed a pattern of performance resembling the use of a WSLS rule, strikingly inconsistent with a time-based account. It seems that temporally regulated behavior can shift to alternative (local) sources of control under particular reinforcement contingencies; although the dynamics of this transition is not yet understood and might constitute the biggest unanswered question of this research.

Consistently, Study II also suggested that behavioral control seems to shift from one cue to another in the MSR according to their relative reliability as indicators of potential reinforcement. Specifically, we found that rats' behavior was mainly controlled by the outcome of the previous responses but was also consistent with temporal regulation. However, neither a pure timing model nor a flexible WSLS rule were enough to explain rats' MSR performance. A mixture model describing a combination of both types of control offered a good account of the data and suggested that subjects might estimate the interval from the beginning of the session to the contingency reversal, the interval since a

particular response lead to a reinforcer, and selectively shift behavioral control from one cue to another. Thus, the dynamic shift in the source of behavioral control according to environmental constraints is not restricted to pigeons' performance but was also observed in rats; hinting on the generality of the phenomena.

Study III showed that when the time to the contingency reversal is variable, time and outcomes share joint control, and that the degree of control by each of the cues seemed related to the momentary probabilities of reinforcement. In this experiment, pigeons' performance was sensitive to the relatively unpredictable change in contingencies within a single session, especially when S1 had a higher payoff than S2. This observation led us to test the generality of the mixture model assessed in Study II, exposing the inability of its non-associative timing component to capture the interactions between the time of reinforcement and the reinforcement rate at those times. Hence, a similar mixture model with a timing component based on the LeT model debuted as a first quantitative step towards unraveling, or at least hypothesizing, about the underlying processes that simultaneously control choice. Yet, the main limitation of this account is that it remains silent about how exactly does the momentary probability of reinforcement determine the relative control of each cue over behavior.

One way in which animals may adapt to regularly changing environments is by regularly changing behavioral control from one cue to another. For such adaptation, animals might also be able to, continuously and simultaneously, learn from more than one contingency relation. Thus, just one more shout for integrated theories of timing and associative learning.

Contrary to the generalized belief that the MSR task is nothing more than another timing task, for pigeons, and a regular simple simultaneous discrimination task, for rats and humans; this dissertation is an example that the MSR task is a suitable preparation to study one of the most puzzling and unresolved issues in behavioral science: how do associative learning and timing mechanisms integrate?

References

Bitterman, M. E. (1965). Phyletic differences in learning. *American Psychologist, 20*(6)*,* 396–410. http://dx.doi.org/10.1037/h0022328

Bizo, L. A., & White, K. G. (1994). The behavioral theory of timing: Reinforcer rate determines pacemaker rate. Journal of the Experimental Analysis of Behavior, 61, 19-33.

Bizo, L. A., & White, K. G. (1995). Biasing the pacemaker in the behavioral theory of timing. *Journal of the Experimental Analysis of Behavior, 64*, 225-235.

Bizo, L. A., Chu, J. Y. M., Sanabria, F., & Killeen, P. R. (2006). The failure of Weber's law in time perception and production. Behavioural Processes, 71, 201-210.

Cambraia, R., Vasconcelos, M., Jozefowiez, J., & Machado, A. (2019). Biasing performance through differential payoff in a temporal bisection task. *Journal of Experimental Psychology: Animal Learning and Cognition, 45*, 75-94. doi: 10.1037/xan0000192

Carvalho, M., Machado, A., & Vasconcelos, M. (2016). Animal timing: a synthetic approach. *Animal Cognition, 19*(4), 707–732. doi: 10.1007/s10071-016-0977-2

Cook, R. G., & Rosen, H. A. (2010). Temporal control of internal states in pigeons. *Psychonomic Bulletin & Review, 17*, 915-922. doi:10.3758/PBR.17.6.915

Cowie, S., Bizo, L. A., & White, G. (2016). Reinforcer distributions affect timing in the free-operant psychophysical choice procedure. *Learning and Motivation, 53,* 24-35. doi:10.1016/j.lmot.2015.10.003

Cowie, S., Davison, M., Blumhardt, L., & Elliffe, D. (2016). Does overall reinforcement rate affect discrimination of time-based contingencies? Journal of the Experimental Analysis of Behavior, 105, 393-408. doi: 10.1002/jeab.204

Crocetti, C. P. (1962). Drive level and response strength in the bar-pressing apparatus. *Psychological Reports, 10*, 563–575. doi: 10.2466/pr0.1962.10.2.563

Davison, M., & Cowie, S. (2019). Timing or counting? Control by contingency reversals at fixed times or numbers of responses. *Journal of Experimental*

Psychology:Animal Learning and Cognition, 45, 222-241. doi:10/1037/xan0000201

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review, 84,* 279-325.

Grondin, S. (2014). About the (non)scalar property for time perception. In: H. Merchant & V. de Lafuente (Eds.), *Neurobiology of Interval Timing.* (pp.17-32). New York, NY: Springer. doi:10.1007/978-1-49391782-2.

Guilhardi, P., McInnis, M. L. M., Church, R. M., & Machado, A. (2007). Shifts in the psychophysical function in rats. *Behavioural Processes, 75,* 167-175. doi:10.1016/j.beproc.2007.02.002

Guilhardi, P., Yi, L., & Church, R. M. (2007). A modular theory of learning and performance. *Psychonomic Bulletin & Review, 14*, 543-559. doi: 10.3758/BF03196805

Izquierdo, A., Brigman, J. L., Radke, A. K., Rudebeck, P. H., Holmes, A. (2017). The neural basis of reversal learning: An updated perspective. *Neuroscience, 345,* 12-26. doi:10.1016/j.neuroscience.2016.03.021

Jozefowiez, J., Machado, A., & Staddon, J. E.R. (2014). Cognitive versus Associative Decision Rules in Timing. In Arstila V. & Lloyd D. (Eds.), Subjective Time: The Philosophy, Psychology, and Neuroscience of Temporality (pp. 355-376). MIT Press. Retrieved from www.jstor.org/stable/j.ctt9qf5dd.29

Jozefowiez, J., Staddon, J. E. R., & Cerutti, D. T. (2009). The behavioral economics of choice and interval timing. *Psychological Review, 116*, 519-539. doi:10.1037/a0016171

Killeen, P. R. & Fetterman J. G. (1988). A behavioral theory of timing. *Psychological Review, 95*(2), 274–295.

Killeen, P. R., & Fetterman, J. G. (1988). A behavioral theory of timing. *Psychological Review, 95*, 274-295. doi:0033-295X/88/I00.75

Killeen, P. R., Fetterman J. G., & Bizo, L. A. (1997). Time's causes. In C. M. Bradshaw, & E. Szabadi (Eds.), *Time and Behaviour: Psychological and Neurobehavioural Analyses* (pp. 70–132). Amsterdam: North-Holland.

Kirkpatrick, K., & Church, R. M. (1998). Are separate theories of conditioning and timing necessary? *Behavioural Processes, 44*, 163-182. doi:10.1016/S0376-6357(98)00047

Laude, J. R., Stagner, J. P., Rayburn-Reeves, R. M., & Zentall, T. R. (2014). Midsession reversals with pigeons: Visual versus spatial discriminations and the intertrial interval. *Learning and Behavior, 42,* 40–46. doi: 10.3758/s13420-013-0122-x

Lejeune, H., & Wearden, J. H. (2006). Scalar properties in animal timing: Conformity and violations. *The Quarterly Journal of Experimental Psychology*, *59*, 1875-1908. doi:10.1080/17470210600784649

Linares, D., & López-Moliner, J. (2016). quickpsy: An R package to fit psychometric functions for multiple groups. *The R Journal, 8*, 122-131.

Machado, A. (1997). Learning the temporal dynamics of behavior. *Psychological Review, 104,* 241–265.

Machado, A., & Guilhardi, P. (2000). Shifts in the psychometric function and their implications for model of timing. *Journal of the Experimental Analysis of Behavior, 74*(1), 25-54.

Machado, A., Malheiro, M. T., & Erlhagen, W. (2009). Learning to time: A perspective. *Journal of the Experimental Analysis of Behavior, 92*(3), 423-458. doi:10.1901/jeab.2009.92-423

McMillan, N., & Roberts, W. A. (2012). Pigeons make errors as a result of interval timing ina visual, but not a visual-spatial, midsession reversal task. *Journal of Experimental Psychology: Animal Behavior Processes, 38,* 440–445. doi: 10.1037/a0030192

McMillan, N., & Spetch, M. L. (2019). Anticipation of a midsession reversal in humans. *Behavioural Processes, 159,* 60-64. doi: 10.1016/j.beproc.2018.12.016

McMillan, N., Kirk, C. R., & Roberts, W. A. (2014). Pigeon *(Columba livia)* and Rat (*Rattus norvegicus)* performance in the midsession reversal procedure depends upon cue dimensionality. *Journal of Comparative Psychology, 128*, 357-366. doi:10.1037/a0036562

McMillan, N., Spetch, M. L., Roberts, W. A., & Sturdy, C. B. (2017). It's all a matter of time: Interval timing and competition for stimulus control. *Comparative Cognition and Behavior Reviews, 12,* 83-103. doi:10.3819/CCBR2017.120007

McMillan, N., Sturdy, C. B., & Spetch, M. L. (2015). When is a choice not a choice? Pigeons fail to inhibit incorrect responses on a go/no-go midsession reversal task. *Journal of Experimental Psychology: Animal Learning and Cognition, 41*, 255-265. doi:10.1037/xan0000058

McMillan, N., Sturdy, C. B., Pisklak, J. M., & Spetch, M. L (2016). Pigeons perform poorly on a midsession reversal task without rigid temporal regularity. *Animal Cognition, 19,* 855–859. doi: 10.1007/s10071-016-0962-9

Orduña, V., Hong, E., Bouzas, A. (2007). Interval bisection in spontaneously hypertensive rats. *Behavioural Processes, 74, Behavioural Processes, 74,* 107-111. doi:10.1016/j.beproc.20106/10/013

Pavlov, I. P. (1927). Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex. G. V. Anrep, Ed. London, UK: Oxford University Press.

Rayburn-Reeves, R. M. & Zentall, T. R. (2013). Pigeons' use of cues in a repeated five-trial-sequence, single-reversal task. *Learning and Behavior, 41*(2), 138–147. doi: 10.3758/s13420-012-0091-5

Rayburn-Reeves, R. M., & Cook, R. G. (2016). The organization of behavior over time: Insights from mid-session reversal. *Comparative Cognition & Behavior Reviews, 11*, 103-125. doi:10.3819/CCBR.2016.110006

Rayburn-Reeves, R. M., & Zentall, T. R. (2013). Pigeons' use of cues in a repeated five-trial-sequence, single reversal task. *Learning and Behavior, 41*, 138-147. doi:10.3758/s13420-012-0091-5

Rayburn-Reeves, R. M., Laude, J. R., & Zentall, T. R. (2013). Pigeons show near-optimal win-stay/lose-shift performance on a simultaneous discrimination, midsession reversal task with short intertrial intervals. *Behavioural Processes, 92,* 65–70. doi:

Rayburn-Reeves, R. M., Mollet, M., & Zentall, T. R. (2011). Simultaneous discrimination reversal learning in pigeons and humans: Anticipatory and perseverative errors. *Learning and Behavior, 39,* 125–137. doi: 0.3758/s13420-010-0011-5

Rayburn-Reeves, R. M., Moore, M. K., Smith, T. E., Crafton, D. A., & Marden, K. L. (2018). Spatial midsession reversal learning in rats: Effects of egocentric cue use and memory. *Behavioural Processes, 152*, 10-17. doi:10.1016/j.beproc.2018.03.005

Rayburn-Reeves, R. M., Stagner, J. P., Kirk, C. R., & Zentall, T. R. (2013). Reversal learning in rats (*Rattus norvegicus*) and pigeons (*Columba livia*): Qualitative differences in behavioral flexibility. *Journal of Comparative Psychology, 127,* 202-211. doi:10.1037/a0026311

Rayburn-Reeves, R., Laude, J., & Zentall, T. (2013). Pigeons show near-optimal win-stay/lose-shift performance on a simultaneous-discrimination, midsession reversal task with short intertrial intervals. *Behavioural Processes,* 92, 65–70. doi:10.1016/j.beproc.2012.10.011

Roberts, W. A. (2002). Are animals stuck in time? Psychological Bulletin, 128(3), 473–489. doi: 10.1037/0033-2909.128.3.473

Roberts, W. A. (1981). Isolation of an internal clock. Journal of the Experimental Analysis of Behavior: Animal Behavior Processes, 7, 242–268.

Sackett, R. S. (1939). The effect of strength of drive at the time of extinction upon resistance to extinction in rats. *Journal of Comparative Psychology, 27*(3), 411–431. doi: 10.1037/h0063123

Sanabria, F., & Kileen, P. R. (2008). Evidence for impulsivity in the spontaneously hypertensive rat drawn from complementary response-withholding tasks. *Behavioral Brain Function.* doi: 10.1186/1744-9081-4-7

Santos, C. & Sanabria, F. (2020). Past outcomes and time flexibly exert joint control over midsession reversal performance in the rat. *Behavioural Processes.*

Santos, C., Soares, C., Vasconcelos, M., & Machado, A. (2019). The effect of reinforcement probability on time discrimination in the midsession reversal task. *Journal of the Experimental Analysis of Behavior, 111,* 371-386. doi:10/1002/jeab.513.

Smith, A.P., Pattison, K. F., & Zentall, T. R. (2016). Rats' midsession reversal performance: The nature of the response. *Learning and Behavior, 44,* 49–58. doi: 10.3758/s13420-015-0189-7

Staddon, J. E. R. (2010). *Adaptive Behavior and Learning* (Internet Edition). Cambridge, UK: Cambridge University Press.

Strang, C. G., & Sherry, D. F. (2014). Serial reversal learning in bumblebees (Bombus impatiens). *Animal Cognition, 17*(3), 723-734. doi: 10.1007/s10071-013-0704-1

Stubbs, D. A. (1980). Temporal discrimination and a free-operant psychophysical procedure. *Journal of the Experimental Analysis of Behavior, 33*, 167-185.

van Horik, J. O. & Emery, N. J. (2018). Serial reversal learning and cognitive flexibility in two species of Neotropical parrots (Diopsittaca nobilis and Pionites melanocephala). *Behavioural Processes, 157,* 664-672.

Williams, D. A., Frame, K. A., & LoLordo, V. M. (1992). Discrete signals for the unconditioned stimulus fail to overshadow contextual or temporal conditioning. Journal of the Experimental Analysis of Behavior: Animal Behavior Processes, 18, 41–55. doi: 0097-7403/92/$3.00

Williams, S. (1938). Resistance to extinction as a function of the number of reinforcements. *Journal of Experimental Psychology, 25*(5), 506–522. doi: 10.1037/h0053675

Zarcone, T. J., Branch, M. N., Hughes, C. E., & Pennypacker, H. S. (1997). Keypecking during extinction after intermittent or continuous reinforcement as a function of the number of reinforcers delivered during training. *Journal of the Experimental Analysis f Behavior, 67,* 91–108. doi: 10.1901/jeab.1997.67-91

Zeiler, M. D., & Powell, D. G. (1994). Temporal control in fixed-interval schedules. *Journal of the Experimental Analysis of Behavior, 61*, 1-9.

Zentall, T. R., Andrews, D. M., Case, J. P., & Peng, D. N. (2019). Less information results in better midsession reversal accuracy by pigeons. *Journal of Experimental Psychology: Animal Learning and Cognition, 45,* 422–430. doi: 10.1037/xan000215

Appendices

Appendix A

LeT model parameters used to simulate individual pigeons' performance.

	LeT simulation parameters			
Birds	**α**	**μ**	**σ**	***CV***
P458	0.02	0.03	0.00675	0.225
P917	0.01	0.03	0.00675	0.225
P935	0.02	0.05	0.01125	0.225
P730	0.01	0.03	0.00675	0.225
P851	0.01	0.03	0.00675	0.225
PG17	0.01	0.03	0.00675	0.225
PG23	0.03	0.04	0.01200	0.300

Note. For all birds $\beta = .95$ and $W_0 = .5$ for all states.

Appendix B

Maximum-likelihood estimates of the location (μ), scale (σ), and range (γ, λ) parameters of Equation 1 and the coefficient of variation (CV) for each simulated pigeon with the LeT model in each experimental condition.

	Conditions									
	Int - Int					High - High				
Birds	***μ***	***σ***	***γ***	***λ***	***CV***	***μ***	***σ***	***γ***	***λ***	***CV***
P458	39.60	10.02	0.01	0.04	0.25	38.62	9.80	0.00	0.01	0.25
P917	39.62	9.64	0.01	0.03	0.24	38.85	10.03	0.00	0.02	0.26
P935	38.93	10.03	0.02	0.05	0.26	37.65	10.42	0.00	0.02	0.28
P730	38.88	10.20	0.01	0.05	0.26	38.46	9.42	0.00	0.01	0.24
P851	38.83	10.12	0.01	0.04	0.26	40.83	10.34	0.00	0.03	0.25
PG17	38.98	10.11	0.01	0.03	0.26	40.39	10.07	0.00	0.02	0.25
PG23	38.22	12.03	0.04	0.08	0.31	39.27	11.88	0.00	0.04	0.30
Average	39.01	10.31	0.01	0.05	0.26	39.15	10.28	0.00	0.02	0.26
95% CI	38.65	*9.73*	*0.01*	*0.03*	*0.25*	*38.33*	*9.70*	*0.00*	*0.01*	*0.25*
	39.37	*10.88*	*0.02*	*0.06*	*0.28*	*39.98*	*10.86*	*0.00*	*0.03*	*0.28*
	High - Low					Low - High				
Birds	***μ***	***σ***	***γ***	***λ***	***CV***	***μ***	***σ***	***γ***	***λ***	***CV***
P458	44.81	9.62	0.00	0.10	0.21	33.69	10.27	0.04	0.01	0.30
P917	45.51	10.20	0.00	0.09	0.22	33.72	10.30	0.04	0.01	0.31
P935	44.79	9.53	0.00	0.13	0.21	33.46	9.93	0.08	0.01	0.30
P730	45.42	10.00	0.00	0.10	0.22	33.73	10.03	0.03	0.01	0.30
P851	45.11	10.59	0.00	0.09	0.23	33.46	10.19	0.03	0.01	0.30
PG17	44.94	10.06	0.00	0.10	0.22	33.86	9.87	0.04	0.01	0.29
PG23	44.90	11.78	0.00	0.19	0.26	32.32	11.86	0.11	0.01	0.37
Average	45.07	10.26	0.00	0.12	0.23	33.46	10.35	0.05	0.01	0.31
95% CI	*44.85*	*9.69*	*0.00*	*0.09*	*0.22*	*33.07*	*9.84*	*0.03*	*0.01*	*0.29*
	45.29	*10.82*	*0.00*	*0.14*	*0.24*	*33.85*	*10.86*	*0.08*	*0.01*	*0.33*

Appendix C

Reversal trials *not* experienced by each bird on the phases where the reversal trial was sampled from the even-number list

		Subject			
Phase	*Condition*	*P816*	*P389*	*PG16*	*PG18*
1	Int-Int	40	40	38	38
3	Low-High	24	26	64	42
		P053	*P435*	*P795*	*P762*
1	Int-Int	46	56	22	26
3	High-Low	16	20	52	40