http://researchspace.auckland.ac.nz

University of Auckland Research Repository, ResearchSpace

Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the Library Thesis Consent Form and Deposit Licence.
The Properties of Reinforcement: Strengthening Versus Signalling

Joshua Bensemann

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

School of Psychology
The University of Auckland
2015
Abstract

The properties of reinforcement are currently being debated in operant psychology. The traditional view is that reinforcers strengthen the behaviour that precedes it, whereas a modern view treats reinforcers as discriminative stimuli that an organism uses to determine which behaviour will produce the next reinforcer. This thesis approached the issue as if both properties exist, and used local analyses of choice procedures to isolate each property for observation.

In Experiment 1, the signalling property was isolated by creating an environment where the previously reinforced response alternative was removed and replaced with a previously unavailable response alternative. This prevented the subjects from making the response that had just been reinforced and therefore limited the usefulness of strengthening based explanations for local response patterns. The subjects responded to those environmental contingencies by showing preference for the locally richer alternative immediately after reinforcement. When Experiment 1 was altered so that subjects could respond to the just-reinforced alternative, there was a small increase in preference for the just-reinforced alternative, but only when that alternative could produce reinforcers.

In Experiment 2, the strengthening property was isolated by removing the relevance of any discriminative stimulus that was produced by reinforcement which meant subjects could use not use an individual reinforcer to guide them towards the next reinforcer. Subjects responded by showing preference for the richest alternative immediately after reinforcement rather than preferring the location of the last reinforcer, as has been observed in local analyses of previous concurrent VI VI procedures. Data from Experiment 2 were used in two additional analyses to provide support for the conclusions reached during that experiment. The first analysis involved computer simulations of responding to ensure that response patterns were caused by the effects of reinforcement, and not due to artefacts created by analysis procedures. The second analysis fitted alternative choice models to test the stability of the response data.

Overall, at local levels of analysis there was more evidence to support signalling properties of reinforcement than strengthening properties, although this does not mean that strengthening properties of reinforcement do not exist.
Acknowledgements

First of all I would like to acknowledge my supervisors: Associate Professor Douglas Elliffe, Dr Brenda Lobb, and Associate Professor Christopher Podlesnik. All three of you have supported me in different ways and subtly (or not so subtly) pushed me towards completing this thesis. In addition, having the three other points of view improved the quality of this research well above anything one student could do on their own.

I also want to acknowledge the extended operant lab family for being part of my postgrad journey as well as letting me share in their own. Rather than risk forgetting anyone I will keep this to the current lab roster but also acknowledge that there are many more names that belong here: John Bai, Vikki Bland, (next week Dr) Denys Brand, Dr Sarah Cowie, Stephanie Gomes-Ng, Stuart McGill and Dr Ludmila Miranda Dukoski. I would especially like to acknowledge Ludmila and Denys as the two people who have put up with me the most. Ludmila and I have shared our mutual postgrad journeys closely and I am thankful for that. Denys and I started postgrad together and as usual he managed to submit well before me. He also thinks that he shouldn’t be counted as part of the lab but should also know by now that I am always right. I’ve also tried to convince Ludmila that I’m always right but she refuses to buy it.

Outside of academia, I would like to acknowledge both the Auckland Uniguides and the Shaolin Gao Can Mun Nam Pai Chuan Kung Fu School. Both of these organisations are comprised of many great people who have shown genuine interest in me and in making the world a better place.

Finally, none of this would be possible without my first teachers, Dianne Bacica-Bensemann and Shane Bensemann. My parents have always supported my interest in science and have continued to encourage me to do what makes me happy rather than what leads to the greatest financial gain.
Contents

Abstract .................................................................................................................................... III
Acknowledgements .................................................................................................................. IV
Chapter 1. Introduction .............................................................................................................. 1
  1.1. The Law of Effect ............................................................................................................... 2
  1.2. Signalling Theory ............................................................................................................. 5
  1.3. Strengthening versus signalling ....................................................................................... 11
  1.4. Other properties of reinforcers ...................................................................................... 14
  1.5. This Thesis ...................................................................................................................... 15
Chapter 2. Experiment 1 .......................................................................................................... 16
  2.1. Introduction ..................................................................................................................... 16
  2.2. Method ............................................................................................................................ 20
    2.2.1 Subjects ...................................................................................................................... 20
    2.2.2 Apparatus .................................................................................................................. 20
    2.2.3 Procedure ................................................................................................................... 21
  2.3. Results ............................................................................................................................ 26
    2.3.1 Data Sets .................................................................................................................... 26
    2.3.2 Issues ......................................................................................................................... 26
    2.3.3 Response ratios between alternatives ........................................................................ 27
    2.3.4 Local analysis .............................................................................................................. 33
    2.3.5 Sets 2 and 3 ............................................................................................................... 34
    2.3.6 Set 4 ........................................................................................................................... 39
    2.3.7 Set 5 ........................................................................................................................... 45
    2.3.8 Set 6 ........................................................................................................................... 49
  2.4. Discussion ....................................................................................................................... 51
Chapter 3. Experiment 2 .......................................................................................................... 58
  3.1. Introduction ..................................................................................................................... 58
  3.2. Method ............................................................................................................................ 65
    3.2.1 Subjects ...................................................................................................................... 65
    3.2.2 Apparatus .................................................................................................................. 65
    3.2.3 Procedure ................................................................................................................... 66
  3.3. Results ............................................................................................................................ 68
    3.3.1 Stability ...................................................................................................................... 68
3.3.2 Constant-Ratio Rule ........................................................................................................ 70
3.3.3 Local Analysis ................................................................................................................ 72
3.3.4 Effect of CODs ............................................................................................................... 79
3.3.5 Changeovers .................................................................................................................... 80
3.3.6 Dynamic interactions between responses and food ......................................................... 88
3.4. Discussion ......................................................................................................................... 96
Chapter 4. Simulations ........................................................................................................... 103
4.1. Introduction ..................................................................................................................... 103
4.2. Method ............................................................................................................................ 107
4.2.1 Simulation ..................................................................................................................... 107
4.2.2 Correction Procedure ................................................................................................. 109
4.3. Results ............................................................................................................................. 110
4.3.1 Simulation ..................................................................................................................... 110
4.3.2 Correction procedure ................................................................................................. 129
4.4. Discussion ....................................................................................................................... 142
Chapter 5. Modelling ............................................................................................................. 147
5.1. Introduction ..................................................................................................................... 147
5.2. Modelling and Analysis ................................................................................................. 150
5.3. Discussion ....................................................................................................................... 155
Chapter 6. Discussion ............................................................................................................ 159
6.1. General results ................................................................................................................ 159
6.2. Signalling properties of reinforcement ......................................................................... 162
6.3. Strengthening properties of reinforcement .................................................................. 165
6.4. Reinforcers and behaviour ......................................................................................... 169
Chapter 7. References ............................................................................................................ 175
Chapter 1. Introduction

A subject’s behaviour is the product of its environment. Because of this all behavioural research involves manipulating an environment to see how that modifies the subjects’ behaviour. More specifically, a subject’s environment is often changed by arranging rewards that the organism receives for its actions. These rewards are known as reinforcers, and there are multiple theories as to why a subject’s behaviour is modified by the presentation of a reinforcer. One theory, the law of effect (Thorndike, 1911), is that reinforcers mechanistically increase the likelihood that the immediately preceding behaviour will reoccur. This increase in likelihood, also known as strengthening, means that whatever biological mechanism caused the subject to emit that behaviour gets primed to occur again. A competing theory is that reinforcers are discriminative stimuli, also known as signposts, which are used as signals to inform the organism as to how it can obtain more reinforcers (Davison & Baum, 2006; Shahan, 2010). In signalling theory, the reinforcer itself is not what is controlling behaviour; rather the organism uses cues from the environment to increase its chances of receiving more reinforcers. Therefore, the main distinction between the two theories is that strengthening compels the subject to emit the strengthened behaviour, whereas signalling suggests which behaviour should be emitted.

This thesis describes several behaviours as being ‘strengthened’. The construct of response strength has been defined in several different ways in the history of behaviour analysis (e.g., Nevin, 1974). Skinner almost always equated response rate with response strength. In Nevin’s behavioural momentum metaphor, response rate is measured as resistance to disruption. The distinction may be important when the experimental contingencies affect response rate independently of underlying response strength. For example, a differential reinforcement of low rates (DRL) schedule can maintain strong
behaviour in the sense of persistent behaviour, but the contingency requiring long inter-
response times means that response rates must be low. Because no experiment in this thesis
manipulated response rate independently of reinforcer rate, as in the example of a DRL
schedule, nothing is lost by assuming response rate to measure response strength, as in
Skinner’s usage.

1.1. The Law of Effect

A major advantage of Thorndike’s (1911) law of effect that is can be used for
parsimonious explanations of the basic principles of operant psychology. One such
demonstration is using reinforcers to manipulate both human and non-human animals to
produce desired behaviours using the method of successive approximations also known as
shaping (Skinner, 1953). The basic process of shaping was demonstrated by Wolf, Risley and
Mees (1964), where therapists used candy as a reinforcer to get a child to wear his glasses.
An explanation of why shaping works is that each time candy was given during shaping, the
likelihood of that behaviour reoccurring increased. Because the child and all other organisms
can only produce a limited number of behaviours within a set time, increasing that behaviour
leads to decreases in others. With each reinforcer the chosen set of behaviours becomes
stronger and will eventually dominate non-reinforced behaviours. Then the therapists chose a
subset of that set of behaviours, and that set was reinforced until those behaviours begin to
dominate the rest. The concept of strengthening explains easily how reinforcers can be used
to manipulate a selected behaviour.

It is also possible to use the law of effect to explain how strengthening increases
indirectly selected behaviour, as demonstrated by experiments of superstitious behaviour
(Skinner, 1948; Wagner & Morris, 1987). For example Skinner (1948) showed that
presenting a subject with response non-contingent reinforcers at regular intervals led to an
increase in the behaviour that happened to occur immediately before that reward. Because that behaviour was more likely to occur again, there was an increased chance that the behaviour would occur before the next non-contingent reinforcer and therefore be increased again. This creates a cycle of repeated strengthening for that behaviour. As with shaping, strengthening one class of behaviours led to a decrease in other behaviours, due to limits in maximum behavioural output, and resulted in the eventual domination of the superstitious behaviour.

As there are limits in behavioural output, an organism is constantly forced to choose amongst behaviours in its repertoire. Therefore all behaviour is the outcome of a choice (Herrnstein, 1970), and consequently choice research can provide evidence for many principles of operant psychology, including the strengthening properties of reinforcement. The basic choice procedure presents the subject with two or more mutually exclusive response alternatives and each alternative provides reinforcers according to different contingencies. Responses to each alternative are compared, and when more responses are made to one alternative than the others, it is said that preference has been shown for that alternative. The term ‘preference’ as used in this thesis is intended to be theory-free, and refers simply to calculated values that represent the relative rates of responding. It does not refer to any mental state of a subject. As with the construct of response strength discussed earlier, no experiment used contingencies that might maintain a low response rate despite being ‘preferred’ by the subject, such as a DRL schedule. Consequently, increases or decreases in responding to an alternative are referred to as increases and decreases in preference for that alternative. Preference can be measured at two levels when using a choice paradigm, local (short-term) and global (long-term). Global preference measures a subject’s overall proportion of responses to each alternative, resulting in a few numbers that represent the many choices that were made. Usually the total number of responses to each alternative
reflects the number of reinforcers received for responding to that alternative (Herrnstein, 1961, 1970). Local preference analyses show short-term changes in the subjects’ behaviour. Doing this shows how reinforcers and other experimental events cause preference to change, and that short-term preference patterns can be modified by many factors (for example, the number of successive reinforcers on a particular alternative, Davison & Baum, 2000). The choice between using a global or local analysis depends on what type of behaviour needs to be observed. In general, it is more beneficial to observe changes to experimental variables using global analyses because fewer numbers are easier to interpret and display graphically. However, local analyses are more beneficial when observing the subjects’ reactions to specific experimental events as those data are destroyed during global analyses. This thesis is focused on the effects of each reinforcement event, therefore, unless stated otherwise all future references to preference will refer to preference measured using local analyses.

When focusing on local preference, the law of effect predicts immediate increases in post-reinforcer preference for the alternative that produced the immediately preceding reinforcer. The law of effect explains this as increased strengthening of responses to the reinforced alternative, and this pattern has been observed in the past. For example Shimp’s (1966) Experiment 1 presented subjects with a two-alternative discrete-trial procedure where the probability of reinforcers being produced by the alternatives was fixed for each response and was determined by stimuli associated with the alternatives. When the probability on the left alternative reinforcer was .75, choosing a left response would produce food on 75% of left choices and choosing right produced food on 25% of right choices. Shimp reported that the probability of responding to the left alternative increased when the immediately preceding response was to the left and had resulted in food. Furthermore, the probability of choosing left was higher still if the two prior responses were to the left and both produced food.
Identical results were observed for the probability of responding to the right. Shimp interpreted these results as being evidence for the strengthening property of reinforcement.

However, not all experiments have produced results that support strengthening explanations. Nevin’s (1969) Experiment 1 used a similar discrete-trial procedure but arranged food reinforcers according to a variable interval (VI) schedule. Therefore, unlike Shimp’s (1966) procedure, not every trial could produce a reinforcer. The results of Nevin’s experiment did not show an increased probability of responding to an alternative when the immediately preceding response was to the same alternative and produced food.

Others have observed increased local preference for the alternative that produced the immediately preceding reinforcer in free operant procedures. However this effect appears to be short-term. For example, Menlove (1975) observed an immediate shift in preference towards the alternative that had just produced a reinforcer, and then a gradual shift back towards a preference pattern that represents a global level of preference. Menlove referred to this pattern as a positive immediacy effect. This pattern was demonstrated by collating responses based on how much time had elapsed since the last reinforcer, and then producing response ratios for each group of collated responses. This response pattern has been observed by others (Boutros, Davison, & Elliffe, 2009; Davison & Baum, 2002; Elliffe & Davison, 2010; Landon, Davison, & Elliffe, 2002) and graphs created by this type of collation are known as preference pulses.

1.2. Signalling Theory

While the majority of examples mentioned so far have been used as evidence for the strengthening property of reinforcement, they can also be explained in terms of signalling. Both shaping and superstition can be explained as the reinforcers signalling to the organism that it should do what it just did to obtain further reinforcers. Even though there was no
causal connection between the behaviour and reinforcer in the superstition example, the subject misinterpreted the reinforcer-based stimulus as a signal for it to repeat the last behaviour to gain more reinforcement, and so to repeat the preceding behaviour. The choice procedure results could also be explained as attempts to gain further reinforcement. However these explanations are not as parsimonious as those that are based on the law of effect.

The advantage of signalling theory is that it explains results that the law of effect cannot. A simple example of this is the typical response pattern observed when using Fixed Interval (FI) schedules of reinforcement where reinforcers are arranged at set times after the previous reinforcer delivery. Predictions based on the law of effect are that subjects should increase their probability of responding (which increases the response rate) immediately after a reinforcer, which is not what is observed. Instead subjects tend to decrease their response rate immediately after a reinforcer and then gradually increase the rate as time passes and time to the next reinforcer arrangement decreases (Ferster & Skinner, 1957). While this response pattern does not support the law of effect, it does support signaling theory. Because reinforcers are arranged at set times after the previous reinforcer, the delivery of a reinforcer signals that the next reinforcer will not be arranged until the scheduled time has passed. Subjects react to this information by decreasing their response rates after the reinforcer because no further reinforcers can be delivered immediately, and increase their response rates as time increases.

Another example of response patterns that cannot be explained by the law of effect was observed in a choice experiment by Krägeloh, Davison and Elliffe (2005). They arranged reinforcement contingencies so that the overall numbers of reinforcers on each alternative were equal, to keep the global levels of preference equal. However, they also arranged reinforcers probabilistically based on the location of the immediately preceding reinforcer. For example if a .9 probability was arranged, as it was in their Condition 4, then there was a
90% chance that the next reinforcer would occur on the location that produced the immediately preceding reinforcer. Under these conditions, Krägeloh et al. observed that shifts in local preference were determined by the probability used. If all reinforcement does is strengthen behaviour, then all post-reinforcement shifts in local preference should have been equal in this situation, because reinforcers produced for each alternative were equal. In addition, Krägeloh et al. found that preference shifted away from the alternative that produced a reinforcer when there was a low probability of the next reinforcer being produced from that same location. Because the law of effect always predicts an increase in preference for the reinforced alternative, any shifts away from that alternative are evidence against strengthening properties of reinforcement. Shifts away from the just-productive alternative were also reported by Boutros, Elliffe, and Davison (2011) when they arranged strict-alternating conditions. Strict alternation meant that the next food would always be produced by responses to the alternative that did not produce the last food reward. While strengthening cannot explain these results, signalling theory can. When the probability of the next reinforcer is associated with the location of the previous reinforcer, reinforcement based-stimuli act as signals which provide information as to the likely location of the next reinforcer. If there is a .9 probability that the next reinforcer will come from the same location as the immediately preceding reinforcer, like in Condition 4 of Krägeloh et al., then the subject should prefer the just-productive location after reinforcement because of the high likelihood of that location producing reinforcement again. However if the probability is 0, then the next reinforcer will never come from the previous location. Therefore, the subject should not prefer the just-productive location after reinforcement and should shift preference to the other alternative.

A related effect was observed by Davison and Baum (2006) using a frequently changing procedure (Davison & Baum, 2000). In frequently changing procedures the
reinforcement contingencies are changed multiple times per session, meaning that the subjects’ behaviour is often readapting to the changing contingencies. In general, results of these procedures show that subjects’ preference is related to the reinforcement ratio in operation at the time – the more reinforcers produced on an alternative, the more preference the subjects show towards that alternative (Davison & Baum, 2000; 2002; Landon & Davison, 2001). Davison and Baum’s (2000) experiment varied the reinforcer ratios after every 10 reinforcers and found that subjects’ preference also shifted constantly in reaction to the new reinforcement ratio, as expected. Additionally, responding to an alternative in their procedure would occasionally deliver a non-reinforcer stimulus (either a key-light or magazine light) that was correlated with the location of the next reinforcer, instead of producing the reinforcer itself. When the stimulus was correlated with a high likelihood that the next reinforcer would occur on the same alternative that produced the stimulus, Davison and Baum observed a small but immediate shift in preference towards that same alternative. If the stimulus meant that reinforcement was likely to occur on the other alternative, the subjects shifted preference to that other alternative. The stimuli were being treated as a signal for the location of the next reinforcer, similar to how subjects were using reinforcers as signals for the location of the next reinforcer in Krägeloh et al.’s (2005) experiment.

Davison and Baum (2006) and others, such as Shahan (2010), have interpreted results that show similarity between the effect of reinforcers and other stimuli to mean that the reinforcer does not strengthen behaviour. Instead a reinforcer is a discriminative stimulus in these interpretations. According to this theory of reinforcement, a reinforcer acts as a signal that informs the subject about how to gain more reinforcers. If a reinforcer signals more reinforcers for repeating the same behaviour, then the subject is likely to repeat that behaviour. In a choice scenario this means that the subject prefers to respond at the location that produced the reinforcer. If the reinforcer is signalling more reinforcers for other known
behaviours, the subject is more likely to emit those behaviours. If food from one location
signals more food on the other alternative, it causes subjects to shift preference to that other
alternative. This theory also explains why some evidence for choice experiments appear to
support the law of effect with results showing shifts in preference towards the just-reinforced
alternative (e.g., Landon et al., 2002; Menlove, 1975). In those experiments the reinforcer
from each alternative signalled more reinforcement from that alternative. This caused more
responses to the alternatives that produced the reinforcers and produced a pattern that was
consistent with predictions based on the law.

If reinforcers are discriminative stimuli as suggested by signalling theory, then they
should have the same signalling properties as discriminative stimuli that are not reinforcers.
One of these properties is that a signal’s effects on preference is contextual. An example of
contextual control with non-reinforcer stimuli was observed by Boutros, Davison, and Elliffe
(2009) when they conducted a steady-state version of Davison and Baum’s (2006) procedure.
Boutros et al. kept the reinforcer ratio constant for an entire condition instead of changing it
multiple times per session. They observed no effects of non-reinforcement stimuli on the
subjects’ choice as seen in Davison and Baum’s experiment. The signals did not cause
changes to preference. One way to explain this is due to information redundancy. Since the
reinforcer ratio was not changing constantly, the subjects could use their experience (all
previously obtained information) to predict the location of the next reinforcer without the
added signals. In other words, the addition of the signal did not provide the subjects with any
new information as to the likely location of the next reinforcer. Therefore the signal did not
influence choice at a local level. In Davison and Baum’s procedure, however, the reinforcer
ratio was unpredictable due to the subjects’ lack of experience with that ratio; additional
signals did provide useful information about the location of the next reinforcer, and so the
subjects reacted to it.
This same pattern of information redundancy has also been observed in experiments where reinforcers were the only stimuli used, suggesting that reinforcers are signals. For example, this is demonstrated when comparisons are made between Landon et al.’s (2002) steady-state concurrent VI schedule experiment and Landon and Davison’s (2001) equivalent frequently changing procedure. Landon and Davison observed large shifts in preference after each reinforcer, whereas Landon et al. observed small shifts in preference after reinforcement. If reinforcers are discriminative stimuli that share the properties of signals then this finding is consistent with the differences between Boutros et al.’s (2009) results and those of Davison and Baum (2006). Landon and Davison’s results are like Davison and Baum’s; the reinforcer ratio was unpredictable so each reinforcer/signal provided useful information to the subjects as to how they should behave to obtain more reinforcers. Landon et al.’s results are like Boutros et al.’s; the addition of the reinforcer/signal did not provide the subjects with any new information about the likely location of the next reinforcer.

Further investigations of the signalling properties of reinforcement have shown that they can be used to signal more than the location of the next reinforcer. Cowie, Davison, and Elliffe (2011) demonstrated that reinforcers can signal the likely time of the next reinforcer. They did this by making the timing of the next reinforcer depend on the location of the previous one. For example, in some conditions a reinforcer from one alternative meant that, if the next reinforcer was from that same alternative, it would be arranged relatively sooner after the previous reinforcer delivery. However, if the next reinforcer was from the other alternative, then it would be arranged at a relatively later point in time. In other conditions, the reverse was true. Cowie et al. observed that subjects preferred the alternative that was linked to earlier reinforcer immediately after food, and then shifted preference to the other alternative as time since the last food delivery increased.
Others have demonstrated that post-reinforcement preference is controlled by complex processes involving time and probabilities that cannot be predicted using the law of effect. Davison, Cowie, and Elliffe (2013) arranged concurrent FI VI schedules and randomly determined which schedule produced the next reinforcer. They found that preference after a reinforcer was not towards the just-reinforced location, instead initial preference depended on the FI schedule. When the FI schedule was short, preference after reinforcement was towards the schedule and when the FI schedule was long, preference after reinforcement was towards the VI schedule. Both Cowie, Elliffe, and Davison (2013), and Miranda-Dukoski, Davison, and Elliffe (2014) arranged procedures where the probability of reinforcement on each alternative changed systematically across time since last reinforcer. Cowie et al. (2013) had one change point, whereas Miranda-Dukoski et al. had one change point per second. Subjects’ post-reinforcement preference in both experiments was not towards the location of just-reinforced alternative. Instead preference was towards the location that had the greater probability of reinforcement during that time period.

1.3. Strengthening versus signalling

Both strengthening and signalling theories predict the same thing for local behavioural patterns when environmental contingences are arranged so that reinforcers are delivered for repeating the reinforced behaviour. Signalling theory predicts an increase in the preceding behaviour when the reinforcers signal further reinforcers for making the same response, and law of effect/strengthening based predictions will always predict an increase in the preceding behaviour after the reinforcer. However, predictions using the theories differ when environmental contingencies are arranged so that the following reinforcer is unlikely to be delivered for repeating the preceding behaviour. In that scenario, signalling theory predicts that a subject’s behaviour will shift to perform the behaviour that produces the next reinforcer, assuming that the next reinforcer is predictable. Recent research has shown that
this is what organisms do (Boutros, Elliffe, & Davison, 2011; Krägeloh et al., 2005). In this second scenario, predictions from law of effect are the same as they are from the first scenario, the subject will be more likely to repeat the preceding behaviour, and therefore, this prediction is incorrect.

Because signalling is able to predict many behavioural patterns that strengthening cannot, it is tempting to say that strengthening processes do not exist. However, there are results where subjects appeared to be responding as if their behaviour has been strengthened even though their preference pattern supports signalling. For example, when the probability of continued reinforcement on an alternative was 0 in both Krägeloh et al.’s (2005) and Boutros et al.’s (2011) procedures, subjects had some preference for the just-productive alternative, even though that alternative would never produce the next reinforcer. After initial preference for that alternative, preference quickly shifted towards to location that was signalling to produce the next reinforcer. While that result can be interpreted in multiple ways, such as a stay bias by the subject, or as a misinterpretation of the signal, it can also be interpreted as a local strengthening effect. If this is the case then the response to the just-productive alternative was strengthened, increasing preference for it, while the reinforcer-based stimuli increased preference for the other alternative. If this was what happened, then the strengthening effect and the signalling effect were interfering with each other.

This suggestion of interference fits with the conclusions of others who have suggested that behaviour might be under more than one form of control that changes depending on factors, such as amount of training. For example Machado and Keen (1999) observed that subjects started with a matching pattern and then appeared to develop a momentary maximization strategy with increased exposure to their environmental contingences. A similar suggestion came from Racey, Young, Garlick, Pham, and Blaisdell (2011), who stated that behaviour that occurs early after a condition change is controlled by explorative
tendencies whereas later behaviour is controlled by exploitative tendencies. Explorative behaviours are those that gain information, such as switching to each alternative in order to determine which has the highest rate of reinforcement. Exploitative behaviours are those that use that information to gain more reinforcement, such as responding more to the alternative that produces the highest rate of reinforcement. The idea of joint control has also been suggested by Balleine and Dickinson (1998) who suggested that goal-directed behaviours are controlled by one region of the brain for rats and humans, whereas habitual behaviours are controlled by another. Goal-directed behaviours are those that direct subjects to gaining more reinforcers and habitual behaviours are trained behaviours that the subjects repeat. Balleine and Dickinson demonstrated that these behaviours are controlled by separate areas of rats’ brains by lesioning, which caused impairments to the associated type of behaviour. It is interesting that exploitative behaviours, habitual behaviours, and behaviours predicted by strengthening explanations of reinforcement are similar to each other. All three are types of behaviour that repeat the same action; exploitative behaviours are repetitions of the most profitable behaviour, habitual behaviours are repetitions of the most trained behaviour, and strengthen behaviours are repetitions of the previous behaviour. There are also similarities between explorative behaviours, goal-directed behaviours, and behaviours that are signalled by reinforcement as they are all reactions to the availability of information. Explorative behaviours attempt to gain information about that environment, signalled behaviours are reactions to the information, and goal-directed behaviours both gain and react to information. It might be the case that behaviours contained within these two groupings are fundamentally the same and the variations within the groups were created by the observer. If that is the case then the neurological evidence from Balleine and Dickinson suggest these two processes are separate from each other. Perhaps the potential strengthening/signalling effect observed by
Boutros et al. (2011) and Krägeloh et al. (2005) shows what happens when neither strengthening nor signalling has full control over behaviour.

1.4. Other properties of reinforcers

There is an alternative explanation for the preference patterns that occurred when the probability of continued reinforcement was 0 in both Krägeloh et al.’s (2005) and Boutros et al.’s (2011) procedures. It might be the case that the increase in preference for the just-productive alternative was caused by the analysis that was used to observe it, because both of results were obtained using preference-pulse analyses. Recently, McLean, Grace, Pitts, and Hughes (2014) suggested that these analyses can artificially create preference patterns that show subjects preferring the just-productive alternative. In their suggestion, reinforcers occur during bouts of responses to each alternative. Because preference pulses typically collate responses based on the time elapsed since reinforcement, the responses that occur after the reinforcer might be part of the response bout, and might not be altered by the immediately preceding reinforcer, as strengthening and signalling theories suggest. The non-altered responses become the first part of the data that is used for determining post-reinforcer preference, and because they occur on the just-productive alternative, the preference pattern shows preference for the just-productive alternative. It might have been the case that subjects did not show preference for the just-productive alternative at all, but appeared to when analysed because reinforcers have been used as time markers by the researchers. As there have been many conclusions drawn based on preference pulse analyses (Examples include: Cowie et al., 2013; Davison & Baum, 2002; 2007; Landon, Davison, & Elliffe, 2003b; Miranda-Dukoski et al., 2014), it is important to determine whether this unintended data organisational effect has mislead these conclusions.
1.5. This Thesis

Because there are multiple explanations for why local preference patterns occur post-reinforcement, it became of interest to examine how a reinforcer works and determine what properties they have. Attempts were made to separate strengthening from signalling properties to observe each property in isolation. Signalling properties were isolated by eliminating the just-productive alternative after reinforcement, which prevented any strengthening-based preference from occurring. Strengthening properties were isolated by making reinforcer-based signals redundant so that subjects did not need to follow the discriminative signal in order to gain further reinforcement. Isolating each property showed whether both properties exist and provided a better understanding of how environmental contingencies control behaviour. It was also of importance to investigate McLean et al.’s (2014) explanations for local preference patterns to determine how much of the conclusions made about reinforcement properties could be explained via other means.
Chapter 2. Experiment 1

2.1. Introduction

Recent research has demonstrated that the law of effect (Thorndike, 1911) does not explain all patterns of local preference. The law of effect states that behaviour followed by a favourable consequence (reinforcer) leads to an increase in the probability of the reoccurrence of that behaviour; this process is known as strengthening. In choice research this phenomenon is detected as an increase in preference for the alternative that produced the reinforcer. Research has shown that post-reinforcer preference shifts towards the just-productive location in some circumstances (Davison & Baum, 2002; Landon, Davison, & Elliffe, 2002; Menlove, 1975). However, this is not always the case as Krägeloh, Davison, and Elliffe (2005) demonstrated that under certain conditions preference will shift away from the just-productive alternative. They manipulated the probability of continued reinforcement on an alternative and observed that subjects shifted preference away from the just-productive alternative when the probability of continued reinforcement was low. Boutros, Elliffe, and Davison (2011) also observed that subjects shifted preference away from the just-productive alternative when the next reinforcer was always from the other alternative. When the probability of continued reinforcement is low as it was in these conditions, the probability of reinforcement on the non-productive alternative is high. In these situations preference was towards the likely location of the next reinforcer, not the location of the previous reinforcer. This means that receiving a favourable consequence did not cause an increase in preference for the just-productive location. Instead the favourable consequence guided behaviour toward the likely location of further favourable consequences in the future.
Findings such as these that contradict the predictions from the law of effect have been used to generate a signalling-based theory of reinforcement, which treats each reinforcer as a discriminative stimulus. Subjects use the presentation of the reinforcer-based stimulus to guide their behaviour toward gaining further reinforcers. Therefore results such as those of Krägeloh et al. (2005), and Boutros et al. (2011), are explained as reinforcers from one location signalling the likely location of the next reinforcer. After the signal, subjects adjust their preference accordingly. This theory also explains results where preference increases for the just-productive alternative, as would be expected from the law of effect. The increase in preference is due to the reinforcer signalling that more reinforcers will be gained by repeating the same behaviour, and not because the immediately preceding behaviour has been strengthened.

However, this does not mean that the strengthening process described by the law of effect does not exist. There might be evidence for both signalling theory and the law of effect in Krägeloh et al.’s (2005) and Boutros et al.’s (2011) results. Krägeloh et al.’s results show that at 0 probability of continued reinforcement, immediately after reinforcement some subjects preferred the just-productive alternative equally as much as they preferred the alternative that was certain to arrange the next reinforcer (i.e. showed indifference). After the initial preference there was a shift in responding to the location of the next reinforcer before shifting back towards indifference. Analyses from experiments prior to Krägeloh et al.’s typically show that responding immediately after reinforcement favours one alternative, and then preference shifts towards indifference as time since reinforcement passes. Perhaps under the conditions of Krägeloh et al.’s experiment, preference was initially at indifference because law of effect based strengthening effects were counter-acting signalling effects. Strengthening effects drive preference towards the just-reinforced alternative, signalling effects direct preference towards the next-reinforced alternative. Both alternatives are
presented to the subject immediately after reinforcement; therefore, it was possible for both preference patterns to compete.

Boutros et al. partially avoided this issue by using a switching-key procedure (Findley, 1958). Only one alternative was presented after reinforcement and that was determined probabilistically. The alternative presented after reinforcement had a noticeable effect on preference during strict alternating conditions, where the probability of continued reinforcement on the just-reinforced alternative was 0. If the just-reinforced alternative was presented, the results were identical to the equivalent conditions from Krägeloh et al.’s (2005) procedure (i.e., indifference immediately after reinforcement, a shift to the next-reinforced alternative, then back to indifference). When the next-reinforced alternative was presented after reinforcement, subjects showed strong preference for that alternative before shifting preference back towards indifference. After comparing Boutros et al.’s with Krägeloh et al.’s results, it appears that the presence of the just-reinforced alternative interferes with preference for the next-reinforced alternative. Whether this is a strengthening effect or something else, the fact remains that the preference patterns were altered.

It also should be noted that the results of Boutros et al. (2011) were similar to those of Hachiga, Sakagami, and Silberberg (2014), who arranged a concurrent variable ratio (VR) extinction procedure using rats as subjects. After each reinforcer in their experiment, a light was illuminated above one of the two alternatives, and that light indicated which alternative would produce the next reinforcer. When the alternative that was illuminated was the just-productive alternative, there was extreme preference for that alternative. However, when the other alternative was illuminated, subjects showed some initial preference for the just-productive alternative before shifting preference to the alternative that would produce the next reinforcer. Therefore it also appears as if there was some sort of interference between a signalling and strengthening effect. It should be noted that in Hachiga et al.’s procedure, the
signals were generated by non-reinforcer-based stimuli instead of by the reinforcer as in Boutros et al.’s procedure. With that said, evidence from Davison and Baum (2006, 2010) showed that non-reinforcer-based stimuli that were correlated with the location of the next reinforcer caused shifts in preference towards that location, similar to reinforcers that also signal the location of the next food. This means that the signals used by Hachiga et al. should have had a similar function to Boutros et al.’s reinforcers, making this an appropriate comparison.

If the correct interpretation of the results from Krägeloh et al.’s (2005) and Boutros et al.’s (2011) experiments is that reinforcers have both signalling and law of effect based effects, then the two theories are not mutually exclusive. In addition, Boutros et al.’s results suggest that it is possible to limit any strengthening effects of reinforcement by presenting the subject with an alternative that did not produce the last reinforcer. From this observation it follows that the signalling property of reinforcement can be isolated from whatever interference is caused by having a just-productive alternative available after reinforcement. This was done in the present experiment by removing the just-productive alternative after reinforcement and replacing it with another alternative, meaning that subjects could not respond to the just-productive alternative post-reinforcement but still had a choice of alternatives to respond to. The conditions of the experiment can be placed into three groups. In one group of conditions (Sets 1, 2 & 3), the reinforcer rates on each alternative were fixed so that the reinforcer ratio operating at any time depended on the available alternatives. In the second group of conditions (Sets 4 & 5), the subjects could respond to the just-productive alternative, but that either provided no reinforcement (Set 4) or was only possible in trials where the just-productive alternative was selected to appear (Set 5). In the final group of conditions (Set 6), the probability of reinforcement was dependent on the location of the last reinforcer, similar to Krägeloh et al.’s and Boutros et al.’s experiments.
2.2. Method

2.2.1 Subjects

Six pigeons, numbered 61 to 66 were used as subjects. They were housed individually. After experimental sessions, subjects were fed mixed grain to maintain them at 85% ± 15 g of their free-feeding weights; water and grit was available at all times.

2.2.2 Apparatus

The subjects were housed in the experimental chambers. These chambers were 375 mm high, 375 mm wide and 370 mm deep. The back and side walls of the chamber consisted of sheet metal. The front wall consisted of metal bars, as did the floor and top of the chamber. Within each cage were two wooden perches which intersected at 90°. One perch was 100 mm from and parallel to the right wall, the other was 100 mm from and parallel to the front wall. Situated on the right wall, 200 mm above the wooden perches were 3 response key alternatives. These keys were 20 mm in diameter and spaced 80 mm apart. A force of 0.1 N or more was required for a response to be registered. Situated directly below the centre key was the magazine aperture. It measured 50 mm by 50 mm, and it was 60 mm above the perches. During food deliveries all key lights were extinguished and the hopper containing wheat was raised for 3 s. The magazine aperture was lit while the hopper was raised. A computer running MED-PC® software controlled all experimental events, and recorded the times that they occurred, to 10 ms resolution. The room lights were switched on at 12:30 am and off at 4:00 pm. Sessions began daily at 1 am; no personnel entered the room while sessions were running.
2.2.3 Procedure

All subjects had previous experience with operant procedures so no pre-training was required. Experimental sessions began with two or three alternatives being lit, dependent on the condition. Only lit keys were active and eligible to produce food for responses. In two-alternative conditions, the alternative inactive/unlit at start-up was randomly determined so that all three alternatives had a .33 probability of being inactive.

During sessions, a programmed exponential VI schedule was active. Every second the programmed schedule had a probability of arranging a reinforcer that depended on the condition and alternatives currently active (lit). In Conditions 1 – 3, 13, and 20 – 22, the overall exponential schedule was a variable interval (VI) 30 s schedule and reinforcers were allocated probabilistically according to the ratio between each pair of alternatives. In all other conditions each alternative was associated with a different VI schedule, and the overall exponential VI schedule in operation was the combined rate of the two alternatives which were active at the time. For example, if one alternative was associated with a VI 20 s schedule (3 reinforcers/min) and the other was associated with a VI 60 s schedule (1 reinforcer/min), the overall schedule in operation was VI 15 s (4 reinforcers/min). The schedule associated with the inactive alternative did not time. When a food delivery was arranged according to the overall schedule, it was allocated to one of the active alternatives based on the ratio of food deliveries available. In the VI 20 s versus VI 60 s example the food would be allocated according to a 3:1 ratio. Then the schedule would stop timing until that food reinforcer had been delivered. The VI schedules associated with each alternative are shown in Tables 2.1 and 2.2. In the first two sets of conditions the subjects needed to respond exclusively to an alternative for 2 s before a reinforcer could be collected (2 s COD; Herrnstein, 1961). The changeover requirement was removed for subsequent conditions. Sessions ended after 60 foods had been delivered or 60 min had elapsed, whichever occurred...
first. All responses were recorded irrespective of whether they were to active or inactive alternative.

Sets 1, 2, & 3

Sessions in all conditions of Sets 1 – 3 started with only two of the three alternatives active. After a food was delivered, the alternative that produced the delivery became inactive, while the previously inactive alternative became active. This meant that subjects could not receive a food delivery from the alternatives that had produced the immediately preceding delivery. Each condition was run for 40 sessions; sessions 21 – 40 were used for analysis.

Set 4

Like all other previous sets of conditions, there were only two alternatives active at any time during Set 4. However the inactive alternative was also lit, meaning that three alternatives were lit after a food delivery but only two of the alternatives could produce the next delivery. After a food was delivered, the alternative that provided the delivery remained lit but became inactive, while the previously inactive alternative became active. Each condition was run for 40 sessions; sessions 21 – 40 were used for analysis.

Set 5

Sessions in all conditions in Set 5 started with only two of the three alternatives active. However unlike other sets of conditions, the alternatives that were active after a food delivery were determined randomly. This means that unlike the first four sets, it was possible for the alternative that produced the immediately preceding food delivery to produce the next delivery, if and only if that alternative was selected to be active post-reinforcement. More data were required for Set 5 compared to Sets 1 – 4, due to the many combinations of
previous reinforcer location and currently available alternatives. Therefore 6 conditions of 60 sessions were run, and sessions 41 – 60 were used for analysis.

Set 6

Set 6 conditions started with only two of the three alternatives active, and after a food was delivered the alternative that provided the delivery became inactive, while the previously inactive alternative became active. However conditions differed from others sets by the way that food allocations were determined. A VI 30 s schedule was operating at all times and the food deliveries were allocated by a single ratio that was associated with the condition. The richer alternative of any pairing was determined by which alternative had been active prior to the immediately preceding reinforcer. For example, if a 3:1 ratio was operating, the alternative that was active before the immediately preceding reinforcer but had not produced it would be three times more likely to produce the next reinforcer than the newly active (previously inactive) alternative. In the above example, if the left and centre key were active before the immediately preceding reinforcer and the centre produced the next reinforcer, the left and right key would then be active and the left key would be associated with the 3 from the 3:1 ratio. At the start of the session the richer of the two starting alternatives was randomly determined. Due to the complicated contingences arranged in this dataset, conditions were run for 100 sessions and sessions 51 – 100 were used for analysis.
Table 2.1

VI values for Sets 2, 3, 4 and 5.

<table>
<thead>
<tr>
<th>Condition #</th>
<th>Key 1</th>
<th>Key 2</th>
<th>Key 3</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>VI 60 s</td>
<td>VI 20 s</td>
<td>VI 180 s</td>
<td>2 s COD</td>
</tr>
<tr>
<td>5</td>
<td>VI 180 s</td>
<td>VI 60 s</td>
<td>VI 20 s</td>
<td>2 s COD</td>
</tr>
<tr>
<td>6</td>
<td>VI 20 s</td>
<td>VI 180 s</td>
<td>VI 60 s</td>
<td>2 s COD</td>
</tr>
<tr>
<td>Set 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>VI 180 s</td>
<td>VI 60 s</td>
<td>VI 20 s</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>VI 20 s</td>
<td>VI 180 s</td>
<td>VI 60 s</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>VI 60 s</td>
<td>VI 20 s</td>
<td>VI 180 s</td>
<td></td>
</tr>
<tr>
<td>Set 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>VI 60 s</td>
<td>VI 20 s</td>
<td>VI 180 s</td>
<td>All Keys Lit</td>
</tr>
<tr>
<td>15</td>
<td>VI 180 s</td>
<td>VI 60 s</td>
<td>VI 20 s</td>
<td>All Keys Lit</td>
</tr>
<tr>
<td>16</td>
<td>VI 20 s</td>
<td>VI 180 s</td>
<td>VI 60 s</td>
<td>All Keys Lit</td>
</tr>
<tr>
<td>Set 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>VI 20 s</td>
<td>VI 180 s</td>
<td>VI 60 s</td>
<td>Random Keys Active</td>
</tr>
<tr>
<td>11</td>
<td>VI 180 s</td>
<td>VI 60 s</td>
<td>VI 20 s</td>
<td>Random Keys Active</td>
</tr>
<tr>
<td>12</td>
<td>VI 60 s</td>
<td>VI 20 s</td>
<td>VI 180 s</td>
<td>Random Keys Active</td>
</tr>
<tr>
<td>17</td>
<td>VI 20 s</td>
<td>VI 60 s</td>
<td>VI 180 s</td>
<td>Random Keys Active</td>
</tr>
<tr>
<td>18</td>
<td>VI 60 s</td>
<td>VI 180 s</td>
<td>VI 20 s</td>
<td>Random Keys Active</td>
</tr>
<tr>
<td>19</td>
<td>VI 180 s</td>
<td>VI 20 s</td>
<td>VI 60 s</td>
<td>Random Keys Active</td>
</tr>
</tbody>
</table>
## Table 2.2

### Ratios used for Sets 1 and 6.

<table>
<thead>
<tr>
<th>Condition #</th>
<th>Key 1:Key 2</th>
<th>Key 2:Key 3</th>
<th>Key 1:Key 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Set 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3:9</td>
<td>9:1</td>
<td>3:1</td>
</tr>
<tr>
<td>2</td>
<td>9:1</td>
<td>1:3</td>
<td>9:3</td>
</tr>
<tr>
<td>3</td>
<td>1:3</td>
<td>3:9</td>
<td>1:9</td>
</tr>
<tr>
<td><strong>Set 6</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>3:1</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>9:1</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>Ext:1</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>1:Ext</td>
<td></td>
</tr>
</tbody>
</table>
2.3. Results

2.3.1 Data Sets

For Sets 1 – 5, all the data from the same set of conditions were collated so that responses to the richest alternative of each condition were combined; responses to the second richest of each condition were combined, and so forth. This process merged the 3 or 6 conditions into a single 9:3:1 ratio dataset. This eliminated the effect of any bias towards individual key locations as each reinforcer rate was assigned to each key. This method has been used for both experiments in this thesis and is identical to the method used by both Davison et al. (2007), and Elliffe and Davison (2010). In these results, responses to the alternatives in the 9:3:1 reinforcer ratio were identified as B₉, B₃, and B₁, respectively. Reinforcers delivered from those alternatives were identified as R₉, R₃, and R₁, respectively.

2.3.2 Issues

There were a few issues with the present experiment that need to be addressed before certain analyses can be conducted. The experiment was designed to prevent subjects from responding to the just-reinforced alternative by deactivating the key light. The only exceptions to this were Set 4 when no key lights were deactivated and Set 5 when the deactivating alternative was randomly determined. The assumption was that subjects would respond as though the deactivated alternative was not available. However, the deactivated alternative was physically present and it was possible for the subjects to respond to it (i.e., to peck the darkened and inactive key). Responses to the deactivated alternative were therefore recorded to confirm that this seldom occurred. No subject made more than 20 responses to the unlit alternatives per set, compared to hundreds of thousands of responses they each made to lit alternatives. This shows that the procedure was an effective way to remove the just-reinforced alternative and was functioning as intended.
A second issue was the arranged reinforcer ratios. In the current procedure, the reinforcer ratio was 9:3:1. If this was a typical concurrent choice procedure, the obtained reinforcer ratio would approximate 9:3:1. However the current procedure caused differences between local and global reinforcer ratios. When B₉ was paired with B₃, B₉ was three times more likely to produce food, and was nine times more likely to produce food when paired with B₁. These results were obtained during pairwise comparisons when using a 9:3:1 ratio. A different pattern occurs at the overall level; B₉ only produced 1.2 times the total reinforcers of B₃ and only 3 times the total reinforcers of B₁. This was due to the uneven number of presentations of each pairing, the B₉:B₃ pairing only occurred after an R₁ reinforcer, which was the least likely event to occur. Due to this issue, the first set of conditions was created to determine whether preference between a pair of alternative was controlled by the obtained pairwise level ratio or the obtained overall level. Figure 2.1 shows preference between the B₉:B₃ and B₃:B₁ pairs of Set 1. There appears to be no systematic differences across subjects, suggesting that the subjects responded according to the 3:1 pairwise ratio, not the overall ratio. To confirm this, a Wilcoxon matched-pairs signed-ranks test was conducted and found no significant difference between preference in the B₉:B₃ and B₃:B₁ pairings (p = .753; Figure 2.1).

2.3.3 Response ratios between alternatives

Figures 2.1 – 2.4 show how subjects allocated responses between pairs of alternatives in the different datasets. The choice allocations of interest were the B₉:B₃ and B₃:B₁ pairings because these pairings had the same ratio of reinforcement. As reported in the proceeding section there was no systematic preference pattern between pairs in Set 1 (2 s COD, Same RF Rates), in addition signed-ranks tests showed that there were no systematic patterns in Set 3 (No COD, Different RF Rates; p = .753; Figure 2.3) or Set 5 where the random pairing of alternatives appeared after reinforcement (p = .116; Figure 2.4). However Figure 2.2 shows
that data obtained during Set 2 (2 s COD, Different RF Rates) produced a systematic pattern. The preference for B₉ over B₃ was clearly greater than preference for B₃ over B₁ for all subjects except Subject 65. Although data obtained from Subject 65 show opposing results, a signed-ranks test reveals that the tendency to prefer B₉ more was statistically significant (p = .046). Both pairings have a 3:1 reinforcement ratio and since the overall reinforcement rate for the B₉:B₃ pair was higher than the B₃:B₁ pair, that result was consistent with previous findings of an overall reinforcer rate effect (Alsop & Elliffe, 1988; Elliffe & Alsop, 1996). However, this effect might also have been expected to occur in Sets 3 and 5, and Figures 2.3 – 2.4 show that it did not.
Log response ratios for responses made between pairs of keys in Set 1. The dashed line represents the arranged reinforcer ratio. Each panel shows data from individual subjects.

Figure 2.1
Figure 2.2

Log response ratios for responses made between pairs of keys in Set 2. The dashed line represents the arranged reinforcer ratio. Each panel shows data from individual subjects.
Figure 2.3

Log response ratios for responses made between pairs of keys in Set 3. The dashed line represents the arranged reinforcer ratio. Each panel shows data from individual subjects.
Figure 2.4

Log response ratios for responses made between pairs of keys in Set 5. The dashed line represents the arranged reinforcer ratio. Each panel shows data from individual subjects.
2.3.4 Local analysis

Response data from each set of conditions (after Set 1) were collected into bins according to time since the last reinforcer. Each bin accounted for 3 s since reinforcement, which means all responses made at $\geq 0$ s and $< 3$ s were placed in the first bin, responses made at $\geq 3$ s and $< 6$ s were placed in the second bin, and so on. Within bins, data were arranged according to which alternative produced the immediately preceding reinforcer and which response was made. The log ratio of responses on the two available alternatives was then calculated for each bin for all sets except Set 4, proportion of responses were calculated for that set. All data points that could not be transformed (usually due to exclusive responding to one alternative) were removed, although this rarely occurred. Responses to the just-reinforced (and therefore deactivated) alternative have also been excluded from analyses. Log ratio data were then plotted against time bins to produce preference pulses (Davison & Baum, 2002). The x-axis represents the minimum value of each time bin, meaning the ratio at 0 s represents the responses that were $\geq 0$ s and $< 3$ s. This process was used for all preference pulses that have been reported.

There are differences in the way that data were presented for preference pulses from different sets of conditions. Preference pulses in Sets 2 and 3 (Figures 2.5 – 2.6) are shown separately for the choice between B9 and B3 (i.e., for choice following an R1 reinforcer, because B1 became unavailable after an R1 reinforcer), for choice between B9 and B1 (i.e., following an R3 reinforcer), and for B3 versus B1 (i.e., after R9). The preference pulses from Set 4 (Figures 2.9 – 2.11) show proportion of responses for all three alternatives as they were all available after a reinforcer. Each panel displays responses after a different alternative reinforcer. Each panel in Set 5 (Figures 2.12 – 2.14) displays responses between a specific pair of alternatives, the pulses within each panel show responses to that pair after different reinforcers. This was done because the previous reinforcer did not determine the next
configuration of alternatives, therefore any pairing could occur after each reinforcer. Finally, the pulses from Set 6 (Figure 2.15) were modified so that data were collapsed across pairs of alternatives with the responses to the old alternative grouped together and responses to the new alternatives grouped together. Pulses have been separated by reinforcer ratio for the old alternative/new alternative.

2.3.5 Sets 2 and 3

Results from Set 2 were different from those of Set 3, as was the case with the overall response ratio comparison data. In Set 2 (Figure 2.5) there is some indication that preference for B₉ over B₃ was greater than for B₃ over B₁ in the first 5–10 s after reinforcement, at least for 5 of the 6 birds – although both pulses were still always smaller than the B₉:B₁ pulse. Figures 2.5 – 2.6 show preference pulses of similar shape to those from previous research, in that preference was strongly towards one alternative immediately after reinforcement and declined gradually towards indifference with increasing time since reinforcement. The alternative preferred was not, however, the just-reinforced alternative – it could not be, because the just-reinforced alternative became unavailable after reinforcement and until the next reinforcer. Rather, preference was always towards the richer alternative of the pair, strongly so just after reinforcement but still in that direction even 30 s after reinforcement. This preference could not indicate a strengthening effect, because the alternative to which responding should have been strengthened was not available. Rather, preference was towards the richer alternative of the available pair. This centrality of the role of the reinforcer ratio is supported by the finding that the pulse for the B₉:B₁ choice, the most extreme reinforcer ratio, was always substantially above those for the B₉:B₃ and B₃:B₁ ratios. Data obtained from all subjects in Set 3 show little difference in shapes of pulses for the B₉:B₃ and B₃:B₁ pairs, with the possible exception of Subject 62 (Figure 2.6).
Preference pulses showing log response allocations over time since reinforcement in Set 2. Preference is shown separately for each pair of alternatives and data from each subject are displayed separately.
Figure 2.6

Preference pulses showing log response allocations over time since reinforcement in Set 3. Preference is shown separately for each pair of alternatives and data from each subject are displayed separately.
Both pulses indicate preference for the richer of the two available alternatives immediately after a reinforcer delivery, after that the subjects’ preference shifted towards a stability point and remained there.

Reinforcer delivery times from Sets 2 and 3 were also analysed. These data were collected into bins according to time since the last reinforcer, using the same procedure detailed for the response data preference pulses. Instead of calculating log ratios, each time bin was divided by the total number of reinforcers obtained for that pair within each dataset. These data were averaged across subjects and then plotted against time bin to produce a graph showing what proportion of total reinforcers was obtained from each alternative within that time bin.

The first noticeable difference between the sets was that reinforcers were obtained in the first 3 s bin of Set 3, but not in Set 2 (Figure 2.7). This is because the 2 s COD would always be in effect at this time during Set 2. If a reinforcer was arranged on an alternative immediately after the preceding reinforcement interval, subjects would be forced to respond to that alternative for 2 s before it could be obtained, as the previous response before reinforcement would have been to the deactivated alternative. This was not the case in Set 3 where a large proportion of reinforcers were obtained in the first 3 s of the B₉:B₃ and B₉:B₁ pairings. Additionally, the highest proportions of reinforcers obtained in a single time bin were much higher in Set 2 than Set 3.
Figure 2.7

Graphs showing the proportion of all reinforcers delivered at specific times since reinforcement for Sets 2 and 3. Each panel shows a different key pairing from each set. Data from all birds have been averaged.
2.3.6 Set 4

The reinforcement contingencies of Set 4 were identical to Set 3. There was no COD, the subjects could not receive reinforcement from the alternative that produced the immediately preceding reinforcer, and the location of the preceding reinforcer determined the reinforcer ratio operating after reinforcement. However, instead of turning a key light off to signal that an alternative was operating on an extinction schedule, all three alternatives were lit and available after the delivery of a reinforcer. Therefore, differences in response patterns between Sets 3 and 4 are due to changes in key lights, not reinforcement patterns.

Table 2.3 compares the mean proportional response allocation to each alternative compared to which alternative is on extinction in Sets 3 and 4. Observations of the data show that all subjects responded to the extinction schedule when the key light was lit, and rarely responded when it was not. An analysis of Set 4 that was similar to the analysis presented in Figure 2.8 shows that choice between B₉ and B₃ when B₁ was on extinction, was very different from that between B₃ and B₁ when B₉ was on extinction, unlike Set 3. However Table 2.3 suggests that subjects were responding to the reinforcer ratio produced by all three alternatives instead of the local rate of reinforcers on active alternatives. As mentioned above in the issues section, the overall obtained reinforcer ratio was not 9:3:1 and more specifically the reinforcer ratio of the B₉:B₃ pairing did not equal the reinforcer ratio of the B₃:B₁ pairing.

The data were arranged into preference pulses using proportions of responses to each alternative as the response measure instead of log transformations. Unlike the pulses from all other sets, responses to all three alternatives have been presented on the graphs. This method was used to display all alternatives simultaneously and was the same method that has been used to display data from more than two alternatives in other chapters.
Table 2.3

Proportion of responses allocated to each alternative in Sets 3 and 4 separated by which key was on extinction. The extinction key was unlit in Set 3 but lit in Set 4.

<table>
<thead>
<tr>
<th>Key on extinction</th>
<th>Alt 1</th>
<th>Alt 2</th>
<th>Alt 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Set 3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alt 1</td>
<td>0.000</td>
<td>0.752</td>
<td>0.248</td>
</tr>
<tr>
<td>Alt 2</td>
<td>0.875</td>
<td>0.000</td>
<td>0.125</td>
</tr>
<tr>
<td>Alt 3</td>
<td>0.756</td>
<td>0.244</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Set 4</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alt 1</td>
<td>0.395</td>
<td>0.399</td>
<td>0.206</td>
</tr>
<tr>
<td>Alt 2</td>
<td>0.447</td>
<td>0.374</td>
<td>0.179</td>
</tr>
<tr>
<td>Alt 3</td>
<td>0.445</td>
<td>0.353</td>
<td>0.202</td>
</tr>
</tbody>
</table>
Figure 2.8

Log response ratios for responses made between pairs of keys in Set 4. The dashed line represents the arranged reinforcer ratio. Each panel shows data from individual subjects.
Preference pulses Subjects 61 and 62 from Set 4 showing the proportion of responses for all three alternatives after a reinforcer. Each panel displays responses after a different alternative reinforcer.
Figure 2.10

Preference pulses Subjects 63 and 64 from Set 4 showing the proportion of responses for all three alternatives after a reinforcer. Each panel displays responses after a different alternative reinforcer.
Figure 2.11

Preference pulses Subjects 65 and 66 from Set 4 showing the proportion of responses for all three alternatives after a reinforcer. Each panel displays responses after a different alternative reinforcer.
Figures 2.9 – 2.11 show that response patterns were similar irrespective of which alternative produced the reinforcer. This suggests that subjects were not using the location of the just-productive alternative as a signal for extinction on that alternative post-reinforcement. That is, they did not discriminate that a reinforcer on one alternative would never be followed by another reinforcer on that alternative.

2.3.7 Set 5

Set 5 differs from all other sets by the fact that the just-productive alternative could be active following a reinforcer. The data have been arranged to show preference pulses based on the alternative that provided the previous reinforcer and which pair of alternatives was active following that reinforcer (Figures 2.12 – 2.14).

The general trend shows preference for the richer alternative of every pairing. In addition, the preference shown was modulated by the location of the previous reinforcer. Initial preference for the richer of the paired alternatives was higher when the immediately preceding reinforcer occurred on the now richer alternative than when it occurred on the now poorer alternative. When the just-productive location was not available following reinforcement, initial preference for the richer alternative tended to be between preference values shown when the richer or poor alternative were the just-productive alternatives. In other words, initial post-reinforcement preference for the richer alternative was, highest after a richer alternative reinforcer, lowest after a poorer alternative reinforcer, and in the middle after a reinforcer from the now unavailable alternative.
Preference pulses showing responses between B₉ and B₃ from Set 5. Each panel displays responses between a specific pair of alternatives, the pulses within each panel show responses to that pair after different reinforcers. Data for each subject have been presented separately.
Figure 2.13

Preference pulses showing responses between B₃ and B₁ from Set 5. Each panel displays responses between a specific pair of alternatives, the pulses within each panel show responses to that pair after different reinforcers. Data for each subject have been presented separately.
Figure 2.14

Preference pulses showing responses between B₉ and B₁ from Set 5. Each panel displays responses between a specific pair of alternatives, the pulses within each panel show responses to that pair after different reinforcers. Data for each subject have been presented separately.
2.3.8 **Set 6**

In Set 6 the local preference ratio depended on which pair of alternatives was available before the immediately preceding reinforcer and which pair was available after that reinforcer. For example, in Condition 13 the alternative in the preceding pair that had not produced a reinforcer remained available after reinforcement and became the richer of the two available alternatives with a 3:1 ratio. Preference pulses were modified so that data were collapsed across pairs of alternatives with the responses to the old alternative grouped together and responses to the new alternatives grouped together.

Results show that preference immediately after a reinforcer was toward the alternative that was currently richer (Figure 2.15). The degree of preference also seems to be related to the overall probability of reinforcement for that alternative. For example there was the most preference for the richer alternative when the probability of it being reinforced was 1 (Ext:1 or 1:Ext), and more preference for the richer alternative when there was a .9 probability than when there was a .75 probability. Preference tended to shift towards indifference after the initial preference for the rich alternative, except when there was a 0 or 1 probability of reinforcement on one of the alternatives. In that situation the preference always remained in favour of the richer alternative, which was also the only alternative that could have produced the next reinforcer. This result is similar to the findings of Boutros et al. (2011).
Figure 2.15

Preference pulses from Set 6. Data have been collapsed across pairs of alternatives with the responses to the old alternative grouped together and responses to the new alternatives grouped together. Pulses have been separated by reinforcer ratio for the old alternative/new alternative.
2.4. Discussion

The general findings show that when subjects could not respond to the alternative that produced the previous reinforcer, they showed preference for the location most likely to produce the next reinforcer. Findings also showed that when only two alternatives were available post-reinforcement, one of which was the location of previous reinforcer, initial preference was shifted towards the just-reinforced location. Finally, when all three alternatives were available, subjects were unable to discriminate that the just-productive location was on extinction.

Most concurrent-schedule experiments measure a subject’s preference between two (and occasionally more) alternatives that provide reinforcers at varying rates. Normally the alternatives and reinforcement contingencies available after a reinforcer has been presented are the same as those available before that reinforcer. Previous findings show that a typical pattern of post-reinforcement responding in this situation is for subjects to have greater initial preference to the location of the just-reinforced alternative and then to shift preference back towards the global reinforcement ratio (Davison & Baum, 2002; Landon et al., 2002). In Sets 2 and 3 of the present experiment, this pattern was prevented by replacing the just-reinforced alternative with the alternative that had produced the reinforcer immediately prior to that. Subjects showed greater initial preference for the richer of the two of the available alternatives before shifting preference back towards indifference. It appears the tendency to show increased preference for an alternative immediately after a reinforcer is not due solely to a local strengthening process because the tendency occurred when the strengthened (just-reinforced) alternative was unavailable.

One possible explanation for the results of the two-alternative sets is that subjects were responding to the local reinforcer rates that were operating on the alternatives that were
currently available. The reason for the initial increased performance could have been due to differences in obtained local reinforcer rates in that time period. Figure 2.7 shows that the proportion of reinforcers obtained by subjects after responding to the richer alternative in Sets 2 and 3 was generally greatest in the period immediately following a reinforcer. Therefore, if the subjects were tracking the local probability of reinforcement, it makes sense that they responded to the richer alternative immediately after reinforcement. However, the reinforcement distribution was dynamically related to the subjects’ response patterns; increasing preference at one time period will lead to a decreased likelihood that a reinforcer will be obtained on the other alternative, making the leaner alternative appear less productive during that same period than it potentially was. That is, the obtained reinforcer rates differed substantially from those arranged. This would have shifted preference away from the leaner alternative, because the rich alternative would have been perceived as locally richer than it actually was. This distribution was also impacted by the presence or absence of a 2-s changeover delay (COD) as can be seen in the differences in reinforcer distribution between Sets 2 and 3. The delay was programmed so that responses to an alternative could not produce reinforcers unless the subject had been exclusively responding to that same alternative for at least 2 s. Because the last response before a reinforcer was always to an alternative that would be unavailable after that reinforcer, the first post-reinforcer response had to be to a new alternative and would start a delay. If a reinforcer was arranged within 2 s after the previous one, that reinforcer could not be obtained immediately as the delay needed to elapse. The earliest these reinforcers could have been obtained was at 2 s after reinforcement, assuming that the subject started responding to the correct alternative immediately after the previous reinforcer. Therefore many of those reinforcers would be obtained after that period, and this explains why there was a large spike in the proportion of reinforcers obtained around that time on the rich alternative compared with Set 3. If the reinforcer was arranged on the
lean alternative in Set 2 then it was unlikely that the subject started responding to that alternative, causing more time to elapse before possible collection, leading to fewer reinforcers being obtained in this period compared to Set 3. This increase/decrease of obtained rich/lean alternatives explains why there was an initial shift towards the richer alternative post-reinforcement in Set 2. Previous research by Shull and Pliskoff (1967), and Krägeloh and Davison (2003) suggested that adding a COD led to long-lasting effects on preference that persisted after the COD was reduced or removed. Perhaps the COD effect carried through to non-COD conditions in the present experiment and was the reason why a consistent pattern of increased initial preference for the rich alternative occurred.

An alternative explanation for these patterns is that subjects were using the location of the previous reinforcer and other experimental events as stimuli to guide them to the likely location of the next reinforcer, similar to the suggestions of Davison and Baum (2006), and Shahan (2010). This would mean that the subject’s preference was not controlled by available alternatives; it was controlled by what had happened previously. There is evidence both for and against this possibility. Set 6 provides the strongest evidence for signalling as the available alternatives alone did not provide information about which was the richer alternative, yet the subjects were able to track this (Figure 2.15). Instead the richer alternative could be determined by a combination of which alternative provided the last reinforcer and which alternatives were active at that time. Another similar way of determining the currently richer alternative would be to keep track of which was the newly active alternative and then shift preference either towards or away from it, depending on the condition. That method would also require using information about events before the last reinforcer (i.e., which alternative was not there) and therefore would support signalling as the main cause of local preference patterns.
Evidence against signalling control as the primary method of preference control can be seen in Set 4 (Figure 2.8). The reinforcer contingencies for Set 4 were identical to Set 3; the alternative that produced the immediately preceding reinforcer became inactive and could not produce the next reinforcer. The only difference was that the inactive alternative remained lit at all times. Under these conditions, the subjects would continue to respond to the just-productive alternative, even though it would never produce the next reinforcer. If the subjects were using the last reinforcer as a signal for more reinforcement, then they should not have responded to that alternative until after the next reinforcer had been delivered and the alternative had become active again. Instead it seems that the subjects’ response patterns were similar after reinforcers from every alternative regardless of the location that produced it. Subjects preferred the alternative that was richest globally the most and the leanest alternative the least. Like all other datasets, preference immediately after reinforcement favoured the rich alternative before decreasing but remained the most preferred alternative. If subjects were following a reinforcer-based signal in Set 4, then they were not discriminating the signal for extinction.

Data from Set 5 show mixed evidence for signalling and possible evidence for a different type of control, strengthening. Unlike other sets, it was possible for subjects to respond to the location that provided the immediately preceding reinforcer, but only when that alternative was selected to appear. The subjects continued to prefer the alternative that was the richer of the two available alternatives even though the available alternatives were not determined by the location of the previous reinforcer (Figure 2.12 – 2.14). This means that preference had to be controlled by the available alternatives and could not be controlled by a signalling effect. If the preference could not be signalled by reinforcement in Set 5 then it is possible that preference in other sets was not controlled by the location of the last reinforcer. However, preference in Set 5 appeared to be controlled by the location that had
produced the immediately preceding reinforcer in a different way. If the alternative producing that reinforcer reappeared as the richer of the two active alternatives, initial preference for the richer alternative was greater than when the alternative producing that reinforcer reappeared as the leaner alternative. This suggests that the immediately preceding reinforcer was enhancing preference for that alternative, exactly as expected with a strengthening effect.

An alternative explanation for the increased preference to the rich alternative has been provided by McLean, Grace, Pitts, and Hughes (2014). They suggested that many features of preference pulses are caused by the way that data are processed to produce that analysis. Using simulations they showed that combining response patterns from alternatives can produce pulses that suggest increased initial preference that decreases over time, when in reality the subjects could be switching between alternatives and responding in bursts. However their simulation requires bursts of responses to an alternative to be interrupted by reinforcers for that alternative, as after the interruption the burst to that alternative continues. This continuation of the response burst is what they believe causes part (if not all) of the preference for the just-productive alternative. This was not possible in most conditions of the present experiment as the removal of the just-productive location ended any response bouts at time of the reinforcer. With that said, McLean et al.’s suggestion might explain the results from Set 5. When the just-productive alternative reappeared after reinforcement, the response bout could continue, which would shift post-reinforcer preference patterns towards that alternative, without that alternative increasing the probability of response the just-productive alternative.

In summary, the results of the present experiment suggest that post-reinforcer preference is not controlled by a single factor. When the just-productive alternative was removed, preference favoured the richer of the available alternatives. Set 5 showed that subjects could do this by using post-reinforcement information; such as which alternatives
were active. However Set 6 shows that subjects could do this using pre-reinforcement information such as which alternative was not active beforehand, or which alternative was active before but non-productive. Set 4 shows that the reinforcement scheduling is not the only thing that controls preference. Keeping the inactive alternative lit altered subjects preference patterns, even though the arranged reinforcement scheduling was identical to Set 3.
Chapter 3. Experiment 2

3.1. Introduction

The primary focus of Chapter 2 was on the signalling properties of reinforcement and how they influence post-reinforcement preference when isolated from possible strengthening properties. This chapter takes the opposite approach, focusing on strengthening properties and limiting signalling properties. This was achieved by decreasing the usefulness of any signal that is produced by the reinforcer. More specifically the reinforcer-based stimulus did not provide information about the likely location of the next reinforcer and so all previously obtained reinforcers could not guide the subjects towards that location.

One way to reduce the information contained within a signal (i.e., a reinforcer in these experiments) is to repeat that same information many times. For example, being told that there is food in the fridge is interesting information the first time that you hear it, but by the tenth time the information is redundant. However when the information changes to ‘there is no more food’, then it becomes interesting again, but can also become redundant when repeated. The same principle can be applied to explain differences in post-signal preference between frequently changing environments and steady-state environments. The generic term of ‘signal’ has been used as it refers to both the signal produced by reinforcers acting as discriminative stimuli and signals produced by any explicit discriminative stimuli that are not also reinforcers.

A frequently changing environment is created when experimental contingencies change multiple times per session. For example Davison and Baum (2000) changed their contingencies after every component of 10 food deliveries, with up to 7 different components being presented in each session. A steady-state environment is produced by a procedure that keeps the contingencies constant within a session and typically repeats the same
contingencies for many consecutive sessions. Due to the different rates of environmental change, the main difference between the two procedures is the amount of experience the subject gains with the contingency before a change. This means that if subjects are learning about their environment through the signals provided by reinforcers, then information in the steady-state environment will be repeated many more times than the frequently changing environment. Therefore, any individual signal should be less informative about the environment in steady-state conditions. In a frequently changing environment, most signals are informative because the information they contain changes frequently.

Differences in post-signal response patterns have been observed between frequently changing and equivalent steady-state procedures and are assumed to be caused by differences in the amount of information that is provided by the signal. Experiments with frequently changing procedures have shown large changes in post-reinforcement preference (Aparicio, 2008; Baum & Davison, 2004; Davison & Baum, 2000, 2002, 2003, 2007; Landon & Davison, 2001). The results of those experiments have shown that subjects shift a lot of preference to the just-productive alternative (i.e., increase relative response rates on that alternative), and will keep making large shifts with successive food deliveries. Results from equivalent steady-state procedures, on the other hand, suggest that shifts in preference are largely determined by which alternative produced the reinforcer (Landon, Davison, & Elliffe, 2002; 2003). If the reinforcer comes from the lean alternative there is a relatively large shift toward that alternative, but there is a relatively small shift if it comes from the rich alternative. In the above scenarios, the reinforcers’ signal will carry different weights. In the frequently changing environment the current reinforcer ratio will be uncertain, and each reinforcer will provide information as to what the likely ratio will be. More reinforcers from one alternative suggest that that alternative is likely to be the rich, therefore preference shifts towards it. However data analysed from steady-state environments are typically taken from
later sessions, at which point the subject has gained a lot of information about the likely ratio. Gaining more reinforcers from the alternative that has been providing the most reinforcers is less useful in determining the current rich alternative.

Other experiments have shown that in general, subjects react differently to signals depending on whether they are informative. For example, Davison and Baum (2006, 2010) arranged a frequently changing procedure where some reinforcer deliveries were replaced by non-food stimuli, such as the magazine light. Each stimulus was correlated with the reward ratio that was associated with the alternative that produced the stimulus. Like other frequently changing experiments, the subjects showed large shifts in preference after reinforcers. In addition, subjects’ reactions to the non-food stimuli were similar to the reinforcers. When the correlation was positive (i.e., the stimulus predicted that responding in that location was more likely to produce food in the future), subjects’ preference would shift towards that alternative. Subjects’ preference shifted away from the alternative producing the non-food stimuli when the correlation was negative (i.e., the stimulus predicted that responding in the other location was more likely to produce food in the future).

In an equivalent steady-state procedure, conversely, Boutros, Davison, and Elliffe (2009) found that preference after a non-food stimulus was not influenced by the stimulus-food correlation. This makes sense because information provided by the non-food signal about the current ratio is much less informative and so exerts less control. More evidence that organisms only react to informative signals comes from Boutros, Davison, and Elliffe (2011). They arranged a steady-state procedure where non-food stimuli indicated which alternative was likely to deliver the next reinforcer, instead of indicating the current reinforcer ratio. The environment was in a steady state, so uncertainty about the reinforcer ratio was assumed to be low, but the location of the next reinforcer was probabilistically determined and therefore always uncertain. Reinforcers delivered in Boutros et al.’s procedure did not reduce
reinforcer ratio uncertainty any further, and did not reduce the uncertainty of the next reinforcer location because that location was independently determined. Signalling the location of the next reinforcer also provided little additional information about the reinforcer ratio, but it did reduce uncertainty about the next location and so the stimuli exerted control over the subjects’ responding. When doing that Boutros et al. (2011) observed that after a non-reinforcer stimulus, preference shifted towards the location that was likely to produce the next reinforcer. Therefore, they demonstrated that subjects’ reactions to signals were determined by the information provided by the signals.

Evidence suggests that creating a steady-state environment removes most of the signalling effect of reinforcers. Therefore examining preference data from that type of environment and comparing it to data from a frequently changing environment provides further answers as to why organisms behave differently when in these different environments. The results that were chosen for comparison were from four-alternative choice procedures by Elliffe and Davison (2010), and Davison, Krägeloh, Fraser, and Breier (2007). Both procedures used four variable-interval (VI) schedules arranged concurrently in a frequently changing 27:9:3:1 reinforcement ratio. The location of each schedule changed after a component of several reinforcers. With pigeons, Elliffe and Davison’s components comprised 10 reinforcers, so that each schedule’s key location changed after 10 reinforcers. With rats, Davison et al.’s components were 5, 10, or 20 reinforcers long, in different conditions. These experiments were chosen because choice with more than two alternatives has not been explored as extensively as choice with two alternatives, which may be limiting the current understanding of complex preference. For example, in two-alternative choice procedures, a subject chooses when to switch to responding to the other alternative. When there are more than two alternatives, the subject must choose when to switch and which of the other alternatives to switch to. Using data from procedures with additional alternatives means that
two comparisons can be made in this chapter; frequently changing compared with steady-state environments, and two-alternative choice, compared with more than two alternatives.

The results from both Elliffe and Davison’s (2010), and Davison et al.’s (2007), experiments were similar. Both compared preference between pairs of alternatives with the same relative rate of reward but different absolute rates. For example, Davison et al. compared preference between a VI 40 s: VI 120 s pair to that of a VI 360 s: VI 1080 s pair, both of these pairs had a 1:3 reinforcer ratio but produced food at different rates. The main finding in both experiments was that preference between any pair of alternatives was influenced by the absolute rate of reinforcers produced by those alternatives. The more reinforcers produced by the pair, the more preference for the richer of the two alternatives, (i.e., there was more preference for Davison et al.’s VI 40 s over the VI 120 s than for the VI 360 s over the VI 1080 s, although in each case the reinforcer ratio was the same). This is known as a reinforcer-rate effect (Alsop & Elliffe, 1988; Elliffe & Alsop, 1996). Further analyses by Elliffe and Davison showed some of the local preference patterns. They found that preference shifted towards the location of the just-productive alternative. In this way, preference was similar to that found in many two-alternative experiments. Additionally, they analysed the subjects’ switching patterns by looking at which alternatives subjects were most likely to switch to after responding on another alternative. They found that the likelihood of switching to an alternative was related to the overall rate of reinforcement on that alternative. Subjects were most likely to switch to the richest alternative, and second most likely to switch to the second richest alternative, and so on.

The findings that preference differed between pairs with the same ratio of reinforcement were important because of their implications for the Generalized Matching Law (GML; Baum, 1974), which is a dominant model of choice. In order for the GML to be an appropriate model of why changes in preference occur, certain assumptions must be
satisfied. Two of the assumptions are “Relative homogeneity: Relative preference is not affected by proportional increases or decreases in all reinforcement rates” (Prelec, 1984, p. 102) and “Relative independence: Relative preference is not affected by the rate of reinforcement for a third alternative” (Prelec, 1984, p.102). These assumptions state that the relative reinforcer rates (or equivalently, the reinforcer ratio) on a pair of alternatives is the sole determinant of preference between those alternatives. If the reinforcer ratio between two alternatives remains constant, then neither increasing the overall rate of reinforcers of those alternatives, nor changing reinforcer rates on other alternatives, should cause a change in preference between those two alternatives. This is known as the constant-ratio rule (Luce, 1959). According to this rule, if two or more pairs of alternatives have the same reinforcer ratio, then the amount of preference shown for the richer alternative of each pair should be the same for all pairs, regardless of the overall rates of reinforcers produced by each pair. Because the constant-ratio rule is the combination of two assumptions of the GML, it becomes an indicator of when it is inappropriate to model data using the GML. If the rule is violated, then it is inappropriate to use the GML. The results of both Davison et al. and Elliffe and Davison found violations of the constant-ratio rule.

It should also be noted that the GML is a model that assumes that all reinforcers carry equal weight in determining overall preference. The effect of a reinforcer on one alternative causes the same mathematical increase in preference as a reinforcer on another alternative. The model assumes a strengthening-like process. The contingencies arranged in both Elliffe and Davison’s (2010), and Davison et al.’s (2007), experiments are those that promote the signalling effects of reinforcement (i.e., frequent change in reinforcer ratios). Perhaps this signalling effect was a reason why the results violated the assumptions of the model. Consequently, if the signalling effect is reduced by arranging a steady-state environment, then the assumptions might hold. Therefore in the present experiment, the 27:9:3:1 reinforcer
ratio used by Elliffe and Davison and by Davison et al. was replicated, but the locations of the VI schedules were kept constant for 50 consecutive sessions per condition instead of changing them several times per session.
3.2. Method

3.2.1 Subjects

The same six pigeons that served as subjects in Elliffe and Davison (2010) were used. They were numbered 81 to 86 and were housed individually. After experimental sessions, subjects were fed mixed grain to maintain them at 85% ± 15 g of their free-feeding weights. Water and grit were available at all times.

3.2.2 Apparatus

The subjects were housed in the experimental chambers. These chambers were 380 mm high, 380 mm wide and 380 mm deep. The back and side walls of the chamber consisted of sheet metal. The front wall consisted of metal bars, as did the floor and top of the chamber. Within each cage were two wooden perches which intersected at 90°. One perch was 95 mm from and parallel to the right wall, the other was 95 mm from and parallel to the front wall.

Situated on the right wall, 215 mm above the wooden perches were four response key alternatives. These keys were plastic discs that measured 20 mm in diameter and they were spaced 65 mm apart. Keys were lit with a white light during experimental sessions and pecks of a force of 0.1 N or more were required for a responses to be registered. Situated directly below the two centre keys was the magazine aperture. It measured 50 mm by 50 mm, and was 125 mm below the keys. During food deliveries all key lights were extinguished and the hopper containing wheat was raised for 3 s. Keylights were turned off and the magazine aperture was lit while the hopper was raised. A computer running MED-PC® software controlled all experimental events, and recorded the times that they occurred, to 10 ms resolution. The room lights were switched on at 12:30 am and off at 4:00 pm. Sessions began daily at 1 am and were run in the order of the subject numbers; no personnel entered the room while sessions were running.
3.2.3 Procedure

All subjects had previous experience with four-alternative procedures in Elliffe and Davison (2010) so no pre-training was required. Experimental sessions began with all four alternatives being lit and the programmed exponential VI 20-s schedule being activated. Every second the programmed schedule had a .05 probability of arranging a food delivery. When the program arranged a reinforcer delivery it would assign it to one of the alternatives with a probability dependent on the experimental condition (Table 3.1). There were nine conditions that were run for 50 sessions each. The first eight conditions used a different permutation of the 27:9:3:1 ratio. The assigned probabilities used in the first eight conditions of this experiment in combination with the VI 20-s base schedule, equated to four exponential VI schedules of VI 29.63 s, VI 88.89 s, VI 266.67 s, and VI 800 s, which formed a 27:9:3:1 ratio. The ninth condition arranged a 1:1:1:1 ratio, and equated to four VI 80 s schedules. When a reinforcer was arranged, the program would not arrange additional deliveries for any alternative until the assigned reinforcer had been delivered (dependent scheduling, Stubbs & Pliskoff, 1969). To reduce the potential impact of concurrent superstition (Catania & Cutts, 1963), food was delivered after a response to the assigned alternative, but only if the previous response was also to the same alternative. Otherwise, the food was delivered after the subject made two successive responses to the assigned alternative. This means that there was a 1 response penalty for switching alternatives, also known as a Changeover Ratio 1 (COR 1).
Table 3.1

The probability of reinforcer allocation for each alternative in each condition.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Key 1</th>
<th>Key 2</th>
<th>Key 3</th>
<th>Key 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.025</td>
<td>.225</td>
<td>.675</td>
<td>.075</td>
</tr>
<tr>
<td>2</td>
<td>.075</td>
<td>.675</td>
<td>.225</td>
<td>.025</td>
</tr>
<tr>
<td>3</td>
<td>.675</td>
<td>.025</td>
<td>.075</td>
<td>.225</td>
</tr>
<tr>
<td>4</td>
<td>.225</td>
<td>.075</td>
<td>.025</td>
<td>.675</td>
</tr>
<tr>
<td>5</td>
<td>.675</td>
<td>.025</td>
<td>.225</td>
<td>.075</td>
</tr>
<tr>
<td>6</td>
<td>.225</td>
<td>.075</td>
<td>.675</td>
<td>.025</td>
</tr>
<tr>
<td>7</td>
<td>.025</td>
<td>.225</td>
<td>.075</td>
<td>.675</td>
</tr>
<tr>
<td>8</td>
<td>.075</td>
<td>.675</td>
<td>.025</td>
<td>.225</td>
</tr>
<tr>
<td>9</td>
<td>.25</td>
<td>.25</td>
<td>.25</td>
<td>.25</td>
</tr>
</tbody>
</table>
3.3. Results

3.3.1 Stability

The data were collated so that responses to the richest alternative of each condition were combined; responses to the second richest of each condition were combined, and so forth. This process merged the first 8 conditions into a single 27:9:3:1 ratio dataset. This eliminated the effect of any bias towards individual key locations as each reinforcer rate was assigned to each key twice, and is the same method that was used by Davison et al. (2007), and Elliffe and Davison (2010). In these results, responses to the alternatives in the 27:9:3:1 reinforcer ratio will be identified as B$_{27}$, B$_9$, B$_3$, and B$_1$, respectively. Reinforcers delivered from those alternatives will be identified as R$_{27}$, R$_9$, R$_3$, and R$_1$, respectively.

Both Davison et al. (2007) and Elliffe and Davison (2010) reported that response allocation had not stabilized within 10 – 20 reinforcers after a reinforcer-ratio change. As the main aim of the present experiment was to produce a steady-state replication of their experiments, it was important to determine how many reinforcers were required before the response allocation stabilized. Unfortunately there were insufficient data to produce an analysis of changes in overall response allocations after each reinforcer in a condition; therefore, an analysis of each session was conducted to determine stability. Figure 3.1 shows the response allocations for Sessions 1 – 50 of all 8 conditions combined, for each subject in the present experiment. In general, the highest proportion of responding was allocated to B$_{27}$, the second highest to B$_9$, and lowest proportion was allocated to B$_1$.

The allocation of responses to alternatives changed progressively across sessions before appearing to stabilize. To determine how long each subjects’ preference took to stabilize, the proportion of responses from the last 5 sessions of data (which were assumed to
be stable) were averaged for each subject and subtracted from the proportion of responses
from each
Figure 3.2 shows the calculated deviations from stability for all subjects. The calculated values show how much a single session of responding deviated from the mean of the last five sessions; responding was assumed to be stable when the value fluctuated around 0. Visual estimations from Figure 3.2 suggest that the responding had stabilized by Session 25 for all subjects, except possibly Pigeon 83, whose deviations reached 0 by Session 25, but then deviated briefly, returning to zero after 10 more sessions. Data from Sessions 26 – 50 were considered stable and were used for steady-state analyses. Therefore we are certain that the comparison between our data and that of both Davison et al.’s (2007), and Elliffe and Davison’s (2010) data is a comparison between responding in steady-state and frequently changing environments. Additionally, it appears that the time taken for responding to reach stability differed for each alternative, with responses to B27 taking more sessions to stabilize than the other alternatives. Responses to B27 also had the greatest deviation immediately after a condition change. The large deviation was not surprising as the physical location of any reinforcer probability could not be the same for two conditions in a row, and the difference in probabilities between B27 and any other alternative was greater than the difference between another other pair of alternatives. Therefore at the current location of B27, there would be a large difference between the proportion of responses allocated at stability of the previous condition and proportion of responses allocated in the current condition. Subtracting the new stabilization point from the previous one resulted in the large negative value that was observed.

3.3.2 Constant-Ratio Rule

Results were analysed in terms of the generalized matching law, in which the log response ratios obtained from responses to each pair of alternatives are plotted against the obtained reinforcer ratios. Sensitivity (slope) values ranged from 0.61 to 0.73 with an average of 0.69; that is, all subjects showed undermatching. Pairs of alternatives with the same
Figure 3.2

Sessional deviations in response allocation from the mean response allocations of the last 5 sessions.
reinforcer ratio, but different absolute reinforcer rates, were compared to determine whether
the data conformed to the constant-ratio rule. Davison et al. (2007) and Elliffe and Davison
(2010) reported that subjects showed greater preference for the richer alternative in the B_{27}:B_{3}
pair than for the richer alternative in the B_{9}:B_{1} pair. Only 2 of 6 subjects clearly showed more
preference for the richer alternative in the B_{27}:B_{3} pair. A Wilcoxon matched-pairs signed-ranks test confirmed that there was no systematic reinforcer-rate effect in the two 9:1 ratio
pairs (p = .345). An equivalent analysis was conducted on the three 3:1 ratio pairs. A
Friedman ANOVA by ranks showed that there was also no systematic pattern among the log
response ratios obtained from responses to the B_{27}:B_{9}, B_{9}:B_{3}, and B_{3}:B_{1} pairs (p = .223).
Overall, the results of the present experiment were consistent with the constant-ratio rule.

3.3.3 Local Analysis

To investigate how subjects responded immediately after reinforcement, the following
analysis was conducted. Data were collected into data bins based on time since the last
reinforcer. Bins were 3-s long, meaning that responses made within the first 3 s following
reinforcement were collected into the first bin; responses made from 3 s to 6 s were collected
into the second bin, and so forth. The proportion of responses made to each alternative was
calculated for each bin and then plotted as a function of time. Proportions calculated from the
0- – 3-s bin were plotted at 0 s, proportions from the 3- – 6-s were plotted at 3 s, and so forth.
This produced preference pulse graphs (Davison & Baum, 2002) using proportions as a
measure of response allocation (Figures 3.3 – 3.8). Data were converted to proportions of
responses and plotted as proportions against time bins. Proportions were used instead of the
ratio measures used by Elliffe and Davison (2010) as this avoided the problem where zero
responses to an alternative would lead to an undefined value instead of a proportion of 0.
Figure 3.3

Preference pulses showing the proportion of responses allocated by Subject 81 to each alternative as time since the last reinforcer increases. The pulses are separated by which alternative produced the last alternative.
Figure 3.4

Preference pulses showing the proportion of responses allocated by Subject 82 to each alternative as time since the last reinforcer increases. The pulses are separated by which alternative produced the last alternative.
Figure 3.5

Preference pulses showing the proportion of responses allocated by Subject 83 to each alternative as time since the last reinforcer increases. The pulses are separated by which alternative produced the last alternative.
Figure 3.6

Preference pulses showing the proportion of responses allocated by Subject 84 to each alternative as time since the last reinforcer increases. The pulses are separated by which alternative produced the last alternative.
Figure 3.7

Preference pulses showing the proportion of responses allocated by Subject 85 to each alternative as time since the last reinforcer increases. The pulses are separated by which alternative produced the last alternative.
Figure 3.8
Preference pulses showing the proportion of responses allocated by Subject 86 to each alternative as time since the last reinforcer increases. The pulses are separated by which alternative produced the last alternative.
As can be seen in Figures 3.3 – 3.8, immediately after a reinforcer the subjects’
patterns of preference after a reinforcer from any alternative were similar for all subjects.
This finding is similar to when all three alternatives were lit in Set 4 of Chapter 2. In general,
the most preference was for B_{27}, then for B_{9}, and least for B_{1}. As time since reinforcement
increased, preference for B_{27} decreased. However, the subjects’ preference for B_{27} remained
greater than all other alternatives at all times since reinforcement. Visual inspection of each
graph shows that subjects increased their preference for whichever alternative was the just-
productive alternative compared with when reinforcers were delivered from other alternatives. Although the subjects’ preference was only slightly affected by which alternative produced the reinforcer. The only exception to this was preference from Pigeons 83 and 86 after an R_{o} reinforcer. These subjects showed greatest preference for B_{9} during the first 3 s after reinforcement, before shifting preference to B_{27}. Overall, the response patterns obtained were different from those obtained by Elliffe and Davison (2010), and additionally, these patterns differed from Landon, Davison, Elliffe (2002) who reported pulses from a two-
alternative steady-state procedure. Results from both of those experiments showed that preference immediately after a reinforcer was greatest for the just-productive alternative.

3.3.4 Effect of CODs

It was of interest to determine whether the differences between the preference pulses
presented above, and those reported by Elliffe and Davison (2010), were caused by the
absence of a 2 s changeover delay (COD 2) in the current experiment that was present in
Elliffe and Davison’s procedure. The following data come from two experimental conditions
that were conducted after those reported by Elliffe and Davison, but before those of the
present experiment. Apart from two exceptions, the procedure used was identical to that of
Elliffe and Davison. The first exception was that a base VI 60 s schedule was used for both
conditions, instead of a VI 30 s schedule. The second exception was that a COR 1, identical
to the one used in the present experiment, was arranged in the second condition, while the first condition used a COD 2 s. The preference pulse analysis used was identical to the process described above. The data analysed and shown in Figure 3.9, are from responses immediately after the 9th reinforcer since a VI location change. The average values across subjects have been presented.

Comparing between conditions shows that immediately after reinforcement, subjects preferred the just-productive alternative more than they preferred the other alternatives. However, it appears that preference was shown for the just-productive alternative when a 2 s COD is operating, suggesting that the changeover penalty used does affect local response data, which is consistent with the results of Chapter 2.

3.3.5 Changeovers

The probability of switching to each alternative based on the location of each subject’s previous response can be seen in Figures 3.10 – 3.15. The general pattern shows that probability of switching to a certain alternative was determined by the relative reinforcer ratio. This is consistent with Elliffe and Davison’s (2010) findings. Subjects were more likely to switch to the richer of the available alternatives than to the others. For example, when switching from responding on B27, subjects were more likely to switch to B9 than to B3, and least likely to switch to B1. This pattern can be clearly seen after switches from B27 and B9. Data from B3 and B1 were not as clear. It should be noted that there were fewer responses to those alternatives overall, which leads to more noise in the data.
Previously unreported preference pulses showing the average proportion of responses allocated by subjects to each alternative as time since the last reinforcer increases when a COD 2 s was operating (left) compared to when a COR 1 was operating (right). The pulses are separated by which alternative produced the last alternative.

Figure 3.9
Figure 3.10

Bar graphs showing the switching patterns of 81 from Conditions 1-8.
Figure 3.11

Bar graphs showing the switching patterns of 82 from Conditions 1-8.
Figure 3.12

Bar graphs showing the switching patterns of 83 from Conditions 1-8.
Bar graphs showing the switching patterns of 84 from Conditions 1-8.
Figure 3.14

Bar graphs showing the switching patterns of 85 from Conditions 1-8.
Figure 3.15

Bar graphs showing the switching patterns of 86 from Conditions 1-8.
Results from Condition 9 (1:1:1:1 reinforcer ratio) show a different pattern (Figures 3.16-3.21). If the subjects were switching according to probabilities of reinforcement, they should have been just as likely to switch to any alternative at any point. Instead it appears that the subjects were following a response pattern. For example, after responding at Alternative 1 (left most alternative), Subject 86 was most likely to switch to Alternative 2 (centre left alternative). After that it was most likely to switch to Alternative 3 (centre right alternative) then Alternative 4 (right most alternative), before returning to Alternative 1. Like a typewriter, Subject 86 was responding from left to right before returning to the left. An equivalent right to left response pattern was produced by Subject 81. Not all subjects produced one-directional switching patterns; no subject showed switching patterns that were related to the reinforcement probabilities.

3.3.6 Dynamic interactions between responses and food

The changeovers and local analyses show an overall pattern of switching between the two richer alternatives while occasionally moving to the leaner alternatives. Responding in this way affected the rate and times that reinforcement was delivered to the subjects. To demonstrate this, the time between arrangement and collection of reinforcement was calculated based on which alternative the reinforcer was arranged on. Results for the last 25 sessions of Conditions 1 – 8 are shown in Table 3.2. The results show that the median delay between arrangement and collection time was shortest for the richer alternative, followed by the second richest, and the delay was longest on the leanest alternative. This means that the uneven responses rates influenced the time at which reinforcers were obtained on each alternative. Reinforcers arranged on the richer alternatives were collected quickly, whereas reinforcers on the leanest alternative remained uncollected for longer.
Bar graphs showing the switching patterns of 81 from Condition 9.

Figure 3.16
Figure 3.17

Bar graphs showing the switching patterns of 82 from Condition 9.
Figure 3.18

Bar graphs showing the switching patterns of 83 from Condition 9.
Figure 3.19

Bar graphs showing the switching patterns of 84 from Condition 9.
Figure 3.20

Bar graphs showing the switching patterns of 85 from Condition 9.
Bar graphs showing the switching patterns of 86 from Condition 9.
Table 3.2
The median collection times for each alternative in Conditions 1-8

<table>
<thead>
<tr>
<th></th>
<th>Alt 27</th>
<th>Alt 9</th>
<th>Alt 3</th>
<th>Alt 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>81</td>
<td>3.39</td>
<td>6.28</td>
<td>13.10</td>
<td>43.00</td>
</tr>
<tr>
<td>82</td>
<td>1.29</td>
<td>4.01</td>
<td>9.44</td>
<td>22.48</td>
</tr>
<tr>
<td>83</td>
<td>2.08</td>
<td>7.57</td>
<td>11.79</td>
<td>25.93</td>
</tr>
<tr>
<td>84</td>
<td>1.80</td>
<td>4.80</td>
<td>13.09</td>
<td>29.37</td>
</tr>
<tr>
<td>85</td>
<td>2.37</td>
<td>6.95</td>
<td>13.47</td>
<td>33.88</td>
</tr>
<tr>
<td>86</td>
<td>1.44</td>
<td>4.25</td>
<td>9.53</td>
<td>24.90</td>
</tr>
</tbody>
</table>
3.4. Discussion

The present experiment provided further evidence that behavioural patterns in steady-state environments differ to those in frequently changing environments at both local and global levels of analysis. Local preference patterns indicated that subjects reacted differently following reinforcers in the current four-alternative steady-state environment than they did in Elliffe and Davison’s (2010) four-alternative frequently-changing procedure. The local patterns described here were also different from those of Landon et al.’s (2002) steady-state two-alternative procedure. In addition, the global response patterns in the present experiment did not show the overall reinforcer rate effect that was detected by both Davison et al. (2007) and Elliffe and Davison. With that said, the present experiment also had findings that were consistent with Elliffe and Davison. The likelihood of a subject switching to an alternative was related to the rate of reinforcement scheduled on that alternative. However, this was only true when the reinforcer rates differed on each alternative.

The local preference patterns in the present experiment differ from those observed by Elliffe and Davison’s (2010) frequently changing four-alternative experiment, and also differ from preference patterns that were observed in related two-alternative choice procedure, both frequently changing (Davison & Baum, 2002, 2003; Landon, Davison, & Elliffe, 2003a) and steady-state (Landon et al., 2002, 2003b; Menlove, 1975). Previous experiments that have arranged concurrent VI VI choice lead to differential preference based on which alternative produced the reinforcer, with the subjects showing increased preference for the just-productive alternative immediately after reinforcement. In the present experiment, the post-reinforcer preference did not appear to differentiate according to the previous reinforcer. Rather, it was always in favour of the richest alternative (Figures 3.3 – 3.8).
The experiment was designed to minimise the signalling properties of reinforcement by making the signal relatively uninformative. This was done to examine preference patterns when there was little or no effect of reinforcer-based signals. One possibility was that the preference would conform to patterns based on strengthening effects of reinforcement. This would be shown by enhanced preference for the just-productive alternative as has been seen in other procedures. However, the obtained preference patterns were inconsistent with that of a strengthening effect because preference after each alternative was similar, with little preference shifted towards the location of the just-productive alternative.

Preference pulses that are not related to the location of the previous reinforcer have been observed before in other two-alternative procedures (Cowie et al., 2011; Miranda-Dukoski et al., 2014). However those procedures did not arrange a typical concurrent VI VI procedure. For example, subjects preferred the locally richer alternative immediately after reinforcement when Cowie et al. arranged contingencies where reinforcers occurring soon after the previous reinforcer were more likely to occur either on the just-productive alternative or always on a specific alternative. Subjects also preferred the locally richer alternative immediately after reinforcement when Miranda-Dukoski et al. manipulated the probabilities of reinforcement at every second so that if a reinforcer occurred immediately, it was highly likely to come from one of the alternatives. In both of these procedures the local probability of reinforcement differed over time and was higher for one alternative over the other immediately after the previous reinforcer. In typical concurrent VI VI procedures that use exponential reinforcer rates, the local probability of reinforcement at each second is invariant. While it is true that the richest alternative is the most likely to produce a reinforcer immediately after reinforcement, it is also the most likely to produce the reinforcer at all points of time. If subjects were responding according to the local probability of
reinforcement, then they should show the same level of preference for the rich alternative at all times. Instead preference levels shift over time since reinforcement.

Other concurrent VI VI steady-state experiments have shown that preference after a reinforcer was shifted towards the location of the last reinforcer. It has been suggested that this is due to the effects of CODs (Boutros et al., 2009, 2011). In some experiments the changeover contingencies are programmed so that it is only possible to receive another reinforcer immediately after the previous reinforcer on the just-productive alternative, meaning that the just-productive alternative is locally richer. A similar effect was shown observed in Set 2 of Experiment. However there is evidence from Set 3 of Experiment 1, and analyses of previously unreported conditions from Elliffe and Davison (2010), suggesting that a COD is not required for this pattern to exist. With that said, both the dataset from Experiment 1, and Elliffe and Davison, occur after the subjects have been exposed to CODs. This is important as both Krägeloh and Davison (2003), and Shull and Pliskoff (1967) found that subject’s behaviour was affected by a COD even after it has been removed. However, the subjects in the present experiment also had experience with CODs and still produced non-differential preference patterns.

Even though the experimental design should have minimised the signalling effects of reinforcement, one question that is of importance is whether these results do contain some form of signalling. Assuming that the subjects had sufficient experience with the reward contingencies, there are two reasons why it is unlikely that the location of previous reinforcement was signalling the likely location of the next reinforcer. The first reason is that the location of the just-productive alternative was not correlated with the location of the next reinforcer. Receiving a reinforcer from one location did not indicate that the next reinforcer was likely to come from a specific other key, unlike experiments from Krägeloh et al. (2005), Boutros et al. (2011), and Set 6 of Experiment 1. The second reason is that subjects show the
same reaction to all rewards. If subjects were reacting to reinforcer-based signals, then reinforcers from all alternatives were signalling that they should go to the same alternative immediately after reinforcement.

If there was any signalling effect of reinforcers in this steady-state environment it was likely to be about signalling changes in conditions; however that would require multiple reinforcers to signal. For example, .675 of all reinforcers come from the rich alternative meaning that .325 of all reinforcers do not come from the richest alternative. Receiving an occasional reinforcer from a non-rich alternative is unlikely to signal that the conditions have changed. However each consecutive reinforcer from a non-rich alternative is increasingly less likely. The probability of two being delivered is ~.106; the probability of three is ~.034, four is ~ 0.011, and so forth. The more consecutive rewards from the perceived non-rich alternatives, the more likely that the location of the rich-alternative has changed. If that is how the reinforcers were acting as signals, then those reinforcers still were not providing specific information about the location of the next reinforcer other than that location is likely to be the same as the rich location.

It appears that subjects were not responding via a strengthening based response pattern, and were responding to the likely location of the next reinforcer without using specific information from the last reinforcer. However the subjects were showing systematic responding, returning to the richest alternative after all reinforcers. One possible reason why subjects were doing this was that they were following an exploitive strategy (see Racey et al., 2011). This strategy is where subjects repeat the response that is most likely to be reinforced. This contrasts with an explorative strategy where subjects vary their response patterns to gain and react to information about the environment; this equates to finding and following a signal. There is evidence that shows that subjects will start with an explorative strategy and then switch to an exploitative strategy (Racey et al., 2011). This would suggest that subjects
in Elliffe and Davison (2010) and Davison et al. (2007) were constantly exploring, while our subjects were exploiting. However this would also suggest that Landon et al.’s (2002) subjects were also using an explorative strategy even after extended exposure to the same contingencies, which seems unlikely.

An interesting consequence of providing subjects with more than two alternatives to choose from is seen in the switching data. In two-alternative procedures a switch from one alternative equals a switch to the other alternative. When there are more than two alternatives a switch from an alternative requires an additional choice of where to switch to. Switching in the combined Conditions 1 – 8 data set was similar to that observed by Elliffe and Davison (2010). The probability of switching to an alternative was related to the rate of reinforcement produced by that alternative (Figures 3.10 – 3.15). This meant that subjects were most likely to switch to the richest alternative after responding to any of the others. If they did not switch to the richest (or were switching from the richest), then they were likely to switch to the second richest, and so forth. A special pattern was observed in Condition 9, where all alternatives produce reinforcers at the same rate. If subjects used the same method of switching from the previous conditions, the expected probability of switching to every alternative would be equal. Instead in that condition, subjects appeared to be following a simple sequence of responding to alternatives in a specific order. Subjects 81 and 86 took this to the extreme by responding in a line before starting the line again (Figures 3.16 & 3.21).

Another interesting consequence of providing subjects with more than two alternatives is seen in the interaction between the response patterns and overall reinforcer rates. As with most VI schedules, a reinforcer arranged on one alternative will stay there until collection. If the subject was responding to the alternative that the reinforcer was arranged on at the time of arrangement, then the reinforcer was delivered immediately. If the subject was responding to a different alternative, then the reinforcer was not delivered until after the
switch (and any switching penalty in operation). In two-alternative procedures, the switch occurs as soon as the response run ends on the alternative that the reinforcer was not arranged on. When more than two alternatives are present, it is possible that the subject switches to another alternative that does not have the reinforcer arranged on it. In that situation the next reinforcer will be delivered after that next response run, but only if the subject switches to the alternative with the arranged reward. The data from the present experiment shows that the likelihood of switching to an alternative depended on rate of reinforcement, meaning that subjects are most likely to switch to the richest alternative. Therefore any uncollected reinforcers on the rich alternative were likely to be collected soon after they were arranged, as shown in Table 3.2. In contrast, subjects were less likely to switch to the leanest alternative, meaning that uncollected reinforcers on that alternative remained uncollected for longer as subjects switched between the other alternatives. These different delays to reinforcer collection have the effect of spreading the distribution of obtained reinforcers, with more rich alternative reinforcers being obtained earlier after the previous reinforcer, and more leanest alternative reinforcers being obtained much later, and more reinforcers from the second richest and second leanest being obtained at times in between.

Perhaps this interaction between switching patterns and obtained reinforcer times caused the subjects to perceive that certain alternatives were locally richer at different times since reinforcement, as in Cowie et al.’s (2011) and Miranda-Dukoski et al.’s (2014) procedures. However unlike their procedures, the differences in local reinforcer distributions were caused by the subjects’ choices and were not arranged by the experimenter. With more reinforcers from the richest alternative being obtained at shorter times since the previous reinforcer, that alternative would appear locally richer than the others immediately after reinforcement. If this is what is happening, then observing that subjects show preference for
the richest alternative immediately after reinforcement in the present experiment is consistent with both Cowie et al.’s and Miranda-Dukoski et al.’s findings.

The results of the present experiment were consistent with the constant-ratio rule and, therefore, at odds with the results of Elliffe and Davison (2010) and Davison et al. (2007). Their results violated the constant-ratio rule due to an overall reinforcer rate effect that was absent in the present experiment. While the local preference patterns can be interpreted in several ways, the present/absence of the overall reinforcer-rate can be explained by local preference patterns. Elliffe and Davison found that responding differed based on the location of the last reinforcer. When response patterns immediately after reinforcement differ according to the location producing the reinforcer, the absolute number of reinforcers from each location will alter the overall response patterns. This produces an overall reinforcer-rate effect. However, in the present experiment the individual reinforcers did not influence patterns of preference. This prevented the overall reinforcer-rate effect because the absolute number of reinforcers did not alter the overall response patterns.

In summary, Experiment 2 has shown that large differences can occur between results of frequently-changing environments and steady-state environments. When limited information was produced by the signalling effects of each reinforcer, subjects shifted their behaviour to a repetitive strategy that favoured the richest alternative, instead of shifting toward the just-productive location as would be expected by a strengthening effect. Due to this pattern, an overall reinforcer rate effect found in other experiments was not present. In addition the Experiment 2 demonstrated how providing subjects with more than two alternatives in a choice procedure leads to complications that are not present in two-alternative procedures.
Chapter 4. Simulations

4.1. Introduction

McLean, Grace, Pitts, and Hughes (2014) suggested that preference pulse analyses (Davison & Baum, 2002) used by recent research might be misleading, and show more preference for some alternatives than was actually occurring. McLean et al. used simulations to show that a large proportion of the pattern created by such analyses can be due to the way data are organised and not necessarily due to the reinforcer at all. This finding is directly relevant to this thesis for two reasons. The first is that this thesis focuses on how reinforcers cause local preference patterns and the data organisation effects are one possible cause. The second is that the two previous chapters and the majority of literature that has been cited by this thesis have used preference pulses extensively as a method of determining how subjects react to reinforcers. If there are large data organisational effects in those preference pulses, then the conclusions drawn from those analysis may need to be altered.

The purpose of the original simulation by McLean et al. (2014) was to show that response patterns normally attributed to reinforcement effects can be produced without reinforcers. Instead they can be created by the way that data are processed to produce a preference pulse graph. McLean et al. demonstrated that by creating random distributions of response runs (or bouts), assigning reinforcers to those responses according to a variable interval (VI) schedule, and then analysing those data using preference pulses, preference initially after the reinforcer will be towards the just-productive alternative. This preference towards the just-productive alternative cannot be due to reinforcement effects because the behaviour was generated by a random process, not by subjects that are receiving reinforcers. From this it follows that preference towards the just-productive alternative observed by others
(for example, Landon, Davison, & Elliffe, 2002) might not be due entirely to the properties of reinforcement, if at all.

The reason that the simulated data create preference patterns that show preference for the just-productive alternative immediately after reinforcement is due to response runs being interrupted by reinforcers. The simulation treats responding as a series of alternating bouts to each alternative. When a reinforcer is simulated, it occurs during one of those bouts, which splits the bout into a run of responses before, and a run of responses after, the reinforcer. Because preference pulses typically collate responses based on time since reinforcement, responses immediately after the reinforcer become the first responses of the post-reinforcement period, and will be to the just-productive alternative. This means that in the simulation, preference to the just-productive alternative is caused by the end of a response bout and not by any strengthening or signalling effect of reinforcement.

In addition to the simulation, McLean et al. (2014) suggested a correction procedure to determine how much of the preference pattern is due to the reinforcer and how much is due to the analysis itself. This procedure involves taking real response data and replacing the obtained reinforcers with reinforcers that are generated by a computer simulation, using the same method that assigned reinforcers to simulated data. The resulting response patterns are generated by responding with actual reinforcers, but the new preference patterns are independent of the reinforcers that have been used to organise the data. Preference pulses that are created with these simulated reinforcers are dubbed as null hypothesis preference pulses (NHPP) as they show how much processing the real data contributed to the pulse without the reinforcement effect. The NHPPs are then subtracted from actual preference pulses, leaving only the effect of the reinforcer behind. McLean et al. used this method to reanalyse Rodewald, Hughes, and Pitts’ (2010) data and concluded that there was a reinforcement effect, but that effect was much smaller than suggested by the uncorrected analysis.
In the present analysis, McLean et al.’s (2014) simulation has been extended from two to four alternatives. There were several reasons why it was of interest to do this. One reason was that many of the conclusions made during Chapter 3 are based on preference pulses, meaning that the results may have been misinterpreted. The second reason was that to see if the simulation would predict non-differential pulses like those that were observed in Chapter 3, or whether the pulses would be towards the location of previous reinforcement as observed by Elliffe and Davison (2010). A third reason was to see what preference pulses from both Elliffe and Davison, and Chapter 3 looked like after a correction procedure. Due to these reasons McLean et al.’s process was first adapted to simulate four-alternative responding to a 27:9:3:1 reinforcer rate. Second, the correction procedure was performed on both Elliffe and Davison’s data and the data from Chapter 3.
Table 4.1

The $k$ and $\theta$ values used for each simulation in this chapter. The first 6 were used for both methods of switching that were described in the method.

<table>
<thead>
<tr>
<th>Simulation #</th>
<th>$k$</th>
<th>$\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>0.5</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>0.25</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>0.125</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>
4.2. Method

4.2.1 Simulation

The four-alternative simulation followed the method described by McLean et al. (2014) wherever possible. Any changes required to adapt the procedure for additional alternatives have been described below.

Response values were created using python’s numpy.random.gamma function to sample values from a Gamma distribution with varying \( \theta \) and \( k \) values (see Table 4.1). The probability density for the Gamma distribution was

\[
p(x) = x^{k-1} \frac{e^{-x/\theta}}{\theta^k \Gamma(k)},
\]

where \( k \) is the shape and \( \theta \) the scale, and \( r \) is the Gamma distribution.

Lists of 80000 values were created for each of the four alternatives in each simulation. Unlike McLean et al. (2014), our \( \theta \) and \( k \) values were the same on each alternative. This was done to simplify the process. The generated values represented the number of repeated responses to an alternative, after the first response, during a run on that alternative. For example, a value of 1 equalled a run length of 2, the first response and then 1 additional response to that alternative before switching. All response values were converted to run lengths by adding 1 to each value.

The lists of run lengths on all alternatives were then combined to create the response data, with each run length on a list being used in order. Data were recorded as a response code and time code. The time between each response was set as 0.4 s to maintain consistency with McLean et al. (2014). For example, if the first response run was 3 responses to Alternative 1, response data were recorded as one Alternative 1 response at time 0 s, a second response at time 0.4 s and a third response at time 0.8 s. At the end of the response run the...
The simulation would switch to a different alternative. The switching process differed from McLean et al.’s due to the additional number of alternatives used. McLean et al. alternated between the lists of responses, switching from Alternative 1 to Alternative 2, and from Alternative 2 back to Alternative 1. This process was inadequate for four alternatives because observations from Elliffe and Davison (2010), and Chapter 3 show that subjects do not simply alternate between alternatives (except under specific circumstances such as a 1:1:1:1 reinforcer ratio). Instead subjects switch based on the probability of reinforcement available between alternatives. At the end of a response run, the simulation determined probabilistically which alternative the next response run occurred on. This switching mechanic was incorporated into the simulation via two methods. In the first method, the probability of switching to each alternative was the same (.33) at all times. In the second method, the probability of switching to each alternative was equal to the reinforcer ratio of those alternatives. For example, if switching from Alternative 1, the probability of switching to Alternative 2 equalled the probability of reinforcement on Alternative 2 divided by the combined probability of reinforcement on Alternatives 2, 3 and 4.

After the response data were created, a VI schedule was then applied to the simulated responses. This process also differed from that of McLean et al. (2014), their VI schedule only arranged reinforcers at times divisible by 0.4 s, meaning that reinforcers were arranged after a specific number of responses instead of at any time. The VI schedule used in the four-alternative simulation was equivalent to that used in both Elliffe and Davison (2010), and Chapter 3; every second had a .05 probability arranging a reinforcer (i.e., an exponential VI 20-s schedule). Once a reinforcer was arranged, it was probabilistically assigned to an alternative according to a 27:9:3:1 reinforcer ratio. When the schedule determined that a reinforcer was set up on an alternative, the next response to that alternative produced the reinforcer. The reinforcer was inserted after that response and given a time code equal to the
preceding response. This simulation used the convention that Alternative 1 was always the richest alternative, Alternative 2 was second richest, and Alternative 4 was the poorest.

4.2.2 Correction Procedure

The correction procedure described by McLean et al. (2014) was conducted on the data from Conditions 1 – 8 of the experiment from Chapter 3, and on the data reported by Elliffe and Davison (2010). The procedure involved taking the original data, removing the reinforcer codes and subtracting any reinforcement time that occurred, leaving only response and time codes. A simulated VI schedule was then applied to the data, similar to the last stage of McLean et al.’s simulation. However an appropriate changeover penalty was added to the VI schedule used in the procedure to make the schedule as close as possible to the original experiments. This meant that a COD 2 s was used with Elliffe and Davison’s data, and a COR 1 was used with the data from Experiment 2. McLean et al. added an additional constraint which prevented simulated reinforcers from occurring at the same time as the actual reinforcers had occurred, however this made little difference to the four-alternative correction procedure and was ignored in all subsequent corrections.
4.3. Results

4.3.1 Simulation

The simulated responses and reinforcers were used to produce preference pulses. Data were collected into data bins based on time since the last reinforcer. Bins were 3 s long, meaning that responses made within the first 3 s following reinforcement were collected into the first bin; responses made from 3 s to 6 s were collected into the second bin, and so forth. The proportion of responses made to each alternative was calculated for each bin and then plotted as a function of time. Proportions calculated from the 0–3-s bin were plotted at 0 s, proportions from the 3–6-s bin were plotted at 3 s, and so forth. This produced preference pulse graphs (Davison & Baum, 2002) using proportions as a measure of response allocation. Data were converted to proportions of responses and plotted as proportions against time bins.

The first set of analyses compares the two different methods of switching used in the simulation. Figures (4.1 – 4.6) show what happened after each reinforcer when the probability of switching to each alternative was the same (Method 1), Figures (4.7 – 4.12) show what happened when the probability of switching equalled the probability of reinforcement from each alternative that was available to switch to (Method 2). The general finding was that the switching probability controlled response proportions after approximately 10 s. When the probability was the same, the responses stabilized at indifference (.25 for four alternatives), whereas the responses stabilized at the overall probability of reinforcement when the switching probability equated to the probability of reinforcement for each alternative. This trend was observed for all comparisons between the two switching methods suggesting that switching probabilities control the stability values in the simulation. All further analyses used the values from switching based on the reinforcer probabilities as that was observed by Elliffe and Davison (2010) and in Chapter 3.
Figure 4.1
Preference pulses showing the proportion of responses allocated during Simulation 1 using the equal probability switching method.
Figure 4.2

Preference pulses showing the proportion of responses allocated during Simulation 2 using the equal probability switching method.
After 27

Figure 4.3

Preference pulses showing the proportion of responses allocated during Simulation 3 using the
equal probability switching method.
Preference pulses showing the proportion of responses allocated during Simulation 4 using the equal probability switching method.
Figure 4.5

Preference pulses showing the proportion of responses allocated during Simulation 5 using the equal probability switching method.

k = 6 Theta = .25 Equal switching
Figure 4.6

Preference pulses showing the proportion of responses allocated during Simulation 6 using the equal probability switching method.
Figure 4.7
Preference pulses showing the proportion of responses allocated during Simulation 1 using the unequal probability switching method.
Figure 4.8

Preference pulses showing the proportion of responses allocated during Simulation 2 using the unequal probability switching method.
Preference pulses showing the proportion of responses allocated during Simulation 3 using the unequal probability switching method.
Figure 4.10

Preference pulses showing the proportion of responses allocated during Simulation 4 using the unequal probability switching method.
Figure 4.11

Preference pulses showing the proportion of responses allocated during Simulation 5 using the unequal probability switching method.
Figure 4.12

Preference pulses showing the proportion of responses allocated during Simulation 6 using the unequal probability switching method.
The pulses from Figures (4.7 – 4.12) show what happened after reinforcers from each alternative as the k parameter remains constant but the θ parameter decreases. Decreasing θ means that both the variance of the Gamma distribution (θ) and average response run length (k x θ) decrease. In general trend shows that pulses after each reinforcer become more differential as the average response runs become longer. The pulses from Figures 4.13 – 4.17 show what happened after reinforcers from each alternative as the θ parameter remains constant but the k parameter decreases. This means that the average response run length still decreases but the variance of the gamma distribution remains the same. The same pattern occurred as when the run length was increased by changing θ which suggests that the run length was the most important feature determining whether simulated pulses are differential or not. The general trend was that when the response runs were shorter, the differences after each reinforcer were minor, with little if any shift towards the location of the last reinforcer. This was similar to the result obtained in Chapter 3. When the response runs get longer, responses after reinforcement shift towards the location of that produced the previous reinforcer. This was similar to the results obtained by Elliffe and Davison (2010).
Figure 4.13
Preference pulses showing the proportion of responses allocated during Simulation 7 using the unequal probability switching method.
Figure 4.14

Preference pulses showing the proportion of responses allocated during Simulation 8 using the unequal probability switching method.
Figure 4.15

Preference pulses showing the proportion of responses allocated during Simulation 9 using the unequal probability switching method.
After 27

After 9

After 3

After 1

K = 2  Theta = 4 Unequal switching

Figure 4.16

Preference pulses showing the proportion of responses allocated during Simulation 10 using the unequal probability switching method.
After 27

K = 1 Theta = 4 Unequal switching

Preference pulses showing the proportion of responses allocated during Simulation 11 using the unequal probability switching method.
4.3.2 Correction procedure

Corrected preference pulses from both Chapter 3, and Elliffe and Davison (2010) were created. Corrected data from Chapter 3 were combined as described in that Chapter to produce one NHPP for each alternative; those NHPPs appeared similar to simulations with short run lengths. This contrasts with the NHPP from Elliffe and Davison’s data that were more differential, similar to the simulations with longer run lengths.

Response proportions from the NHPP were subtracted from the original pulses to create corrected pulses Figures (4.18 – 4.29). These residual values range from -1 to 1, showing how much of the original pulse is left over from the process. Values greater than 0 mean that more preference was shown for that alternative than would be predicted by the pulse creation process, values below 0 mean less preference was shown than would be predicted. Figures 4.18 – 4.23 show that in the experiment from Chapter 3, the constant preference towards the rich alternative immediately after any reinforcer was greater than predicted by the NHPP. Likewise the preference for the just-productive alternative in Elliffe and Davison’s (2010) results was greater than predicted (Figures 4.24 – 4.29). However after the preference, response proportions stabilized around 0, suggesting that the extended preference patterns can be explained by processes used in the creation of the preference pulses.
Corrected preference pulses for Subject 81's responses during Conditions 1 – 8 of Chapter 3.
Corrected preference pulses for Subject 82’s responses during Conditions 1 – 8 of Chapter 3.
Figure 4.20
Corrected preference pulses for Subject 83’s responses during Conditions 1 – 8 of Chapter 3.
Corrected preference pulses for Subject 84's responses during Conditions 1 - 8 of Chapter 3.
Corrected preference pulses for Subject 85’s responses during Conditions 1 – 8 of Chapter 3.
Figure 4.23
Corrected preference pulses for Subject 86’s responses during Conditions 1 – 8 of Chapter 3.
Figure 4.24

Corrected preference pulses for Subject 81’s responses during Elliffe & Davison (2010).
Corrected preference pulses for Subject 82’s responses during Elliffe & Davison (2010).
After 27

E&D (2010) 83

After 9

Corrected preference pulses for Subject 83’s responses during Elliffe & Davison (2010).
Corrected preference pulses for Subject 84's responses during Elliffe & Davison (2010).

Figure 4.27
Figure 4.28 Corrected preference pulses for Subject 85's responses during Elliffe & Davison (2010).
Corrected preference pulses for Subject 86’s responses during Elliffe & Davison (2010).
4.4. Discussion

Adapting McLean et al.’s (2014) simulation to four alternatives and varying the average run length produced both differential and non-differential preference pulses. When the average run length was short, pulses generated after each reinforcer were very similar. When the average run length was increased, the pulse after each reinforcer depended on which alternative produced that reinforcer. In addition, results from the correction procedure suggest that preference patterns reported by Elliffe and Davison and in Chapter 3 were not simply the result of the analyses.

Average run length was identified as the most important parameter when determining whether the preference pulses differ based on the alternative that produced the immediately preceding reinforcer. This occurred due to the way that reinforcers from the simulation interact with the responses. The longer the average run length, the more likely it is that the simulation will produce more responses to the just-productive alternative immediately after reinforcement. For example, if a particular run length was 10 responses long and the first response produced a reinforcer, then the following 9 responses contribute to the first few time bins of a pulse. Even if the reinforcer was arranged on that alternative during a response run, there are 8 possible responses that the reinforcer could occur after that will still produce responses that contribute to the first time bins after the reinforcer. These responses will be detected as a shift to the just-productive alternative. Only reinforcers occurring on the last response in that run will produce no responses for other alternatives, as the simulation will switch alternatives at that point. However, if the run length was 2 responses, then at most only one response could contribute to the first time bin, meaning that a smaller shift towards the just-productive alternative will occur, if any at all.
It was not surprising that the null hypothesis preference pulses produced from Elliffe and Davison’s (2010) data resembled the longer run length simulations and that pulses from the data from Chapter 3 resembled the short run length simulations. Elliffe and Davison’s experiment imposed a stricter changeover penalty than Chapter 3. Previous studies have shown that increasing the penalty increases the average run length on an alternative (Allison & Lloyd, 1971; Pliskoff, 1971). This means that it was more likely that there would be more responses in a run after the reinforcer, as in the simulation these responses would contribute to the first few bins of the pulse. However, the changeover penalty in the procedure also limited the number of responses that contributed to the post-reinforcement preference as the penalty prevented the first response (and maybe more in Elliffe and Davison’s experiment) from producing the reinforcer. It appears that the increased run length produced by the stricter penalty increased the run length by more than was required to overcome the penalty.

The corrected pulses show that the conclusions made in the discussion of Chapter 3 were valid. The pulses show that subjects in that experiment were responding to the rich alternative immediately after reinforcement more than would be expected by simply processing the data into preference pulses (Figures 4.18 – 4.23). In addition, the subjects were also shifting towards the just-productive alternative in Elliffe and Davison’s (2010) experiment more than would be expected (Figures 4.24 – 4.29). An interesting aspect of the corrected pulses was what happened after the initial preference shift. The pulses suggest that although the initial shift was caused by reinforcement effects, the rest of the pulse pattern was mostly due to the data processing. In other words, the reinforcer effects were short term, the rest of the response pattern was created by the switching pattern that occurred due to the overall reinforcement ratios.

This chapter has focused on the implication of McLean et al.’s (2014) correction procedure for the pulses reported in Chapter 3, but has not attempted to correct the pulses
reported in Chapter 2. This was deliberate as the procedure used in that chapter was atypical and would cause issues for the correction procedure. In several of the conditions, the alternative that produced the last reinforcer became inactive during the next inter-reinforcer interval. This means that if a simulated VI schedule was applied to the real responses, there would be cases where the VI schedule produced a reinforcer but the reinforced alternative was still available in the real data and subjects had responded to it. The reverse of that would be when the VI schedule arranged a reinforcer on an alternative that was not available at that time in the real data. It would not be appropriate to limit the simulated VI to data that occurred between changes to alternatives as that change only occurred after reinforcement, meaning that the preference patterns would be virtually identical to the real preference pulses. While other steps could be taken to eliminate this problem, they would require constraining the real data so much that it would not produce NHPP that represent the data.

Even though the correction procedure could not be applied to Chapter 2’s data, McLean et al.’s suggestion may explain some of the conflicts between strengthening and signalling that were discussed in that chapter. For example, there was a shift in preference towards the location of the just-productive alternative when that alternative was available post-reinforcement in Set 5 of that chapter. Perhaps that was caused by the continuation of the response bout that was not possible in other conditions. The same logic could be applied to the apparent interference caused by the presence of the just-productive alternative that occurred in Krägeloh, Davison, and Elliffe’s (2005), Boutros, Elliffe, Davison’s (2011), and Hachiga, Sakagami, and Silberberg’s (2014) results. If that is the case, then the local strengthening effect of reinforcement was not causing preference for the just-productive alternative when signalling was shifting preference away from that alternative.

When McLean et al. (2014) applied the correction procedure to Rodewald et al.’s (2010) data, they concluded that there was an effect of reinforcement and that the effect was
being overestimated by the use of preference pulse analyses. Because Rodewald et al. used Davison and Baum’s (2000) frequently changing procedure, it would be expected that the same result would occur if the correction procedure was applied to Davison and Baum’s data, as well as to the data from other procedures that are based on it (Aparicio, 2008; Davison & Baum, 2002; Landon & Davison, 2001). Overestimating the reinforcer effect size would not change the validity of the conclusions from those experiments as they all conclude that reinforcers caused preference for the just-productive alternative. It would also be difficult to see how the conclusions of Cowie, Davison, and Elliffe (2011), and Miranda-Dukoski, Davison, and Elliffe (2014), would be affected by the use of preference pulses. Preference was to the locally richer alternative immediately post-reinforcement in both experiments, and tracked the local reinforcer rates across time since reinforcement to varying degrees. Miranda-Dukoski et al.’s pulses were non-differential after reinforcement, and although the present analysis shows that it was possible to obtain non-differential pulses with a more-than-two-alternative procedure, it also shows that more than two alternatives are needed to produce them. Also it would be difficult to obtain Cowie et al.’s preference reversal by alternating response bouts, unless the bouts changed over time, and changes over time imply that the subjects are tracking time as concluded. However, it would be interesting to apply the correction procedure to the results of Krägeloh et al. (2005), Boutros et al. (2011), and Hachiga et al. (2014), as those experiments contained preference for the just-productive alternative that appeared to interfere with the signalling effects of reinforcement. If that interference disappears after a correction procedure, then the signalling effect may have been stronger than was concluded by the researchers.

The overall findings of this chapter suggest that although some patterns that are observed in preference pulse graphs may be due to the way that data were processed, the general patterns of post-reinforcement preference that were discussed in Chapter 3 were
primarily caused by the effects of reinforcement. It also appears that if data processing was causing the response patterns after reinforcement, then the run length of those data interacts with the shape of the pulse that is produced.
Chapter 5. Modelling

5.1. Introduction

The experiment in Chapter 3 was a steady-state version of a frequently changing procedure used by both Davison, Krägeloh, Fraser, and Breier (2007), and Elliffe and Davison (2010). The results of the experiment were compared directly with those of Elliffe and Davison, and differences were found between the local and global patterns of preference. While the differences in local preference were covered in detail, only one global model of preference was used for comparison. Comparisons of results showed that the assumptions of the generalised matching law (GML, 1974) were satisfied in the steady-state procedure but violated in Elliffe and Davison’s experiments. However, the GML is not the only method of global analysis that can be used to compare between the experiment in Chapter 3, and Elliffe and Davison. A different type of analysis was conducted on Elliffe and Davison’s data, and it was one by Navakatikyan, Murrell, Bensemann, Davison, and Elliffe (2013). The data from Chapter 3 can be reanalysed accordingly to see if there are any other differences between the frequently changing and steady-state procedures.

Navakatikyan (2007) proposed a component-functions model of choice behaviour. This model related the absolute rate of behaviour on each alternative to the absolute rate of obtained reinforcers. The term ‘component-functions’ refers to that fact that the model was comprised of two different functions:

\[ B_1 = F_{enh}(R_1) \times F_{red}(\Sigma R_{oth}) + B_a \]  \hspace{1cm} (5.1)

One function describes the enhancing effect that the rate of reinforcement has on the rate of behaviour made to an alternative; the other function describes the reducing effect that the rate of reinforcement from other alternatives has on the rate of behaviour made to that alternative.
same alternative. Different measures of behaviour can be used. For example, response rates, changeover rates, and residence time (the amount of time a subject responds to an alternative before switching) can and have been used. The model also incorporates parameters that describe the baseline rate of behaviour, and biases, on each alternative. It should be noted that Navakatikyan substituted different functions into each component when testing the model; these included power, exponential and hyperbolic functions. The naming convention used for these models uses the enhancing function’s name first and the reducing function’s second. For example, Power x Hyperbolic would describe behaviour using an enhancing power function with a reducing hyperbolic function.

The component-functions model was first tested on datasets from Alsop and Elliffe (1988), and Elliffe and Alsop (1996; see Navakatikyan, 2007). Navakatikyan compared the performance of several component-function models to other absolute models of behaviour, including two from Davison and Hunter (1976), using the Akaike information criterion (see Burnham & Anderson, 1998). The criterion showed that several of the component-functions models performed better than the other models that were included. Davison and Hunter’s models were the best of the rest of the models compared. The component-functions model was then extended by Navakatikyan and Davison (2010) to model dynamical datasets. Once again the component-functions models performed better than other existing models.

Finally, Navakatikyan et al. (2013) extended the model to behaviour from experiments that had more than two alternatives. Elliffe and Davison’s (2010) experiment was included in that set of analyses. In general, the component-functions models performed better than the other models that they were compared with. Additionally, the models without bias parameters performed better than those that included bias. This indicated that there was no bias in the data sets, and therefore including bias parameters did not improve the models in any way. One exception to the general finding was the dataset from Davison and Hunter
(1976), where the models from Davison and Hunter performed just as well as the component functions models.

One advantage that the component-functions model has over the GML and other matching-based equations, such as Davison and Hunter’s (1976) models, is that it does not make the same assumptions as those described by Prelec (1984). Many of the datasets analysed using the component-functions models are known to violate at least one of these assumptions. Examples include datasets that violate the assumption of relative homogeneity (Alsop & Elliffe, 1988; Elliffe & Alsop, 1996), assumption of relative independence (Davison & McCarthy, 1994; Murrell, 1995), as well as Elliffe and Davison’s (2010) dataset that violated multiple assumptions. Navakatikyan et al. (2013) suggested that this was why the component-functions models performed well with these datasets compared to other absolute models. They also suggested that differences in performance in other datasets may be an indication that the assumptions have been violated. In addition, the component-functions models had similar performance to Davison and Hunter’s model when modelling Davison and Hunter’s dataset, a dataset that is known to satisfy the assumptions of the GML. If differences in performance of the models do indicate violations in the assumptions of the GML, then it will be of interest to compare performance of the models on another dataset that is known to satisfy these assumptions – the dataset from Chapter 3. In addition, it will be of interest to see if the relationship used to describe the data from the steady-state experiment is similar to that of the frequently changing experiment of Elliffe and Davison. The model that performed best on Elliffe and Davison’s data was the Hyperbolic x Power-Hyperbolic with a baseline constant.
5.2. Modelling and Analysis

Navakatikyan (2007) described a full list of component-functions models that were used for the original analysis. However, Navakatikyan et al. (2013) only reported the results from six models when reanalysing Elliffe and Davison’s (2010) results. In addition to these models, the results of a few competing models were also reported, the best performing of these being from Davison and Hunter (1976). After the full set of models were fitted to the dataset from Chapter 3, it was determined that the best performing models were the component-functions models reported by Navakatikyan et al. and the models from Davison and Hunter. Due to this, the results of all other models were not reported. Additionally, Navakatikyan’s reanalysis of Elliffe and Davison’s data used response rates, changeover rates, and residence time as response measures. Of these three measures, response rates provided the best fits for their model. This was also found to be true for the steady-state dataset and only those data have been reported here.

Like Navakatikyan et al. (2013), models were fitted with and without bias. In the models with bias (c), the reinforcer rate from each alternative was multiplied by its bias parameter, for example, $c_iR_i$. There was one fewer bias parameter than number of alternatives because the parameters are scaled to each other, meaning that one of the alternatives bias values could be set to 1 and therefore does not need to be part of the equation. Navakatikyan et al. determined that the models with bias performed worse than the models without bias and so excluded them. This was not the case for the present analysis, and so both bias and non-bias models have been included.

The models were assessed using the same methods as Navakatikyan et al. (2013). A corrected Akaike information criterion (AICc) was used to determine the performance of each model. The AICc value is calculated using the model’s goodness of fit and complexity,
with lower values indicating better models. If two models provide equal goodness of fit, the
less complex model (fewer free parameters) will have the lower AICc value. However,
having a lower value does not automatically make a model better than another; generally
models must have a large difference in value before one is considered superior. To remain
consistent with Navakatikyan et al. we have used a difference of 10 as this criterion (see

Results show the values for 14 of the 16 models that were tested (Table 5.1). The
analysis program could not find solutions for the last models and so these have been
excluded. Table 5.1 ranks the performance of the models according to mean \( \Delta AIC \), which is
the difference between a model’s AIC score and the best score. The \( \Delta AIC \) values of the top 8
models are within 3 of each other, indicating that no model is reliably better than the others.
These were all of the bias models. The best of the no-bias models were outside of the
criterion of value of 10. This means that bias parameters were the largest determinant of a
model’s performance. This also suggests that there was detectable bias in the response data.
After bias parameters, the total number free parameters was the second largest determinant of
performance. Bias models with fewer parameters did better than those with more. That is,
according to AIC, the improvement in goodness of fit produced by adding more parameters
was not justified. The same pattern can be seen in the no-bias models. One of Davison and
Hunter’s (1976) models

\[
B_1 = a \left( \frac{R_1}{R_{x} + R_{oth} + k} \right)^s, \tag{5.2}
\]

placed 1st in the rankings. The best of the component-function models was the Power
x Hyperbolic model without the baseline constant

\[
B_1 = aR_1^k \left( \frac{1}{k_{oth}^{R_{oth}}} \right). \tag{5.3}
\]
The median fitted parameters for both of these models can be seen in Table 5.2.
Table 5.1

The performance statistics of the different models. Note Davison & Hunter 1 is labelled as Davison & Hunter B in Navakatikyan et al. (2013).

<table>
<thead>
<tr>
<th>Model Name</th>
<th>Ranking</th>
<th>Parameters</th>
<th>Δ Mean</th>
<th>Mean AIC</th>
<th>Mean vac</th>
</tr>
</thead>
<tbody>
<tr>
<td>Davison &amp; Hunter 1</td>
<td>1</td>
<td>6</td>
<td>0.000</td>
<td>95.359</td>
<td></td>
</tr>
<tr>
<td>Power × Hyperbolic</td>
<td>2</td>
<td>6</td>
<td>0.391</td>
<td>95.390</td>
<td></td>
</tr>
<tr>
<td>Davison &amp; Hunter 2</td>
<td>3</td>
<td>6</td>
<td>0.763</td>
<td>95.223</td>
<td></td>
</tr>
<tr>
<td>Hyperbolic × Hyperbolic</td>
<td>4</td>
<td>6</td>
<td>1.212</td>
<td>95.207</td>
<td></td>
</tr>
<tr>
<td>Power × Hyperbolic + b</td>
<td>5</td>
<td>7</td>
<td>1.308</td>
<td>95.536</td>
<td></td>
</tr>
<tr>
<td>Hyperbolic × Hyperbolic + b</td>
<td>6</td>
<td>7</td>
<td>1.512</td>
<td>95.461</td>
<td></td>
</tr>
<tr>
<td>Hyperbolic × Hyperbolic-Power + b</td>
<td>7</td>
<td>8</td>
<td>2.379</td>
<td>95.639</td>
<td></td>
</tr>
<tr>
<td>Power × Hyperbolic-Power + b</td>
<td>8</td>
<td>8</td>
<td>2.627</td>
<td>95.620</td>
<td></td>
</tr>
<tr>
<td><strong>No Bias Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power × Hyperbolic</td>
<td>9</td>
<td>3</td>
<td>19.062</td>
<td>88.791</td>
<td></td>
</tr>
<tr>
<td>Davison &amp; Hunter 1</td>
<td>10</td>
<td>3</td>
<td>19.536</td>
<td>88.543</td>
<td></td>
</tr>
<tr>
<td>Davison &amp; Hunter 2</td>
<td>11</td>
<td>3</td>
<td>19.959</td>
<td>88.312</td>
<td></td>
</tr>
<tr>
<td>Hyperbolic × Hyperbolic</td>
<td>12</td>
<td>3</td>
<td>20.245</td>
<td>88.261</td>
<td></td>
</tr>
<tr>
<td>Power × Hyperbolic + b</td>
<td>13</td>
<td>4</td>
<td>20.522</td>
<td>89.081</td>
<td></td>
</tr>
<tr>
<td>Hyperbolic × Hyperbolic + b</td>
<td>14</td>
<td>4</td>
<td>21.682</td>
<td>88.521</td>
<td></td>
</tr>
</tbody>
</table>
Table 5.2

The median fitted parameter values for the top two performing models

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Value 1</th>
<th>Value 2</th>
<th>Value 3</th>
<th>Value 4</th>
<th>Value 5</th>
<th>Value 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hunter &amp; Davison 1</td>
<td>a</td>
<td>101.8122</td>
<td>6.23102</td>
<td>0.704989</td>
<td>1.134817</td>
<td>1.301113</td>
<td>1.029343</td>
</tr>
<tr>
<td>Power x Hyperbolic</td>
<td>s/k_red</td>
<td>26.58395</td>
<td>0.652692</td>
<td>106.3877</td>
<td>1.138469</td>
<td>1.268217</td>
<td>0.982881</td>
</tr>
</tbody>
</table>
5.3. Discussion

The component-functions models from Navakatikyan (2007) performed as well as Davison and Hunter’s (1976) models when applied to the dataset from Chapter 3. In addition, the models with bias parameters performed better than those without such parameters, and less complex models generally did better than more complicated versions.

Navakatikyan et al. (2013) suggested that superior performance of component-functions models compared to those that are based on the GML can be used to indicate violations of the assumptions of GML. The results of the present analyses analysis show that the component-functions models did not perform better than Davison and Hunter’s (1976) models (Table 5.1). If Navakatikyan et al.’s suggestion is followed this result means that the data from Chapter 3 are consistent with the assumptions of the GML, as was also indicated by analyses performed in that chapter. In addition to that, Navakatikyan et al.’s results indicate that the component-functions models performed better on Elliffe and Davison’s (2010) dataset, which was believed to have violated the assumptions. This converging evidence supports the conclusion that data from Chapter 3 satisfied the assumptions, whereas Elliffe and Davison’s data did not.

The difference between bias and no bias in the reanalysis of Elliffe and Davison’s (2010) results and that of Chapter 3 can be explained by the way that data are arranged. The reason that there is no bias in Elliffe and Davison’s data is that data from each alternative were combined and contributed to each point of in the analysis. This combination process eliminated any location bias in that dataset. It was not possible to combine the data from Chapter 3 in that manner, as that would reduce the usable data down to one point and would prevent the analysis. This means that the data would retain any bias the subjects had for each alternative.
The difference between the best equation for Elliffe and Davison’s (2010) data and Chapter 3’s data is not an issue. As reported in the results, many models performed as well on the Chapter 3 data as each other. Contained within this grouping was the Hyperbolic x Hyperbolic-Power model that provided the best fit for Elliffe and Davison’s data; thus, the use of this model could be justified for both datasets. As Navakatikyan et al. (2013) stated there is no theoretical justification for one model over the other, therefore using the same model for both should not be an issue. The main reason the Hyperbolic x Hyperbolic-Power model did not achieve the highest ranking for Chapter 3’s dataset is its complexity. The model’s goodness of fit is as good as all the other top ranking models; however the AIC value is larger because it has more parameters.

Although the data from Chapter 3 were well modelled by Davison and Hunter’s (1976) GML based equations, the fact that other datasets are better modelled using Navakatikyan’s (2007) models shows a theoretical weakness of the GML. In order for the GML to model data appropriately, those data must satisfy certain assumptions (See Prelec, 1984). If these assumptions are not satisfied the data should not be modelled using the GML. This causes problems for the GML if it is accepted as a general model of choice behaviour, as any occasion that it is not appropriate is a reason not to treat the model as general. The biggest issue is that the model does not provide any mechanism to account for violations of the assumption of homogeneity (for example Alsop & Elliffe, 1988; Elliffe & Alsop, 1996) or violations of relative independence (for example, Davison et al., 2007; Elliffe & Davison, 2010). Because of this issue the GML can be treated as either incomplete (Jensen, 2014) or inaccurate (Navakatikyan, 2007).

Jensen (2014) suggested that the log-ratio version of the GML is incomplete and that trying to model more than two-alternative data with log ratios introduces errors that can be avoided by using matrix algebra. This is because log ratios describe relationships between
only two alternatives, and so are inadequate for describing the simultaneous relationships for when there are more than two alternatives. With that said, Schneider and Davison (2005) suggested a multi-alternative equivalent to log ratios that was used by both Davison et al. (2007), and Elliffe and Davison (2010). Their suggestion was a generalized version of Natapoff's (1970) multi-alternative log ratios. However, using Jensen’s suggestion of modelling all alternatives simultaneously using matrices allows for changes to one alternative to cause changes to the relationships between other alternatives and when the number of alternatives is reduced to two, behaviour can be modelled using log ratios. This is because behaviour between alternatives is treated as being on a sample space rather than a continuum (like a log ratio), and increases in behaviour on one alternative force shifts within that space. In other words, behaviour between alternatives is not independent of other alternatives as assumed. Although it should be noted that these assumptions were made only to justify using log ratios to model behaviour with any number of alternatives (Prelec, 1984), treating log ratio models as a special case of a more complex model of behaviour avoids this assumption.

While Jensen (2014) treats the GML as if it were incomplete, Navakatikyan (2007) treats the GML as inaccurate. The GML models changes in preference for one alternative over another as being caused by changes to the reinforcement contingences of those alternatives, and only by changes to those alternatives. This means that changes to anything else, such as reinforcement for other alternatives, should not cause changes to preference between of the modelled alternatives. Because Navakatikyan’s component-functions model states that changes to preference are caused by all sources of reinforcement, a change to one alternative can cause changes to preference between other alternatives. This occurs because Navakatikyan suggests that behaviour on an alternative is the result of enhancement effects by reinforcement on that alternative, plus the inhibition effects of reinforcement on all other
alternatives. Therefore, Navakatikyan treats the GML as erroneous because it does not consider all sources of reinforcement when modelling behaviour between alternatives.

When considering both Navakatikyan’s (2007) and Jensen’s (2014) suggestions, it seems that the GML in its current form is inadequate in principle to model behaviour from datasets with more than two alternatives. It does not mean that it cannot be used for all such datasets, as all analyses indicate that it does model data from Chapter 3. Rather, it does not accurately model all datasets with more than two alternatives. This means that treating the GML in its current form as a general model of choice is unwise. Instead the model should be adapted or discarded.

In summary, the use of Navakatikyan’s (2007) component-functions models did not provide a better fit than other models of absolute choice behaviour. However this fact is not an issue with the model, because it is a confirmation of the validity of Chapter 3’s conclusion that the steady-state dataset provided data that were consistent with the GML, and a frequently changing version of the same procedure did not.
Chapter 6. Discussion

6.1. General results

The aim of this thesis was to investigate the local properties of reinforcement. Chapter 2 detailed a procedure where the response location that produced the immediately preceding reinforcer was removed and replaced by a new alternative. In that situation, subjects showed preference for the richest of the available alternatives immediately after reinforcement. This occurred both when the richest alternative was fixed within a condition, and when the richest alternative was signalled in a way that was similar to Krägeloh, Davison and Elliffe (2005). Chapter 3 was a steady-state replication of Elliffe and Davison’s (2010) frequently changing procedure. The results showed that subjects’ preference was always for the richest alternative, instead of towards the location of the previous reinforcer as it was in Elliffe and Davison’s results. This was likely due to the differences between how behaviour is controlled in steady-state and frequently changing procedures. Chapters 4 and 5 provided additional analyses that supported the conclusions made in Chapters 2 and 3.

The main properties of reinforcement investigated were strengthening and signalling effects. Strengthening is a traditional idea from the law of effect (Thorndike, 1911), whereas signalling is a more recent idea (Davison & Baum, 2006; Shahan, 2010). Signalling, also known as signposting, is where the reinforcer is treated as a stimulus, and subject uses that signal to gain information about how to obtain more reinforcers. This property causes subjects to prefer the just-productive alternative if the signal indicates that alternative as likely to produce the next reinforcer, or to shift preference to another alternative if the signal indicates the other alternative as likely to produce that reinforcer. Strengthening, on the other hand, has one basic effect, increase the probability of the previous response and therefore increase preference for that alternative. There are two ways for the probability to increase,
locally and globally. An increase in the global probability of behaviour leads to greater overall preference for an alternative, whereas increases in local probability increase preference immediately after the reinforcer. Local increases are also known as local enhancement (Menlove, 1975), and unless otherwise stated, any reference to a strengthening effect refers to local enhancement.

It was evident that a strengthening effect did not explain all of the obtained results. For example, in Chapter 2 the location of the previous reinforcer was removed – this was done to remove possible strengthening effects. Doing so showed that the previous reinforcer was not required in order to produce preference for the richer alternative immediately after a reinforcer. In other words a strengthening effect was not required for preference to favour one alternative immediately after reinforcement. In addition, observations from both Chapters 2 and 3 showed that steady-state procedures with more than two alternatives produce preference for the richest alternative immediately after reinforcement. This provides further evidence that reinforcers do not necessarily shift preference towards the location of the previous alternative, as has been found in two-alternative steady-state procedures (e.g., Davison & Baum, 2002; Landon, Davison & Elliffe, 2002; Menlove, 1975). However, there was some evidence for a local strengthening effect. Set 5 of Chapter 2 showed that when the location of the immediately preceding reinforcer was available post-reinforcement, preference was somewhat shifted towards that location, although was still towards the richest alternative. Therefore, while strengthening was not compatible with all of the obtained results, it may still have had an effect on local preference.

On the other hand, signalling could explain most of the local preference patterns that were observed. Preference for the richest alternative post-reinforcement was observed in virtually all reported results. If reinforcers were providing information about the location of the next reinforcer then the subjects always reacted appropriately to that information. The
strongest evidence for signalling comes from Set 6 of Chapter 2, where experimental contingencies were similar to those of Krägeloh et al. (2005). In those conditions the probability of reinforcement on each alternative depended on the location of the immediately preceding reinforcer. Not only did subjects prefer the richest alternative post-reinforcement, the preference for that alternative increased as the probability of reinforcement on that alternative increased. The signal was providing more information than just location of the richest alternative.

However, even though treating the reinforcer as a stimulus can explain most of the obtained results, it is not required to do so in all cases. Preference for the richer of the available alternatives in Sets 2 and 3 of Chapter 2 could be due to the subjects’ history with the available alternatives and not the stimulus provided by location of the immediately preceding reinforcer. In those sets, subjects could have determined which of the two alternatives was the richest by which alternative had produced more reinforcers when those two were paired together previously. Even though the available alternatives were linked to the location of the immediately preceding reinforcer, that link was not required. Further evidence that reinforcer-based signals were not required for subjects’ choice came from Set 5 of Chapter 2, where the available alternatives were not linked to the immediately preceding reinforcer. Under those conditions, subjects still preferred the richer of the available alternatives even though the location of the immediately preceding reinforcer could not signal which alternative was richer.

Aside from strengthening and signalling, a different approach to how reinforcers modify local preference patterns was investigated. McLean, Grace, Pitts, and Hughes (2014) suggested that preference pulses that show an immediate increase in preference for the just-productive alternative may be due to the way that data are organised in those analyses. Their suggestion was that if subjects respond in bouts to each alternative then these bouts will be
occasionally interrupted by the delivery of reinforcers. In preference pulse analyses data are
typically categorised by when they occurred after a reinforcer, either by time elapsed or
sequence since reinforcement (see Boutros, Elliffe, & Davison, 2010 for a discussion of pulse
types). This means that response data are being organised by the reinforcers for these
analyses and this organisation may be creating patterns of preference that are misleading. The
results of Chapter 4 show that even if preference pulses are partially created by the analysis
and not the subjects’ responding, the conclusions of Chapters 3 were still valid. Although
McLean et al.’s suggestion could explain preference patterns from both Chapter 3, and Elliffe
and Davison (2010), preference immediately after reinforcement was greater than was
expected by the data arranging process. With that said, preference after the initial preference
shift was the same as was expected by the data arranging process. This suggests that initial
preference in Chapter 3 was controlled by the properties of reinforcement, and after that
preference was the result of a repeated response pattern. The conclusions of Chapter 2 were
also assumed to be valid, even though no reanalysis could be performed in Chapter 4. If
reinforcers are merely interrupting response bouts as suggested by McLean et al., then
removing the just-productive alternative would effectively end that bout at the reinforcer. The
location of then next response would be due only to the subjects’ preference for the two
alternatives that were available. This means that the preference for the richer alternative
immediately post-reinforcement had to be caused by choice, and not by data organisation.
Therefore, subjects were choosing the richest alternative immediately after reinforcement, as
was concluded. The only possible exception to this was Set 5, where the slight shift to the
just-productive location, when possible, might be explained by McLean et al.’s suggestion.

6.2. Signalling properties of reinforcement

The results of Chapter 2 are consistent with the signalling theory of reinforcement. Of
all the results reported in that chapter, the results from Set 6 were the clearest demonstration
of that property. The contingencies arranged during Set 6 were similar to those of Krägeloh et al. (2005) and Boutros, Elliffe, and Davison (2011). They arranged the probability of reinforcement according to the location of the just-productive alternative; if the probability was .9 then there was a 90% chance that the next reinforcer was arranged on that same location. Since the just-productive alternative was removed in Set 6, the probabilities determined the likelihood that the next reinforcer would occur on the alternative that was present before the immediately preceding reinforcer, but did not produce that reinforcer. The results were similar to those of both Krägeloh et al. and Boutros et al.: Subjects were reacting to the reinforcer by showing preference for whichever alternative was signalled as the richest and the amount of preference was related to overall probability of reinforcement for the rich alternative.

An interesting finding of Set 6 was the results from the $p = 0$ and $p = 1$ conditions. The location of the next reinforcer was certain in these conditions and subjects showed consistent preference for that location at all times since reinforcement. The subjects did not show exclusive preference as there were responses to both alternatives; however that preference remained above indifference. Due to the removal of the just-productive alternative, these conditions were more similar to $p = 0$ conditions of Krägeloh et al. than the $p = 1$ conditions because the just-productive alternative never produced the next reinforcer. Therefore, these conditions were similar to the strictly alternating conditions of Boutros et al. as well. Comparing their results to those of Set 6 reveals some inconsistencies. Immediately after reinforcement in Krägeloh et al.’s and most of Boutros et al.’s $p = 0$ (or strictly alternating) conditions, there was weak preference for the location of the next reinforcer. After that, preference shifted towards the location of the next reinforcer, then to indifference. The only condition where the results resembled those of Set 6 was when immediately after reinforcement Boutros et al. presented the alternative that would be next arrange
Because Boutros et al. used a switching-key procedure, that alternative was the only one present after reinforcement. This situation occurred under two different conditions of their experiment: When that alternative was arranged to appear always, and when that alternative was selected to appear by chance. For some reason, having the just-productive alternative present after reinforcement causes less preference for the location of the next reinforcer. Possible reasons will be discussed in the strengthening section.

The results of Chapter 3 also support the signalling properties of reinforcement by showing how the usefulness of a discriminative stimulus can be modified. This was achieved by arranging a steady-state procedure so that any information gained from a signal would eventually become redundant and then comparing results to a frequently changing procedure where information was still relevant to the subject. The differences between the results of Chapter 3 and Elliffe and Davison (2010) were consistent with the differences between Davison and Baum (2002) and Landon et al. (2002), and with the differences between Davison and Baum (2006), and Boutros, Elliffe and Davison (2009): When contingencies frequently change, any signal (reinforcer or not) causes a larger shift in the subjects’ preference, compared with the same signal in a steady-state environment.

Although there was clear evidence for the signalling properties of reinforcement, and those properties can explain many of the findings in Chapters 2 and 3, it does not mean that the reinforcer-based stimulus was being used to guide preference at all times. For example, in Chapter 3 when the signal was redundant, the subjects did not need any information as to where the richest alternative was. The subjects may have been using previous experience with the contingencies to guide them to reinforcement instead of the signal. The same is true for Sets 2, 3 and 5 of Chapter 2. The subjects could not have been using a reinforcer signal to locate the richer alternative in Set 5 because there was no relation between the location of the just-productive alternative and the richer alternative. As the subjects consistently responded
to the richer alternative of any pair in that condition, they must have been using a different method to determine which alternative was the richest. That method had to be the current pairing of alternatives as it was the only constant. If the subjects were identifying the richest alternative by the pairings in Set 5, then it is also possible that they were doing that in Sets 2 and 3, even though it was also possible to identify the richer alternative using the location of the just-productive alternative in that situation. It is even possible for the subjects to have been identifying the richer alternative in Set 6 without a reinforcer signal. The subjects could have achieved that by tracking which alternative was present before and after the immediately preceding reinforcer. Although, there is evidence that reinforcers interfere with subjects’ memories of events before the reinforcer (Killeen & Smith, 1984), tracking which alternative was present before and after would be difficult and therefore unlikely to be the primary method that subjects used. Reinforcers may have signalling properties, but this does not necessarily mean that all preference is controlled by reinforcer-based stimuli.

6.3. Strengthening properties of reinforcement

While the signalling properties of reinforcement can account for most of the preference patterns that were observed in Chapters 2 and 3, very few of the results can be explained by local strengthening effects. Chapter 3 aimed to investigate strengthening effects by producing a steady-state dataset which limited the usefulness of signals that were provided by reinforcers. However, limiting the signal did not produce preference for the just-productive alternative as would be expected by a strengthening effect. Instead subjects produced patterns of preference that were similar after every reinforcer, regardless of which alternative was just-productive. This was unexpected as preference pulses from two-alternative concurrent VI VI procedures have shown preference for the location of the last reinforcer (Landon et al., 2002), and while there are explanations for those data, none of them are due to local strengthening effects.
There was some evidence for a strengthening effect in Set 5 of Chapter 2. Results showed a slight shift in preference towards the just-productive alternative when that alternative was available post-reinforcement. There was more preference for the richer alternative when the rich alternative was the just-productive alternative, compared to when the lean alternative was the just-productive alternative. Unfortunately, this pattern can also be explained by McLean et al.'s (2014) suggestion about how data organisation in preference pulses may cause misleading results. As it was not possible to perform a correction procedure on those data, it was not possible to identify whether that shift in preference was greater than would be expected by the preference pulse analysis.

While neither Chapter 2 nor 3 produced a conclusive strengthening effect, one result from Set 6 of Chapter 2 might be related to such an effect. The resemblance between Set 6 and the data from some conditions of Boutros et al. (2011) suggests that subjects did not show strong preference for the location of the next reinforcer in Boutros et al. due to the presence of the just-productive alternative immediately after reinforcement. There are multiple explanations for that result, one of which is strengthening. The simplest explanation is that pigeons have a win-stay bias. This means that the subjects in these experiments have the tendency to prefer the alternative that produced the immediately preceding reinforcer. Although this sounds similar to a strengthening effect, it is different because it is species-specific. Some animal species may have an innate tendency to stay after winning (receiving a reinforcer); other species may have the tendency to shift to a different source of reinforcement. However, strengthening does not account for species-specific differences. Possible reasons why pigeons have a win-stay tendency have been given by Baum (2012), and others including Boutros et al. (2011) have suggested this as an explanation of this data pattern. The general explanation is that pigeons evolved a win-stay tendency for foraging because their primary food source (grain, seeds, and fruit) tends to be found grouped together.
in the wild. This means that finding one piece of grain usually signals that more grain can be found in the immediate vicinity and that the pigeon should stay where it is to exploit the current location. This foraging strategy then transfers into operant procedures where the pigeons also stay in an attempt to gain more food. However, this win/stay tendency does not explain why the pigeons in Boutros et al. did not switch back to the just-productive alternative immediately after reinforcement. If they prefer the winning alternative, then they should show preference for it, even when it is not presented immediately.

A different explanation for the above result is to do with contingency discriminability. Perhaps the subjects were not detecting the strictly alternating pattern of reinforcement or were confusing/misremembering the source of the immediately preceding reinforcer. If that was the case then preference would be expected to be less extreme for the location of the next reinforcer. Additionally, when Boutros et al. (2011) presented the next-to-be-reinforced alternative after reinforcement, that could be used as a cue and therefore produce the observed preference for that alternative. However, if subjects were using the alternative that was presented after reinforcement as a cue, they did not do the same thing in conditions when the just-productive alternative was presented after a reinforcer because that alternative would also have been a reliable cue to reinforcement. Also this does not explain why post-reinforcer preference differed depending on which alternative was presented after reinforcement when that alternative was randomised, as there was no reliable cue in that circumstance. Finally, subjects shifted towards the location of the next reinforcer after initial post-reinforcement preference, which seems inconsistent with a poor discrimination of the contingencies.

What might have been happening was that the presence of the just-productive alternative produced a strengthening effect that was interfering with the signalling effect. Strengthening would have been driving preference towards the just-productive alternative and signalling would have been driving preference towards the other alternative. When the just-
productive alternative was not present, as in some of Boutros et al.’s (2011) conditions and in Set 6, there was no interference from strengthening and therefore, clear preference for the location of the next reinforcer. However, the lack of a detectable strengthening effect in both Chapters 2 and 3 makes it difficult to support this suggestion by using the data from this thesis.

With that said, evidence of interference between a local enhancement effect and a signalling effect has been observed by Hachiga, Sakagami, and Silberberg (2014). They signalled which alternative would produce the next reinforcer by use of non-reinforcer-based stimulus, which as available immediately after the delivery of the previous reinforcer. Preference differed based on whether the next reinforcer was signalled on the just-productive alternative or the other alternative, and results resembled those from Boutros et al. when the just-productive or other alternative was presented after reinforcement. When the next reinforcer was signalled to occur on the just-productive alternative, preference patterns were similar to when the next-to-be-reinforced alternative was presented in Boutros et al.’s experiment; that is, there was clear preference for the location of the next reinforcer. When signalled on the other alternative, preference started at indifference before shifting towards the location of the next reinforcer, similar to what happened when the just-productive alternative was presented post-reinforcement in Boutros et al.’s experiment. The results of Hachiga et al. could also be explained by McLean et al.’s (2014) suggestion, however. The only thing that can be said with certainty is that the presence of the just-productive alternative does interfere with preference for the next-to-be-reinforced alternative, when that alternative is not the same as the just-productive alternative.

Overall, there was little evidence for local enhancement/strengthening effects of reinforcement in this thesis. Due to this, it is difficult to support a strengthening explanation for local effects obtained by others. For example, it seems unlikely that Shimp’s (1966)
conclusion of a strengthening effect in a discrete-trial procedure was accurate. It also seems unlikely that strengthening adequately explains results that show increased preference for the just-productive alternative in free-operant procedures (Davison & Baum, 2002; Landon et al., 2002; Menlove, 1975). With that said, the global effects of reinforcement were not investigated in this thesis. The strengthening effect of reinforcement may only function at an overall level and may not contribute to local preference patterns. For example, a reinforcer might signal to a subject where and when a reinforcer will occur (Cowie et al., 2011; Miranda-Dukoski et al., 2014), but might strengthen the probability of responding at all. This idea is similar to that of arousal and coupling from the mathematical principles of reinforcement (see Killeen & Sitomer, 2003). Arousal is similar to strengthening because a general increase of behaviour is caused by reinforcement, whereas coupling is similar to signalling because it relates the contingencies of reinforcement to subjects’ patterns of behaviour. Of course this would then raise the issue of whether the local and global preference patterns are independent of each other, and if not, which level of preference controls behaviour, which has been debated. For example, this thesis, as well as many frequently changing procedures (e.g., Baum & Davison, 2004; Boutros et al., 2009; Cowie et al., 2011; Davison & Baum, 2002; Davison et al., 2007; Elliffe & Davison, 2010; Krägeloh & Davison, 2003; Krägeloh et al., 2005; Landon & Davison, 2001; Miranda-Dukoski et al., 2014), assume control by local level choice, whereas others (e.g., McLean et al., 2014) suggest that the local patterns are a consequence of global preference.

6.4. Reinforcers and behaviour

This thesis has treated reinforcers as tangible objects which have their own unique properties that alter local patterns of behaviour. However, the biggest factor determining the effect of a reinforcer on preference was the experimental environment, which includes how the reinforcement contingencies were arranged and altered. For example, preference patterns
from Elliffe and Davison’s (2010) frequently changing procedure were different from those reported in Chapter 3. The reinforcers and reinforcement contingencies for both experiments were practically identical – only the rate at which the contingencies changed differed. Another example is the differences between Sets 3 and 4 from Chapter 2. The only procedural difference between the two was that in Set 4 the inactive alternative was lit. Attributing properties to reinforcers is not the best way to explain preference changes in these results. Instead it would be better to view these preference patterns and their relation to reinforcers as context-dependent.

The two examples mentioned above do not comprise a complete list of behaviours that appear to be context dependent. There were differences between Sets 2 and 3 of Chapter 2, caused by the presence or absence of a changeover delay (COD). There were also differences in Set 5 depending on which alternative produced the previous reinforcer. Although these results could be attributed to a strengthening effect, it would be the only context in which this effect was observed. In Set 6, the preference depended on the probability of reinforcement associated with the condition, and the location of the immediately preceding alternative. Control by reinforcement in that dataset was different from the others as the subjects needed to track the location of the immediately preceding reinforcer in Set 6, but did not need to in the other sets. As argued previously, subjects did not need to track that location in Sets 2, 3, and 4, and gained no benefit by doing so in Set 5. When these differences, and those between Sets 3 and 4 are considered, Chapter 2 as a whole demonstrated how a mixture of reinforcement contingencies (e.g., CODs), and environmental events (e.g., whether an alternatives were active and/or lit), had more control over local preference than reinforcer-based stimuli that were presented.

Contextual effects on preference can also be seen when comparing frequently changing procedures to steady-state versions. Local preference results differed between
Davison and Baum (2002), and Landon et al. (2002), and between Davison and Baum (2006),
and Boutros et al. (2009), as well as between Elliffe and Davison (2010) and Chapter 3 of this
thesis. The rate of contingency change determined how subjects behaved during different
environmental events. When changes occurred frequently, the subjects’ preference shifted
more after any signal (reinforcer or otherwise) than it did when the contingencies were kept
constant for extended periods of time. In Chapter 3 it was argued that the signal produced by
reinforcement was made uninformative in the steady-state environment. While that argument
was for evidence of the signalling property of reinforcement, it is also a demonstration of
how effects are contextual. When a reinforcer was uninformative about the location of the
next reinforcer, subjects did not appear to be using the signal as a guide. Therefore, the
signalling property was not needed to determine preference, as it was not needed in some
datasets from Chapter 2.

Additionally, results from Chapter 3 also show how simple behaviours can be
contextual. The probability of switching to each alternative in Conditions 1 – 8 was related to
the rates of reinforcement that were obtained from those alternatives. This was consistent
with the results of Elliffe and Davison (2010). However, that pattern was not observed for
Condition 9, where the rates of reinforcement were equal. Instead of being equally likely to
switch to each alternative, subjects switched according to a sequence. It is not known why
this was the case, but it does support the idea that behaviour is contextual.

If behaviour is viewed as contextual, then it still conforms to the core belief of
behaviourism that a subject’s behaviour is a product of its environment. It also conforms to
Herrnstein’s (1961, 1970) view of all behaviour being a form of choice behaviour. What the
contextual view does change is the behaviour’s relationship to reinforcers. All choice models
(e.g., Baum, 1974; Herrnstein, 1970; Hinson & Staddon, 1983; Jensen, 2014; Navakatikyan,
2007; Vaughan, 1981) describe behaviour as a function of reinforcers. In other words,
reinforcer rates are the independent variable and behaviour is the dependent variable. On the other hand, contextual behaviour would mean that the environment as a whole is the independent variable that determines the patterns of behaviour that are emitted by the subject, which in turn determines the amount of reinforcers gained for each behaviour. This idea fits with research that treats behaviours as goal-directed strategies (e.g., Racey et al., 2011) to gain reinforcers. The reinforcer becomes an inducer of behaviour (Baum, 2012), not a directly controlling factor.

The previous paragraphs in this section are not intended to imply that reinforcers do not have properties that control behaviour. Rather, the argument is that local preference does not depend solely on those properties. Reinforcers do act as signals to guide subjects to more reinforcers in some circumstances, but they also do not guide subjects to more reinforcers in other circumstances. The same applies to strengthening, although evidence suggests that the circumstances under which that property applies is limited, if it is indeed a property of reinforcement and not a form of signalling. It appears that what determines preference in any context is whatever provides a reliable cue to the next reinforcer. For example, in Set 5 of Chapter 2, a reliable cue was the alternatives available post-reinforcement, whereas in Set 6, a reliable cue was a combination of what alternatives were available before the immediately preceding reinforcer and what alternatives were available after. A reliable cue in Elliffe and Davison’s (2010) procedure was the location of the immediately preceding reinforcer, whereas the past experience provided a reliable cue in Chapter 3. Therefore, treating preference as a function of reinforcement is limited because preference is determined by the whole environment.

Before concluding this thesis, it is worth briefly discussing why people who are not experimental researchers should be concerned about the properties of reinforcers. A major advantage of determining the properties of reinforcement is that it allows a deeper
understanding of when the presentation of a reinforcer should lead to an increase of a behaviour and when it should not. This understanding can help those (like applied behavioural analysts) who use reinforcers to modify the behaviours of others. For example, Wolf, Risley and Mees (1964) successfully used candy as a reinforcer to get a child to wear his glasses. As far as Wolf et al. were concerned, the presentation of the candy strengthened the target behaviour and only the traditional understanding of reinforcement was required to accomplish their goal. But what if the presentation of candy had not lead to an increase in the target behaviour? The simple explanation would be that candy was not a reinforcer for the child and Wolf et al. needed a different reinforcer. However, that explanation would not help if Wolf et al. tried every possible reinforcer and still failed to increase the rate of the target behaviour. At that point a more complex understanding of reinforcement would be needed to solve the problem. In this scenario Wolf et al. might find that the child responded to social attention and that the reinforcers inadvertently signalled that no one would pay attention to the child immediately after the reinforcers had been delivered. The child would be unlikely to work for the reinforcers in this situation, and Wolf et al. should occasionally and unpredictably pay attention to the child after reinforcement in order for the reinforcers to work. Understanding that reinforcers have signalling properties would reduce the likelihood of the above scenario happening.

In summary, the experiments in this thesis were conducted with the belief that both strengthening and signalling properties of reinforcement exist, and used local analyses of choice procedures to isolate each property for observation. In Experiment 1, the signalling property was isolated by creating an environment where the previously reinforced response alternative was removed and replaced with a previously unavailable response. The subjects responded to those environmental contingencies by showing preference for the locally richer alternative immediately after reinforcement. When the experiment was altered so that
subjects could respond to the just-reinforced alternative, there was a small increase in preference for the just-reinforced alternative, but only when that alternative could produce reinforcers. In Experiment 2, the strengthening property was isolated by removing the relevance of any discriminative stimulus that was produced by reinforcement. Subjects responded by showing preference for the richest alternative immediately after reinforcement rather than preferring the location of the last reinforcer, as has been observed in local analyses of previous concurrent VI VI procedures. Overall, at local levels of analysis there was more evidence to support signalling properties of reinforcement than the strengthening properties. However this does not mean that strengthening properties of reinforcement do not exist.
Chapter 7. References


