Georgia State University

ScholarWorks @ Georgia State University

Psychology Theses

Department of Psychology

5-2-2018

Randomness and Structure to Humans and Rhesus Macaques

William Whitham

Follow this and additional works at: https://scholarworks.gsu.edu/psych_theses

Recommended Citation

Whitham, William, "Randomness and Structure to Humans and Rhesus Macaques." Thesis, Georgia State University, 2018.
doi: https://doi.org/10.57709/11928966

This Thesis is brought to you for free and open access by the Department of Psychology at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Psychology Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

RANDOMNESS AND STRUCTURE
TO HUMANS AND RHESUS MACAQUES

by

WILL WHITHAM

Under the Direction of David A. Washburn, PhD)

ABSTRACT

Random, unpredictable, unstructured stimuli are an everyday part of life. Yet despite this breadth of experience and sophisticated statistical learning mechanisms, humans report patterning even in stimuli that are paradigmatically random. In two experiments, participants evaluated structured and random environments presented in a common statistical learning paradigm, the Serial Reaction Time task. I presented random and specifically nonrandom sequences to humans (Experiments 1 and 2) and rhesus macaques (*Macaca mulatta*, Experiment 2) to explore the seemingly antagonistic relationship between explicit, intuitive beliefs about these sequences and implicit statistical learning of sequence properties. Sequence predictability and experience with a given sequence type significantly predicted reaction times only weakly

and inconsistently across the two experiments. Accordingly, participant choices scarcely deviated from chance, and in those rare cases they deviated from chance largely without directionality, and were not significantly predicted by either sequence predictability or experience with a given sequence type.

INDEX WORDS: Statistical learning, Implicit learning, Randomness, Comparative psychology, Rhesus macaques, Decision making

RANDOMNESS AND STRUCTURE

TO HUMANS AND RHESUS MACAQUES

by

WILL WHITHAM

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2014

Copyright by
William Taylor Whitham II
2014

RANDOMNESS AND STRUCTURE

TO HUMANS AND RHESUS MACAQUES

by

WILL WHITHAM

Committee Chair: David A. Washburn

Committee: Michael J. Beran

Christopher M. Conway

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2018

TABLE OF CONTENTS

RANDOMNESS AND STRUCTURE ..... **i**

TO HUMANS AND RHESUS MACAQUES ..... **i**

RANDOMNESS AND STRUCTURE ..... **iii**

TO HUMANS AND RHESUS MACAQUES ..... **iii**

RANDOMNESS AND STRUCTURE ..... **v**

TO HUMANS AND RHESUS MACAQUES ..... **v**

LIST OF TABLES ..... **viii**

LIST OF FIGURES ..... **ix**

1 INTRODUCTION **1**

1.1 Studying Random Events **2**

1.2 Humans and Random Events **4**

1.3 Randomness Failures in Broader Context **9**

1.3.1 Animal literature ***9***

1.3.2 Statistical learning ***11***

1.4 Randomness and Structure **13**

2 EXPERIMENT 1 ..... **15**

2.1 Methods **15**

2.1.1 Participants and Apparatus ***15***

2.1.2 Serial Reaction Time task ***15***

2.1.3 SELECT task *18*

2.2 Results **20**

2.2.1 SRT modeled *23*

2.2.3 SELECT modeled *27*

2.3 Discussion **29**

3 EXPERIMENT 2 ..... **29**

3.1 Methods **30**

3.1.1 Participants *30*

3.1.2 Training *31*

3.1.3 Training results *32*

3.2 Results **32**

3.2.1 SRT modeled. *36*

3.2.2 SELECT modeled. *44*

3.3 Discussion **50**

4 GENERAL DISCUSSION **50**

REFERENCES **58**

APPENDICES **66**

Appendix A: Sequence generation procedure for overalternating sequences **66**

Appendix B: Training progress for each Rhesus macaque **67**

LIST OF TABLES

Table 1 Research Questions and Hypotheses 16

Table 2. Linear Mixed-Effects Model Fit 24

Table 3. Training Summary 33

Table 4. Linear Mixed-Effects Model Fit 45

LIST OF FIGURES

Figure 1. Task Stimuli..... 19

Figure 2. Task appearance. 19

Figure 3. Mean RTs aggregated across participants.. 21

Figure 4. Aggregated participant choices 22

Figure 5. Averaged observed log(RT/ms) for humans 25

Figure 6 Visualization of model fit:..... 26

Figure 7. Histograms of individual participants' parameter estimations..... 28

Figure 8. Mean RTs aggregated across human participants 34

Figure 9. Mean RTs to stimuli for each monkey.. 35

Figure 10. The mean RT plots from Figure 9, rearranged..... 37

Figure 11. Aggregated human participant choices. 38

Figure 12. Aggregated choices monkey Hank..... 39

Figure 13. Aggregated choices for monkey Murph..... 40

Figure 14. Aggregated choices for monkey Lou 41

Figure 15. Aggregated choices for monkey Chewie..... 42

Figure 16. Aggregated choices for monkey Obi..... 43

Figure 17. Averaged observed log(RT/ms) for humans 46

Figure 18. Averaged observed log(RT/ms) for monkeys 47

Figure 19. Histograms of frequency of deviations of individual human participants' parameter estimations 48

Figure 20. Individual monkeys' parameter estimations..... 49

Figure 21. Mean RTs for two monkeys from this manuscript in a previous SRT task 53

Figure 22. Mean RTs to different sequence positions for the monkeys from Experiment 2 in their final 5 sessions 54

Figure 23. Mean RTs to different sequence positions for the monkeys from Experiment 2 across all sessions 55

1 INTRODUCTION

Human animals (hereafter, humans) and nonhuman animals (hereafter, animals) process sequential information in a variety of contexts. Videos and songs are the compilation of sequences of individual images or sounds that combine to form meaningful, distinct wholes. Spatial navigation, whether through maze, metropolitan street, or vast foraging environment, is the concatenation of a sequence of decisions made at choice points (Tolman, 1938). Language is made up of only a noisy stream of speech sounds from which humans, from infancy, nevertheless acquire a functional understanding.

Humans and animals are also exposed to a subset of sequences that have no inherent structure. Such stimuli will be described here as random or unstructured. The concept of randomness is notoriously difficult to define and, following Nickerson (2002) and others, the term will be used here to describe stimuli produced by a process with specific properties. First, the process must produce each potential output with equal probability. A process that outputs according to chance, but is not equally likely to output each possible alternative, would be probabilistic, but not truly random. Second, each output must be independent of all previous and future outputs. Combined, these properties ensure that any given sequence of outputs generated by such a process is as likely as any other sequence of outputs generated by this process.

Fair coin flips, dice rolls, roulette wheel spins, and lottery drawings are examples of processes that approximate the described properties. But examples of paradigmatically random processes are not limited to the outputs of games of chance. The typical determination of human biological sex is very nearly a random process, with female and male sexes equally likely to be conceived and each person's biological sex independent of that of any other person (Nickerson, 2002; Scheibehenne, Todd, & Wilke, 2011). Protean predator-prey interactions are often given as

another, more ecological, example of unstructured patterns of behavior (Neuringer & Jensen, 2013; Sanabria & Thrailkill, 2009). Consider a hungry shark chasing a seal in the open ocean. This chase creates a dynamic in which the optimal response of the seal is to be as unpredictable in its zig-zags as possible and, ideally, be perfectly random, to elude the sharp teeth that gnash behind it. The shark's optimal response is to match the proportion of zigs and zags emitted by the seal without becoming predictable. Any deviations from randomness in the responses of either animal will be exploited, and thus each must be prepared both to discriminate any degree of nonrandomness and to produce a random series of outputs.

Any sophisticated sequence processing comes first from the detection that meaningful relations between stimuli exist. For example, language learning is possible given exposure to natural language at a critical developmental period, but will never be facilitated by ungrammatical auditory stimuli. The constant challenge to the observer is to infer the properties of sequences, and of the sequence-generating process, on the basis of very limited information. This challenge has long been of interest to psychologists.

1.1 Studying Random Events

Because of the various connotations of the word 'random,' two distinctions should be made. First, random does not necessarily connote uniqueness. A sequence of outputs from a process with the aforementioned properties need not be unusual or unique, so long as this sequence was as likely to be generated as any other. A random sequence can appear highly structured, and indeed any sequence of sufficient length is extremely likely to appear structured for some subset of the sequence (Nickerson, 2002). Similarly, no absolute judgment of randomness may be made on the basis of analysis of the outputs alone. There is no proof that a sequence of events is random. Even highly unlikely outputs will be produced by a truly random

process, and unlikely events never occurring would be a highly salient marker for nonrandomness (Paulos, 1989).

No single, standardized method for assessing the degree to which humans can produce or discriminate a random process has been firmly established (for a review, see Bar-Hillel & Wagenaar, 1991). Nevertheless, some general types can be identified. One methodology uses versions of economic games with mixed strategy equilibria (for examples, see Mookherjee & Sopher, 1994; Rapoport & Budescu, 1992; Sanabria & Thrailkill, 2009). Such games are well-suited for use with both humans and animals, and the apocryphal predator-prey dynamic outlined above can be thought of as a naturalistic version of the game Matching Pennies (MP). In such a game, two players repeatedly compete in a head-to-head, winner-take-all game. One player wins a round by matching the responses of the second player (i.e., the shark bites where the seal is predicted to be), while the second player wins by preventing this from happening (i.e., the seal eludes the shark).

Another methodology, discrimination experiments, required participants to determine whether a sequence was the result of a random process (e.g., Lopes & Oden, 1987). For example, participants were given 100-item sets of heads and tails outcomes and asked to determine whether the set was generated by a random or nonrandom process. In a third, but complementary, methodology, production experiments present participants with a set of possible stimuli (for example, lefts and rights, heads and tails, ones and zeroes, X's and O's, letters of the alphabet, etc.; hereafter 'items') that they are instructed to produce a random sequence (for an example, see Chapanis, 1953). In such an experiment, a participant might be asked to simulate the outcomes of flipping a fair coin 100 times (e.g., *Heads-Heads-Tails-Heads-Tails-Tails-Tails…*). Experimenters can then use the objective properties of the sequences that participants

discriminated or produced – the relative frequencies of each item, any serial dependencies across items - to determine their understanding of randomness (for an example, see Lopes & Oden, 1987).

Though there is no test of randomness, experimenters can explore the conditions in which participants make the judgment that a sequence is random and probe the participants' responses for evidence that they are not truly approximating the likely outputs of a random sequence. Serial dependencies, in which a given item is more or less likely based on the items that come before or after it, are often used for this purpose (for an example, see Neuringer, 1986). The relative frequencies of each of the alternative stimuli are used as another measure, as are the relative frequencies of series of two or more items.

1.2 Humans and Random Events

A few conclusions may be drawn from the experimental frameworks outlined above. Humans, especially outside of the competitive environment of the MP tasks, are repeatedly found to emit a common set of faulty intuitions about random sequences. One such intuition is that random sequences will be less likely to repeat the same item than to alternate to another item. This is termed the overalternation bias or negative recency effect. Participants are consistently found to judge sequences that alternate from one item to other items approximately 60% to 70% of the time as most random (for example, Bar-Hillel & Wagenaar, 1991; Falk & Konold, 1997; Scholl & Greifeneder, 2011; Zhao, Hahn, & Osherson, 2014). Sequences that alternate at this rate are frequently reported as more random than are even paradigmatically random sequences. And this report does not change when the sequences are presented in alternate ways, as in the randomly and nonrandomly generated pixel arrays and clocklike motion

of the stimuli of Zhao and colleagues (2014) or the checkerboard grids of Falk (1975, unpublished dissertation cited in Falk & Konold, 1997).

The overalternation bias is a more general version of the gambler's fallacy, the mistaken intuition that random sequences are self-correcting with regard to item frequency or other qualities (Lyons, Weeks, & Elliott, 2013). For example, a typical expression of the gambler's fallacy states that after a string of coin flips turn up heads, humans will often judge that the subsequent flip is more likely than not to be tails. Furthermore, the number of repetitions dictates the strength of the bias. For example, a random sequence that produces *Tails-Tails-Tails-Tails-Tails* will make a subsequent *Heads* outcome seem more likely than would a sequence that produces *Tails-Tails*.

A byproduct of this tendency to alternate is that occasions in which many of the same items appear in a sequence consecutively (hereafter 'runs') are less frequent, and less extreme, than should be expected in a random sequence. The length, frequency, and timing of these runs make a unique contribution to judgments of randomness (Sun & Wang, 2010). Run lengths and the frequency of alternation between stimuli are correlated, but not perfectly so. Scholl and Greifeneder (2011) used participant ratings of the randomness of sequences to determine which of the properties was more responsible for participants' judgments of the randomness of sequences. They found that the sequences rated as most random overalternated, but that this effect was weaker as the length of any individual run increased. In effect, the ratings suggested that overalternation was viewed as evidence of randomness and long runs of a single item were viewed as evidence for structure. Combined, the two properties do much to describe the characteristics that participants used to judge sequences as random or nonrandom.

Explaining the mechanism(s) for why such a bias would exist has proven more challenging. One line of inquiry seeks to use the qualities of the individuals who performed the randomness perception and production experiments to help explain their outputs. The statistical education and training of the participants would seem, at face, to have a large effect on their functional understandings of the concept of randomness. However, experiments that have compared statistically naïve participants and those with extensive experience with mathematics, probability theory, or statistics do not find large differences between the groups (Chapanis, 1995; Lopes & Oden, 1987; Nickerson & Butler, 2012). The cognitive capacities and psychopathologies of participants are of greater importance. The attention and working memory that participants apply to perception and production of random events play a critical role, as evidenced by attentional biases towards regularities (Zhao, Al-Aidroos, & Turk-Browne, 2013) and the ways participants detect patterns in sequences stored in working memory (Kareev, 1995a, 1995b, 2000; Kareev, Lieberman, & Lev, 1997). Nonrandom, but not necessarily overalternating, sequences are generated by autistic and intellectually disabled participants (Williams, Moss, Bradshaw, & Rinehart, 2002), participants with closed-head injuries (Azouvi, Jokic, Van Der Linden, Marlier, & Bussel, 1996), and participants afflicted by Parkinson's disease (Brown, Soliveri, & Jahanshahi, 1998), Alzheimer's disease (Brugger, Monsch, Salmon, & Butters, 1996), hemispatial neglect (Loetscher & Brugger, 2009) or alcohol dependency (Rosenberg, Weber, Crocq, Duval, & Macher, 1990), and schizophrenia (Rosenberg et al., 1990).

A classic review of this literature compiled by Tune (1964), and later updated by Brugger (1997), listed the numerous factors that contribute to human randomness perception and production and concluded that working memory limitations are the principal determinant of participant behavior with random sequences. When working memory was disadvantaged by

asking participants to produce random sequences rapidly (Baddeley, 1966) or by increasing the weight of task load that participants bear by imposing a concurrent serial recall, semantic category generation, digit generation, or fluid intelligence task (Baddeley, Emslie, Kolodny, & Duncan, 1998), the sequences that participants produced were less random. On this view, a random sequence is one that does not fail the participants' tests of randomness within the stretch of items stored in their working memories (Diener, 1985). These tests are not the formal tests described earlier, and largely seem to test that the frequencies of individual items in working memory are consistent with what the item frequencies should be for the sequence as a whole. The overalternation bias is therefore a specific version of the representativeness heuristic of Tversky and Kahneman (1974). Though a sequence that alternates more often than by chance and goes on fewer, shorter runs is nonrandom when considered as a whole, the shorter subsequences that participants evaluate in working memory are representative in that they are very likely to contain equal proportions of the different items and are unlikely to contain any long run that would indicate nonrandomness. Interesting supporting evidence comes from demonstrations that the "seven-plus-or-minus-two" capacity for short-term memory is in fact a sweet spot for the early detection, or lack thereof, of meaningful relationships (Kareev, 1995a, 1995b, 2000; Kareev et al., 1997). That is, Kareev and colleagues argued on a statistical basis that sampling 5 to 9 items is ideal (enough, but not excessive) for swift detection of strong correlations in sequences.

However, participants have the same erroneous beliefs about randomness when sequences are perceived in their entirety and working memory need not be engaged to the same degree in order to compile the sequence (Falk, 1975; Zhao et al., 2014). A second theoretical account of the overalternation bias thus focuses instead on the prospect of encoding sequences

into long-term memory. This encoding hypothesis posits that judgments of randomness are made when participants are unable to meaningfully encode a sequence into long-term memory (Falk & Konold, 1997). Falk and Konold demonstrated that both participants' subjective ratings of how difficult a sequence will be to encode and also their objective memorization performance better predicted judgments of randomness than did the objective properties of the sequence. The relationship between long-term memory and randomness judgments works in the opposite direction as well: Olivola and Oppenheimer (2008) demonstrated that when participants believed that a sequence was random, they later remembered it as having fewer runs and greater alternation. In synergy with the availability heuristic of Tversky and Kahneman (1974), runs of single items are easily chunked together and recalled from memory and therefore their presence or absence is disproportionately emphasized in judgments of randomness. Groups of items that are more difficult to encode (like those that frequently alternate) are less easily recalled, harder to describe, and more likely to be perceived as random.

Finally, Brugger (1997) offered a cheeky observation as a counterpoint to memory-based accounts of randomness understanding: coins and dice have no memory at all, and they are used to produce perfectly random stimuli. An alternative, or perhaps supplemental, hypothesis is that the overalternation bias is not only an issue of executive control or working memory limitations, but also reveals weaknesses in participants' beliefs about the properties of random events. This faulty concept hypothesis is exemplified by individuals' explicit adherence to the gambler's fallacy, and is part of a much larger family of mistakes that humans make in probabilistic scenarios (Falk, 2014; Nickerson, 2004; Tversky & Kahneman, 1974). The word "random" is poorly defined, used in a variety of contexts, and highly loaded, such that the mere inclusion of the term is enough to influence participants' memories and perceptions (Nickerson, 2002;

Olivola & Oppenheimer, 2008). In all of these tasks, participants were asked, either implicitly or explicitly, to express their idea of the concept of randomness. Their failures to be truly random may thus reveal less about inadequacies of memory or executive pattern detection and more about how poorly they understood what it means for an event to be random in the first place.

1.3 Randomness Failures in Broader Context

The specificity of the overalternation bias is curious in light of evidence from animals and a growing body of research on statistical learning.

1.3.1 Animal literature

Animals interact with structured sequences of information in many of the same ways as humans do (Conway & Christiansen, 2001). From repeated and prolonged experiences in an environment, they learn about the relationships between events, including relationships of a probabilistic nature. Animals must also deal with unstructured information. If the description of protean behaviors described earlier is accurate, animals reckon with sequences that are as close to random as can be approximated by other animals (Neuringer & Jensen, 2012; Sanabria & Thrailkill, 2009). Indeed, it is not difficult to position many ecological decisions as analogous to the choice presented to humans in the discrimination experiments presented earlier (e.g., was the process that generated these events random?). An animal that puts itself in a position to learn underlying relationships of a structured environment will reap the rewards of this learning. Animals in environments with more arcane structure, or no structure at all, will not have access to this reward.

Experimental evidence from the MP design described earlier has demonstrated that animals can at least match the variability of their behavior to the level of variability in their environments. This involves two processes: ascertainment of the level of variability in the

environment and generation of responses that are sufficient to exploit the environment. To explore these capabilities, Barraclough, Conroy, and Lee (2004) analyzed the performance of rhesus macaques (*Macaca mulatta*) on an oculomotor MP task. The monkeys' "competitors" in this task were computer-simulated opponents with different properties. One simulation produced responses randomly. Another analyzed the animals' choice histories (from the last 0 to 4 trials), tested for sequential dependencies in this subset of trials, and weighted responses accordingly. For example, if a significant bias was found in the monkeys' choices (for example, 80 percent of responses were to the right) the simulation chose the opposite option at the same rate to exploit this tendency. The final simulation analyzed both the choice history and reward history of the animals. Contrary to human performance in MP games, in which the overalternation bias is muted but still present (for example, in Rapoport & Budescu, 1992), the monkeys' responses were very nearly optimal for all three simulations. The monkeys' sequences of responses were not random, as an inclination to win-stay, lose-shift was present throughout and only exploited by the third simulated opponent. But as the simulations increased in complexity, animals matched this variability in kind to remain at the Nash equilibrium for the game.

Sanabria and Thrailkill (2009) used a MP game with pigeons (*Columba livia*) in direct competition with other pigeons. Each animal in a competitive pair was placed in an operant chamber with two response options. Each day, a pigeon was rewarded for either matching the responses of the opponent or for choosing the opposite response. Unlike humans and rhesus macaques, pigeons demonstrated the opposite bias towards response perseveration. Indeed, this bias towards perseveration was to such robust degree that it was included as a parameter of models of pigeon MP performance. With regard only to the capacity to generate random behaviors, however, pigeons can be exceedingly variable. Experimental environments that

systematically reward animals for generating sequences that have not been generated within the last *n* trials can shape pigeons' responses into a form indistinguishable from the output of a random Bernoulli process (for a review, see Neuringer, 2002).

The capacity to produce highly variable, highly unstructured sequences is latent within these animals, and perhaps in humans as well (Neuringer, 1986). At issue is why baseline human performance would be so markedly different from that of other species tested on similar tasks with regard to the specific overalternation bias. Whereas pigeon and rhesus monkey performances on MP tasks were different, the species both do not have an overalternation bias and could not be argued to be expressing their erroneous idea of a higher-order concept (as in the faulty concept hypothesis of human foibles with random stimuli). At face, the faulty concept hypothesis is supported, yet other explanations exist. Perhaps, through difference in mechanism or sheer number of trials, the animals received a much more robust opportunity to learn the structure (or lack thereof) of their environments than did humans.

1.3.2 Statistical learning

Humans, and likely many animal species, implicitly internalize the embedded structures of their environments via the mechanism of statistical learning. From repeated exposures to sequences that have an embedded structure, these species learn about the regularities present in their environments and use this learning adaptively. These regularities can take a number of forms. In humans, statistical learning is prominent as a mechanism by which children acquire the phonetic and syntactic regularities of their native languages (Kidd, 2012; Saffran, Aslin, & Newport, 1996). But it need not only be a mechanism for human language learning. Other species – rhesus macaques (Heimbauer, Conway, Christiansen, Beran, & Owren, 2012b), cotton-top tamarins (*Saguinus oedipus*, Locurto, Dillon, Collins, Conway, & Cunningham, 2013;

Locurto, Fox, & Mazzella, 2015; Locurto, Gagne, & Nutile, 2010), pigeons (Locurto et al., 2015), and rats (Toro & Trobalón, 2005) – have demonstrated substantive levels of statistical learning as well. Furthermore, stimuli presented in other modalities are subject to statistical learning (Conway & Christiansen, 2005, 2006). Statistical learning is a general mechanism by which humans and nonhumans learn the meaningful structure of events and objects in their environments.

The statistical learning literature reveals a robust ability for humans to learn the regularities and patterns of auditory and visual stimuli to which they are repeatedly exposed. Critically, these exposures do not require conscious attention nor explicit reward. When participants were given no explicit instruction other than to look at a screen, they nevertheless implicitly learned the underlying regularities of the stimuli presented onscreen and reported its familiarity (Fiser & Aslin, 2002). Moreover, participants learned statistical structure that was embedded in background noise even as they concurrently completed a primary task (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). In a series of experiments, Turk-Browne, Jungé, and Scholl (2005) demonstrated that statistical learning both is and is not automatic: Some level of selective attention is required, but the process is still largely unconscious, passive, and implicit. This finding is bolstered by fMRI evidence that suggests that statistical learning occurs even without explicit awareness (Turk-Browne, Scholl, Chun, & Johnson, 2009). Parallel research with nonhuman primates reveals a similar pattern of results. Locurto and colleagues demonstrated statistical learning in cotton-top tamarins even when the animals' responses were divorced from pellet rewards (Locurto et al., 2013) and specific motoric behaviors (Locurto et al., 2010).

1.4 Randomness and Structure

Randomness is, in many ways, a very difficult concept for people to comprehend. They have pervasive biases about what the properties of randomness are and are not, and maintain these biases despite a lifetime of evidence that stimuli like coin flips or dice rolls are exactly as paradigmatically random as they are purported to be. In some ways this is unsurprising. Though the costs of holding mistaken beliefs about randomness are difficult to quantify, they are unlikely to be especially severe (Kac, 1983). Indeed, they may actually be adaptive, and allow for useful, early pattern detection in noisy sequences (Kareev, 1995a, 1995b, 2000; Kareev et al., 1997). Finally, probability theory is a high-level branch of mathematics and few people, even in university samples, would be expected to have a substantive level of exposure to it. Individuals are not described as 'biased' or even 'irrational' for failures to understand other high-level concepts in mathematics, so it perhaps makes little sense to fault people for their weak understandings of probability theory and randomness (Goldman, 1986).

But the discrimination that must be made is between structure and a lack of structure in the environment, and a review of the statistical learning and randomness literatures reveals striking inconsistencies. Though humans and animals have a robust ability to learn exceedingly subtle properties of sequences implicitly in statistical learning tasks, they stubbornly maintain faulty intuitions about the properties of random sequences when asked to discriminate or produce random stimuli. Statistical learning methodologies give participants much greater exposure to sequences, and perhaps this explains the difference. Were random stimuli to be presented in this mode, with this number of exposures, participants may adaptively use a more correct understanding of random sequences.

In this pair of experiments, I presented different types of sequences – some that were paradigmatically random, some with consistent structure, and some with those properties that participants typically report as random (e.g., overalternation, few long runs) – in a common statistical learning paradigm. Participants had the ability to choose the type of sequence that they would work on for each block of trials. This methodology allowed me to explore humans and nonhumans' ability to find and learn the underlying structures of their environments, and to analyze how this learning affects what sequences they chose to exploit. This methodology, and its use with both humans and animals, also has implications for the encoding and faulty concept hypotheses of randomness failures.

Wilke and colleagues have used a similar methodology with undergraduate participants, community participants with problem or pathological gambling traits, and rhesus macaques (Blanchard, Wilke, & Hayden, 2014; Scheibehenne et al., 2011; Wilke, Scheibehenne, Gaissmaier, McCanney, & Barrett, 2014). In these studies, participants were given a series of choices among simulated slot machines that either produced random outcomes, sequences that alternated between outcomes more often than by chance, or sequences that were less likely to alternate between outcomes and more likely to go on streaks of a single outcome. In each population, participants exhibited a preference for the slot machine that alternated least and streaked most. In addition, participants made little distinction between a random slot machine and one that alternated more than chance, and in some cases even chose the random machine at a higher rate than the overalternating machine.

The unique contribution of the following experiment was simultaneously to assess participant discrimination between random and structured sequences and implicit learning of underlying statistical structure in sequences. Unlike in the experiments of Wilke and colleagues,

participants received a great deal of experience with all sequence types and responses were divorced from specific rewards. Participant learning of the underlying sequence properties was assessed on the basis of decreasing response times with experience with structured sequence variants. Participant choice behavior was therefore determined by two factors: (1) a subjective judgment on the part of the participant of the structure, patterning, ease of responding, or predictability of the presented sequences, and/or (2) implicit, statistical learning of the underlying sequence properties. This experiment was designed to serve as a novel probe of how these two factors interact (see Table 1 for Research Questions and Hypotheses).

2 EXPERIMENT 1

2.1 Methods

2.1.1 Participants and Apparatus

Thirty-four Georgia State University undergraduate psychology students were invited to participate in the study for course credit. Completion of the task took 30 minutes to an hour. This research was approved by Georgia State University's Institutional Review Board. Testing took place in laboratory space on the Georgia State University campus. The computer monitor was positioned approximately 46 cm from the participant, with the joystick positioned on a table immediately in front of the participant.

2.1.2 Serial Reaction Time task

The Serial Reaction Time (SRT) task is a relatively simple task that has long been used in sequence learning and statistical learning experiments (Hale, 1969; Nissen & Bullemer, 1987). Targets appear on a computer screen in multiple locations in succession, and participants must rapidly respond to the location of the target. When an underlying structure dictates the locations of the targets, participants can learn this structure and respond more quickly to correctly-

Table 1 Research Questions and Hypotheses

	Research Question	Hypothesis	Implication
1	RQ1: Do participants learn the structure of sequences presented in the mode of SRT tasks?	H1: Exposure to the task will significantly predict reduced reaction times for structured sequence variants relative to those of the random sequence. RTs for nonrandom stimuli that do not decrease over time relative to RTs for random stimuli will be interpreted as evidence that the participants failed to learn the sequence, and thwart interpretation of SELECT choice behaviors dependent on sequence learning.	Participants learn the properties of binary sequences they experience in SRT tasks, and perhaps in other modes of binary sequence presentation (e.g., the perception and production experiments reviewed above).
2	RQ2: Do participants act on the basis of experience in the task by distinguishing and preferentially choosing an explicitly observable structure?	H2: Exposure to the task will significantly predict choices of the sequence that is the same each time over a sequence that is random. Failure to do so may indicate a lack of preference, or lack of motivation to act on preference.	That organisms would selectively sample from more structured stimuli would be a novel finding, indicating a preference for structure, information content, or predictability of sequences.
3	RQ3: What are the relative contributions of intuitions about random sequences and statistical learning of sequence properties on human's choices?	H3: Exposure to the task will significantly predict human participants' choices of the the random sequence over the sequence with a statistically embedded, but not explicitly observable, structure. Failure to do so may indicate a lack of preference, lack of motivation to act on preference, or, in the case of preference for the statistically structured sequence, indication that implicit statistical learning contributes to explicit choice behavior.	This pattern of responses would indicate that sequence choices conform to subjective reports of what seem like random sequences despite demonstrable learning of the underlying structure the nonrandom sequence (H1).
4	RQ4: Do monkeys share humans' intuitions about sequence structure?	H4: Exposure to the task will significantly predict macaques' preferentially chosing structured sequence types over a random sequence. Failure to do so may indicate a lack of preference, lack of motivation to act on preference, or, in the case of preference for the random sequence, indication that intuitions about random sequences contribute even to monkey choice behaviors.	In conformation with the faulty concept hypothesis, monkeys' choices will be made strictly according to learning of sequence properties, and they will be indifferent to the overalternation and lack of runs of the statistically structured sequence.

predicted targets. Reaction times can be used as an index of the degree to which participants have learned the properties of the sequence of targets and of their ability to predict any given target. This methodology was employed for its simplicity, its visual and nonverbal nature, and its previous application to nonhuman primates (e.g., Heimbauer et al., 2012a). This experiment used an SRT task with two possible target locations, consistent with the frequent use of binary response options in the randomness literature (for examples, see Nickerson & Butler, 2012; Scholl & Greifeneder, 2011; Zhao et al., 2014). Targets appeared on either the left or right of the screen according to one of three algorithms:

1. A **random** algorithm that dictated target locations pseudorandomly using the `random.choice()` function from the `random` module of python 2.7. The output is considered pseudorandom for its use of the Mersenne Twister as the generator, but can be considered random to the human and monkey observers of this experiment (Python Software Foundation, n.d.).
2. A **fixed** algorithm that dictated target locations absolutely. On each selection of this algorithm, the same sequence is presented. This sequence was produced by the author using the 'Coin Flipper' utility of random.org.
3. An **overalternating** algorithm that has specified properties, and thus is structured, but that has those properties that humans commonly report of random stimuli. Targets generated by this algorithm alternated 65% of the time and were disproportionately unlikely to go on runs of more than a few items (for a detailed explanation of this algorithm, see Appendix A).

Participants made joystick deflections in the direction of the target within 2 seconds of target onset. To discourage holding the joystick in a single direction (thereby simplifying the

task), targets did not appear on the screen so long as the joystick remained depressed in a single direction. Deflections of the joystick in the opposite direction of the target did not end the trial, but negated joystick responses for 1.5 seconds (and thus sharply decreased the amount of time in which a correct deflection could occur). The total trial duration was twice the sequence length, in seconds (42 seconds). Failure to deflect the joystick properly within the 2 second response window resulted in a blank screen that persisted for the remainder of the trial duration. For example, if a participant received a blank screen after 8 seconds of a 21-item sequence, they waited 34 seconds before starting a new trial.

2.1.3 SELECT task

The SELECT task was originally developed to offer nonhuman primates the opportunity to choose the order in which they complete a series of tasks (Perdue, Evans, Washburn, Rumbaugh, & Beran, 2014; Washburn, Hopkins, & Rumbaugh, 1991). In this experiment, humans were offered a choice among the different algorithms of the SRT task. The SELECT menu appeared with two of three symbols signifying the three algorithms— the alchemical symbol for sulfur for random sequences, the alchemical symbol for mercury for fixed sequences, and the astrological symbol for Jupiter for overalternating sequences (see Figure 1) - in counterbalanced locations. Choice of one of three symbols resulted in presentation of SRT targets, presented visually as the same symbol that was chosen on the SELECT screen. After the completion of this preferred 21 item sequence, the menu reappeared with another selection of symbols. To distinguish the choice of the SELECT menu from the iterative responding required of the SRT task, presentation of the menu was accompanied by a change in background color of the task (see Figure 2). To ensure that even less-preferred algorithms were experienced, the SELECT screen also offered a forced choice of only one of the algorithms on 20% of trials.

Stimulus	Description
	Alchemical symbol for sulfur; Used in selections and presentations of random sequences.
	Alchemical symbol for Mercury; Used in selections and presentations of fixed sequences.
	Astrological symbol for Jupiter; Used in selections and presentations of overalternating sequences.

Figure 1. Task Stimuli

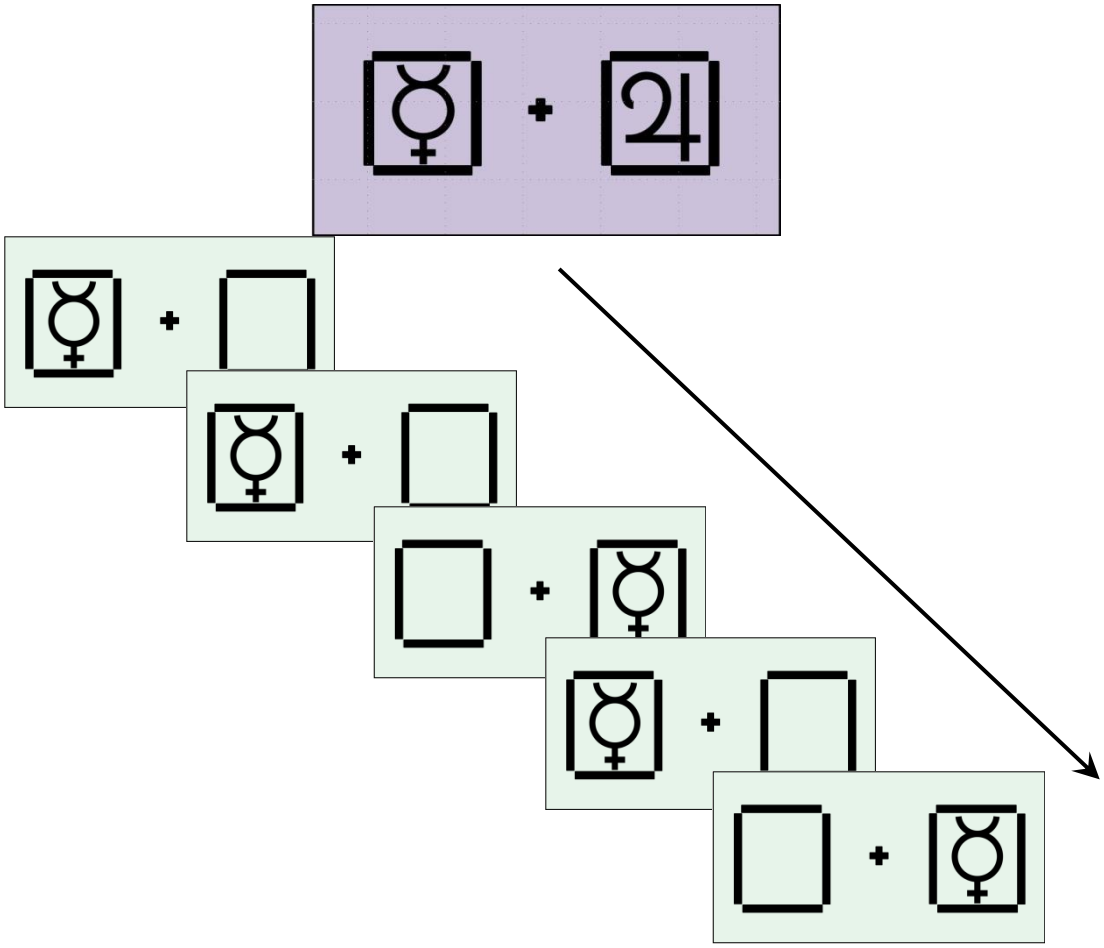


Figure 2. Task appearance. The top image is an example of a SELECT screen with options to choose either the fixed (left) or overalternating (right) sequence. The bottom set of images illustrates the way the first five stimuli of a fixed sequence would appear onscreen

2.2 Results

In this SRT design, experience may be defined as the number of successful deflections a participant had prior to the current exposure or as the total exposures to stimuli a participant's program provided. The difference between the two measures is in whether the participant's behavior is observed or assumed. A successful deflection constitutes a record that the participant observed the stimulus, whereas the count of stimulus transitions also credits the participant with gaining information from stimuli to which they did not deflect. The former definition is more conservative, and is used for all analyses.

Increased exposures to the different sequence types did not change participant reaction times in meaningfully different ways (see Figure 3). The characteristic shape of the RT curve is visible for all three sequence types in the upper plot of Figure 3, with RTs that are relatively high in earliest trials yet rapidly asymptote at a lower mean RT. The grand means of participants' mean RTs for the different sequence types were nearly identical overall (fixed = 381 ms, overalternating = 383 ms, random = 386 ms). When comparing RTs to stimuli appearing before and after the participant's 175th exposure to a sequence type (an approximation of the midpoint of the session), participant mean RTs decreased on average by 10 ms for fixed stimuli (*SD* = 55), decreased by 3 ms for overalternating stimuli (*SD* = 43), and increased by 2 ms for random stimuli (*SD* = 55). These differences represent a negligible change in RT with experience in the task.

Choice behaviors in the SELECT task also did not change in the predicted directions. Visualizations of the change in proportion of sequence types chosen at different blocks of 42 relevant stimulus exposures is presented in Figure 4. In this figure, exposures to both sequences presented on a given SELECT screen (for example, both experience with both fixed and random

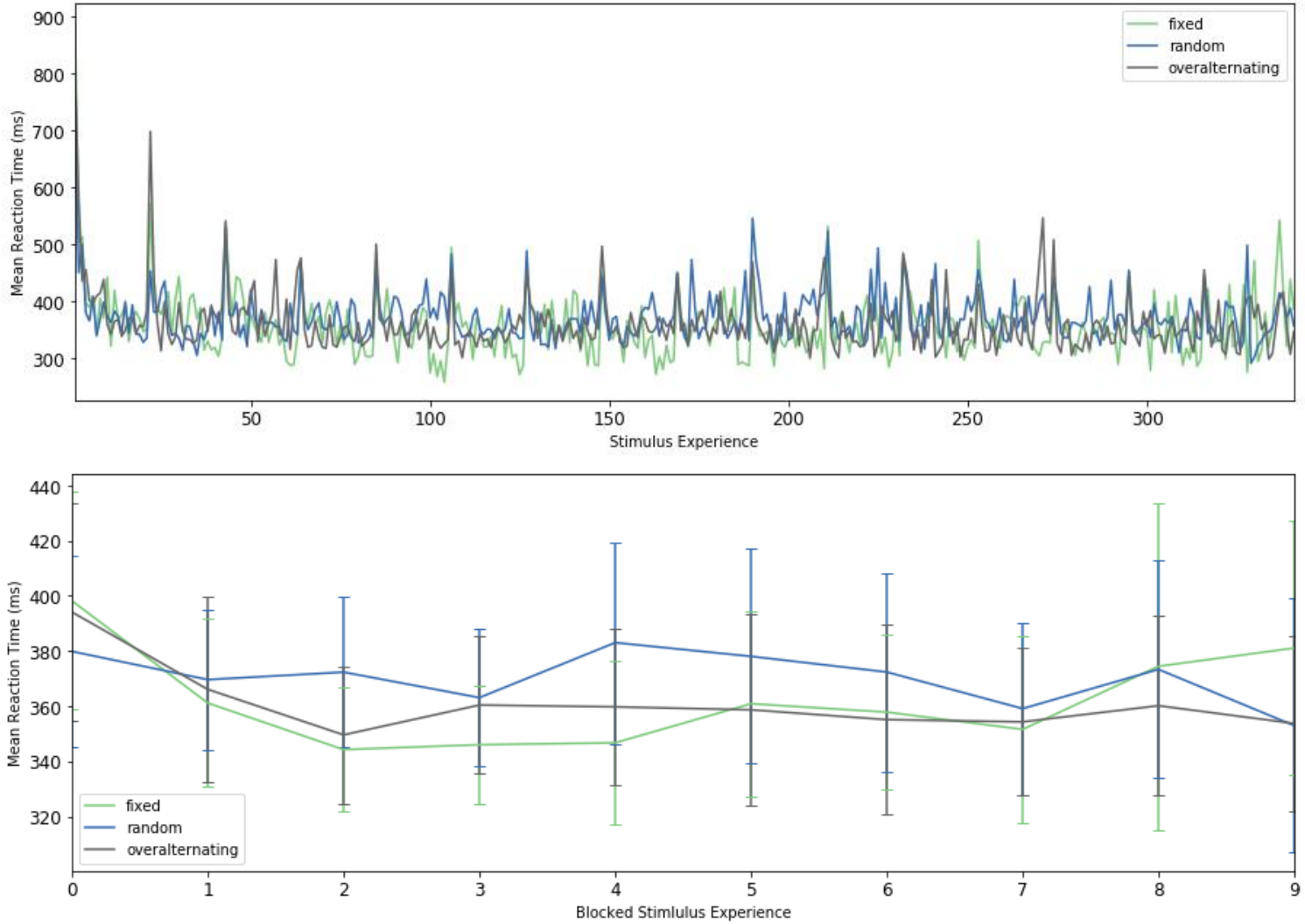


Figure 3. Mean RTs aggregated across participants. The lower plot aggregates sequence exposures into blocks of 42 and includes Standard Error bars. Note that due to participant control over how many sequence exposures they received, not all participants are included in each aggregation.

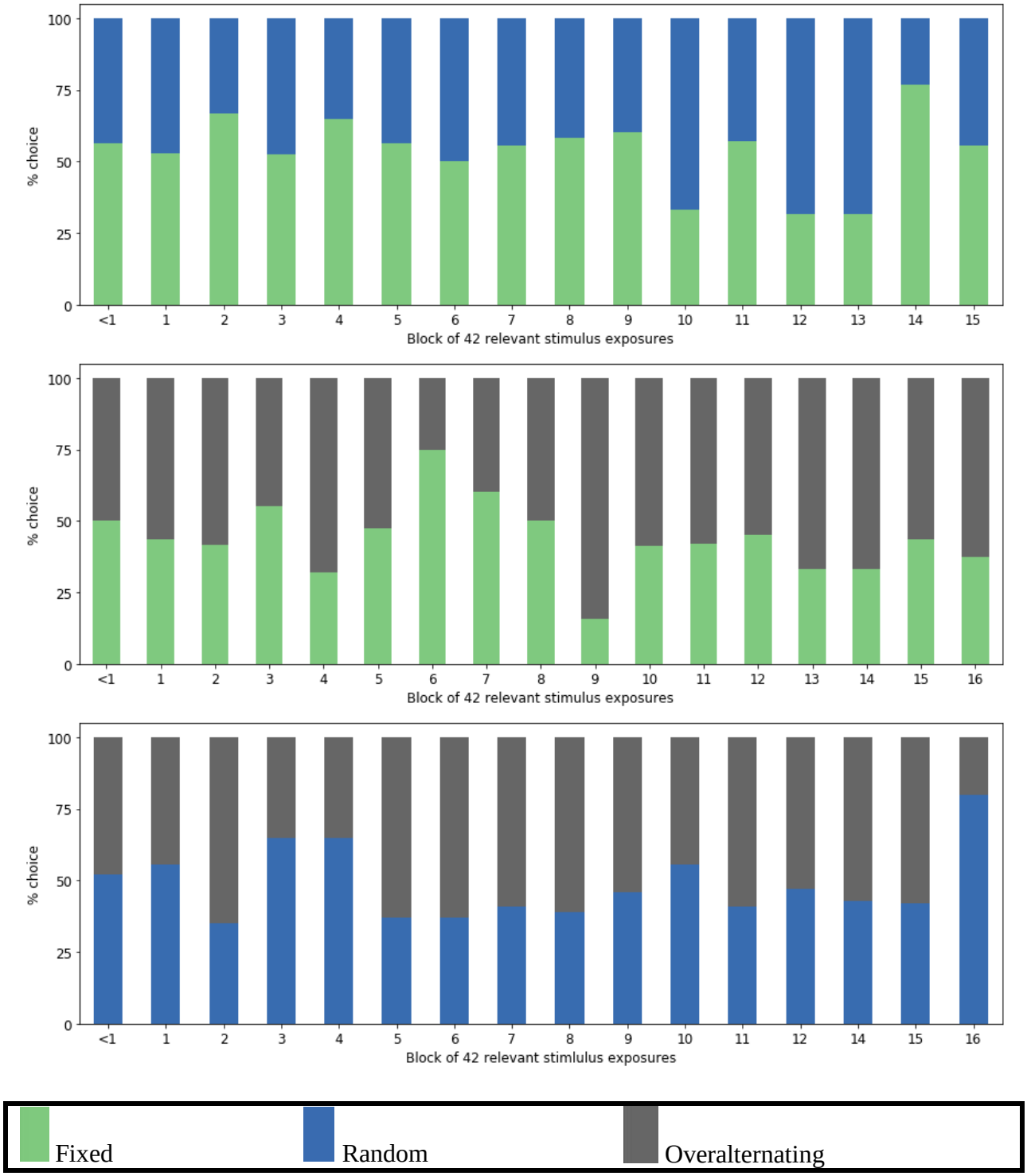


Figure 4. Aggregated participant choices in different SELECT menus across different levels of relevant experience with the task. Blocks are the size of two completed trials (42 stimulus exposures). Note that due to participant control over how many sequence exposures they received, not all participants are included in each aggregation.

sequences) are blocked. Predominance of the fixed sequence (dark blue bar) over alternatives, an assertion of Hypothesis 2 and important bellwether for viability of interpretations of Hypothesis 3, was not observed. To clarify the effect that experience with the different sequence types had on choice behaviors and reaction times, mixed-effects models were used.

2.2.1 SRT modeled

SRT data were analyzed by fitting a linear mixed-effects model to predict natural log-transformed reaction times (hereafter, log(RT/ms)) for a trial from the trial's sequence type and how much experience a participant previously had with that sequence type. Arbitrary names that identified participants were included as random effects in the model to control for predictable individual differences of each participant. Mixed-effects modeling techniques are relatively tolerant of discrepancies of numbers of observations, and thus were used to account for the disparate numbers of trials for each sequence type that participants completed (given that participants had agency over the number and types of trials they experienced). Models were fit using the Statsmodels (Seabold & Perktold, 2010) and Pandas (McKinney, 2010) packages for Python 2.7 (Python Software Foundation, n. d.).

The results of the linear mixed-effects model (*AIC* = 9644.96, *df* = 16) fit for SRT are presented in Table 2. The fixed effects of the fixed sequence type (β = -.006, *SE* = .007, *p* = .402), the overalternating sequence type (β = .004, *SE* = .007, *p* = .609), and the interaction term representing the effect of increasing experience with the fixed sequence (β = -.00004, *SE* = .00003, *p* =.140) were not significant. The effect of greater experience with sequences (β = -.0002, *SE* = .00008, *p* = .008), and the interaction representing the effect of increasing experience with the overalternating sequence type (β = -.00006, *SE* = .00003, *p* = .020) were significant.

The coefficient of the significant interaction term of experience in the overalternating

Table 2. Linear Mixed-Effects Model Fit

Parameter	B	*SE*	*p*
Intercept	5.914	.024	<.001
Experience	-.0002	.00008	.008*****
Fixed	-.006	.007	.402
Overalternating	.004	.007	.609
Fixed x Experience	-.00004	-.00004	.140
Overalternating x Experience	-.00006	-.00006	.020*****

*Notes.***p* < .05. The intercept is the grand mean for log(RT/ms) across sequence types. The *p* values are computed from Wald z-tests testing whether the model coeffeicient is significantly different from zero.

sequence (bottom plot of Figure 5) can be interpreted to mean that 100 trials of experience with the overalternating sequence constitutes a .6% decrease in log(RT/ms). For reference, a .6% decrease in the grand mean log(RT/ms) in these data, 5.914 (an RT of approximately 370 ms), would constitute a 23 ms decrease in RT. The effect of the nonsignificant interaction term of experience in fixed sequences (top plot of Figure 5) would constitute a .4% decrease in log(RT/ms) for 100 trials of experience in the fixed sequence, a difference from the grand mean of approximately 9 ms. The interaction term of experience in random sequences (middle plot of Figure 5) would constitute 1% increase in log(RT/ms) for 100 trials of experience in the random sequence, a difference from the grand mean of approximately 22 ms. Figure 6 illustrates the overall fit of the estimations of model to the data. The relatively normal distribution of the deviations of predicted values from observed around zero indicates a relatively clean fit of model to data. The discrepancy between model and data at higher numbers of sequence exposures (see

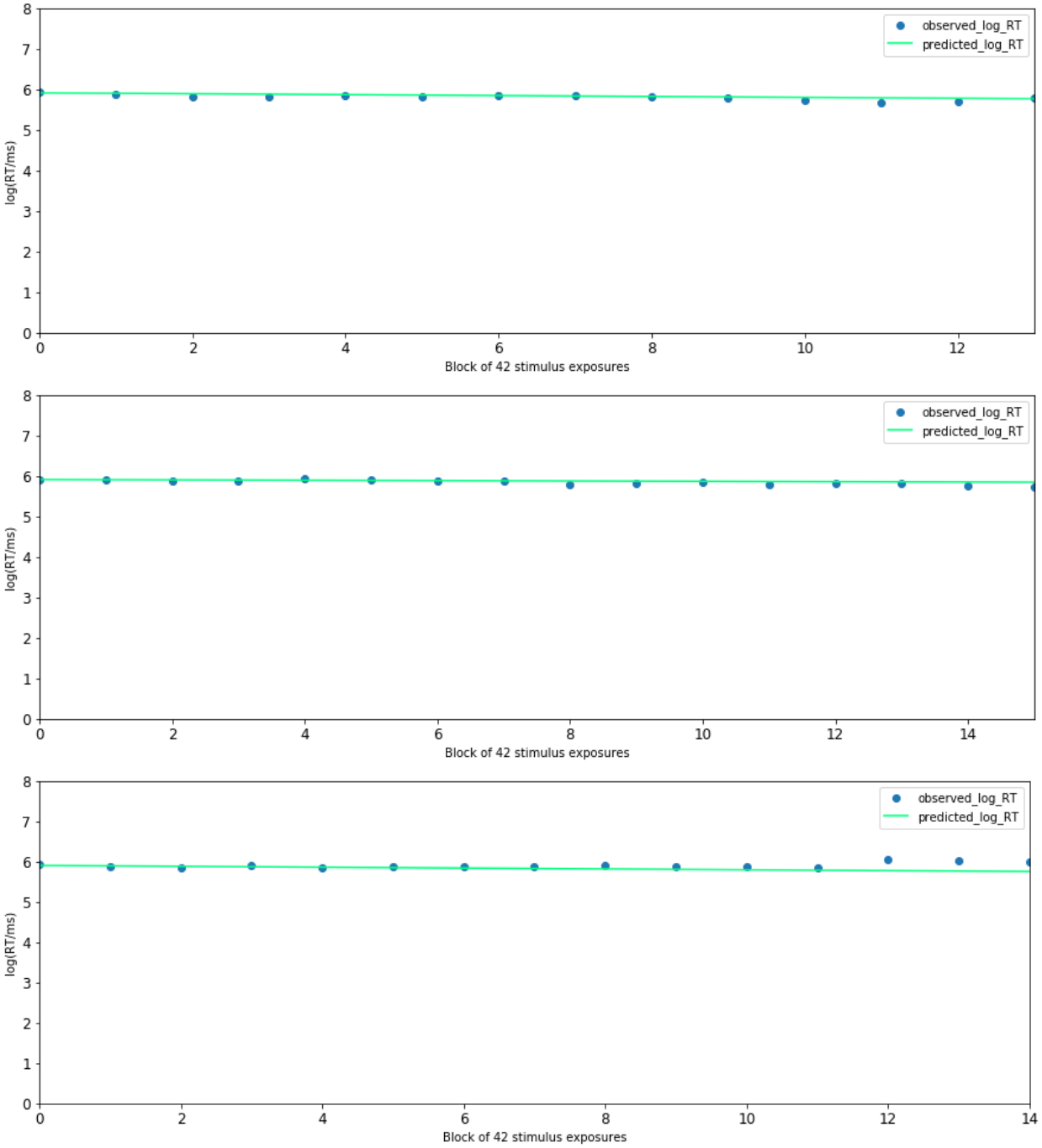


Figure 5. Averaged observed log(RT/ms) for humans across increasing blocks of experience with the fixed sequence (top), random sequence (middle), and overalternating sequence (bottom). Each block was equal to two full sequences, 42 stimulus exposures.

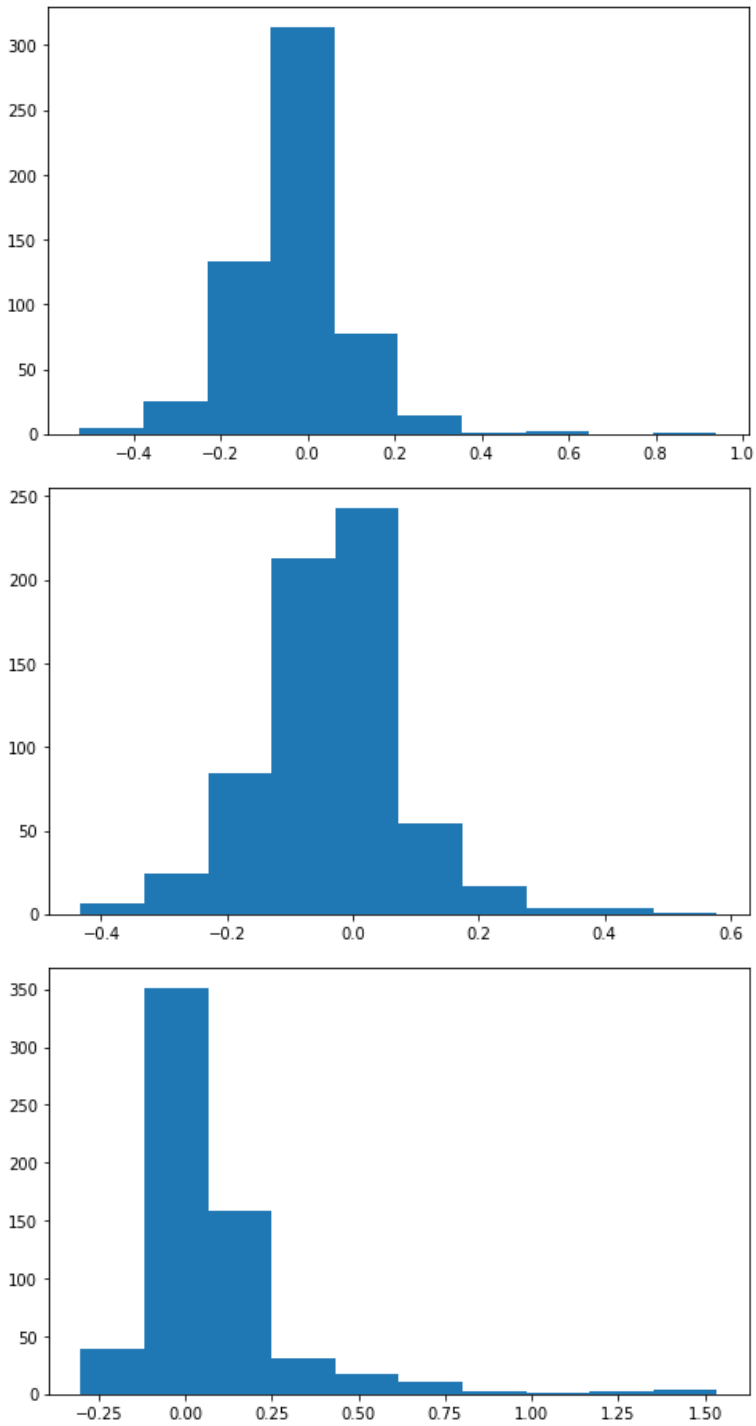


Figure 6 Visualization of model fit: histograms of frequency of deviations of predicted log(RT/ms) from observed at different levels of sequence experience in the the fixed sequence (top), random sequence (middle), and overalternating sequence (bottom). Positive values indicate that predicted values were higher than observed values.

the right side of plots of Figure 5) are likely the result of only a subset of participants experiencing these trials, and of the modeling procedure assigning greater weight to earlier trials with higher participant numbers. Visualizations of aggregations of the estimated random effects parameters are presented in Figure 7. These histograms illustrate the degree to which individualized parameter estimates for each participant improved model fit. The seemingly bimodal distribution around zero for the estimated random effect of experience (second plot from top) illustrates that one subset of participants' RTs increased with experience, while another subset of participants' RTs decreased with experience. This parameter estimate might capture a conflict between learning of sequence properties and boredom due to extensive time working on the task, an idea that will be considered further in the General Discussion.

2.2.2 SELECT modeled

SELECT data were analyzed by fitting a generalized linear mixed-effects model to predict binomial participant choices of the more structured option from the options available on the menu and the amount of experience the participant had with the options available on that SELECT menu. For example, a participant choosing the structured option (say, the fixed sequence) was modeled by what the less structured alternative was (say, the random sequence), and the total number of successful deflections the participant had with fixed and random sequences. As with the SRT data, arbitrary participant identifiers were included as random effects in the model to control for any individual differences across participants. The coefficients of greatest interest, the interactions effect of greater experience that is relevant to a SELECT menu on choices in that menu, were not significant (fixed v random menu x experience: β = .0004, SE = .0004, p = .335; fixed v overalternating menu x experience: β = .0002, SE = .0004, p = .626). No variable significantly predicted the log odds of choosing the structured or

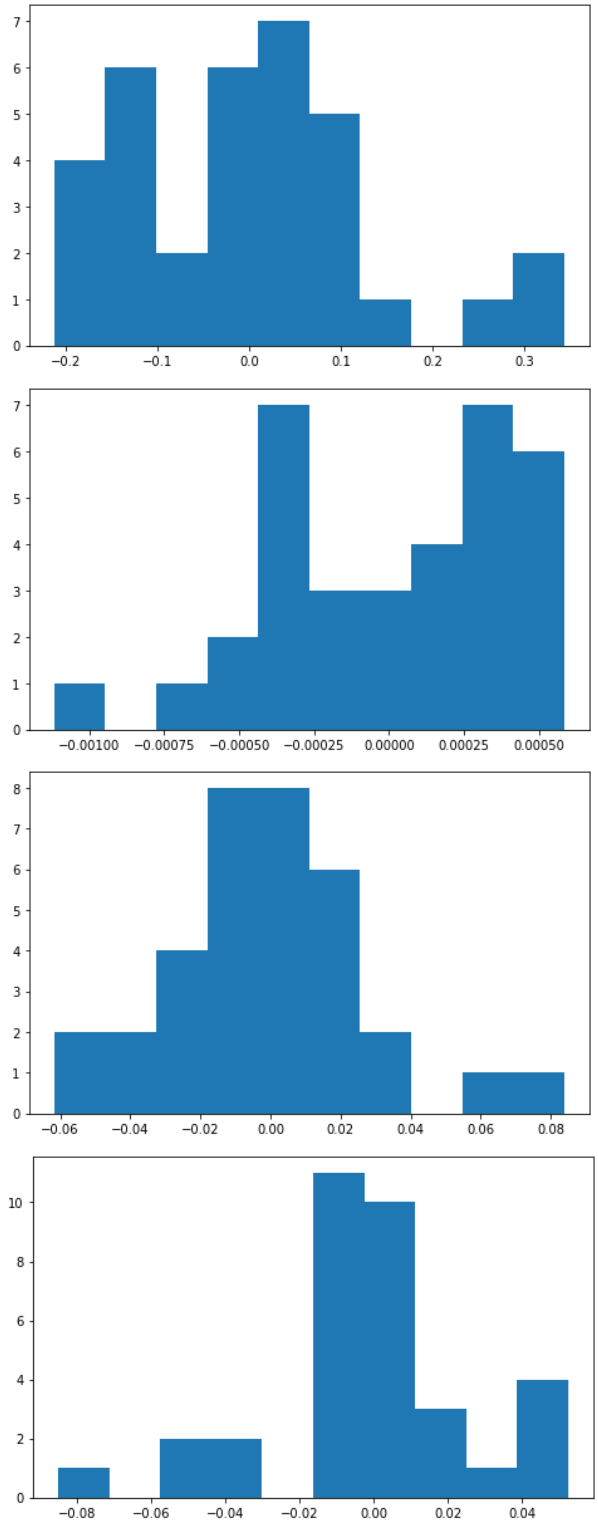


Figure 7. Histograms of individual participants' parameter estimations for the grand mean log(RT/ms) (top), the effect of increasing experience with the different conditions (second from top), the effect of the sequence being fixed (second from bottom), and the effect of the sequence overalternating (bottom).

unstructured option on the SELECT menu.

2.3 Discussion

Participants did not show any evidence of change in sequence choice behaviors across experience in the task, even as reaction times subtly decreased with greater experience in the overalternating sequence and increased with greater experience in the random sequence. This result may suggest an indifference on the part of the participants to the differences in relative information content, statistical regularities, and potential completion speed across the different sequence types.

Some considerations burden this interpretation. A statistically observable decrease in reaction time to stimuli from the fixed sequence, a sequence that is identical every time, was an assumption of task design that was not evidenced in these data. The statistically significant effect that was observed, of decrease in time to complete the overalternating sequence, was slight.

3 EXPERIMENT 2

In an effort to make the features of the fixed sequence more explicit, a new sample of human participants were recruited for a second experiment. In this experiment, the sequences were shortened from 21 to 9 (with the trial duration shortened from 42 sec to 18 sec, accordingly). Shortening the sequences yielded several advantages. The working memory load to encode the sequence properties was much less substantial, with greater proportions of the sequences able to be tracked at once and more frequent breaks (through more frequent intertrial intervals (ITIs) and SELECT menu presentations). Participants could make far more choices on the SELECT menus, a relevant dimension if choice preferences are the result of mapping choices to sequence outcomes rather than associative experience to stimuli. More SELECT menu choices also yielded more data, and more opportunities for changes in choice behavior to emerge and be

observed. Finally, 9 item sequences were a more appropriate sequence length for completion by the rhesus monkeys.

3.1 Methods

3.1.1 Participants

Twenty-nine Georgia State University undergraduate psychology students participated in the study for course credit. Ethical considerations and apparatus were identical to those of Experiment 1.

In addition, five rhesus macaques were tested in their home enclosures at the Language Research Center of Georgia State University. The monkeys used a version of the Language Research Center – Computerized Testing Apparatus to make responses (Richardson, Washburn, Hopkins, Savage-Rumbaugh, & Rumbaugh, 1990). Stimuli were presented on a computer monitor in front of the animals' home enclosures and responses were made using joysticks mounted to the front of the animals' home enclosures. Food rewards of 45 mg banana-flavored fruit pellets were dispensed by a pellet dispenser. This population of animals had extensive experience with use of this apparatus to complete a variety of different computer tasks including serialized economic games (e.g., Parrish, Brosnan, Wilson, & Beran, 2014), a computerized Monty Hall problem (Klein, Evans, Schultz, & Beran, 2012), numerical judgments (e.g., Harris, Gulledge, Beran, & Washburn, 2010), and, of particular note, a Serial Reaction Time task (Heimbauer et al., 2012a and Heimbauer et al., 2012b). In the work by Heimbauer and colleagues, a subset of this population of animals demonstrated the ability to complete familiar sequences up to 8 items long more quickly than random sequences, and to complete grammatical 4, 6, and 8 item sequences more quickly than control sequences.

3.1.2 Training

Because the monkeys were less accustomed to tasks in which many responses are required before reinforcement, only some animals had previously experienced an SRT task, and those animals that had previous experience with an SRT task had not seen the task in several years, all monkeys were first presented with a training version of the task. In this version of the task, the random algorithm was used to generate one item sequences. If seven out of an animal's previous ten trials were successfully completed, the sequence length increased to three. Each subsequent time that this criterion was met, the sequence length increased by three, up to a maximum of 21. Once an animal successfully completed 15 trials of this length in total, it was transitioned to the full version of the task. If seven out of an animal's previous ten trials were not successfully completed, the sequence length decreased by three. After 15 repetitions of the sequence length incrementing to a number, then decrementing from it, the animal was considered to be at asymptote and was transitioned to the full version of the task at the highest number of deflections to reward at which it established proficiency. This criterion was established in order to limit any exposure to the task above and beyond what was necessary for training, and prevent overtraining with an exclusively random version of the task.

Completion of a sequence through consecutive, timely deflections rewarded monkeys with a number of 45 mg pellet rewards commensurate with the length of the sequence completed (sequence lengths of 1 and 3 yield 1 pellet; lengths of 6 and 9 yield 2 pellets; lengths of 12 and 15 yield 3 pellets; lengths 18 and 21 yield 4 pellets), and an ITI equal to the difference between the maximum sequence length-times-two and the time it took to finish the sequence. For example, completing a 21-item sequence in 35 seconds would yield 4 pellets and an ITI of 7 seconds.

3.1.3 Training results

This criterion for establishing asymptotic performance for the monkeys was excessively strict. No monkey was able to pass criteria above a sequence length of one before being judged as asymptotic. For this reason, the cap for asymptote was quadrupled, to 60 repetitions of the the animal incrementing to a new sequence length, then decrementing from it. To continue to limit excessive experience with the training version of the task, monkeys were automatically moved to the full version of the task once proficiency was established with sequences of length six. Three pellets were awarded for completing six item sequences, two pellets for three item sequences, and one pellet for one item sequences. Some descriptive information on each monkey's training history is presented in Table 3 and in Appendix B. All monkeys passed these new training criteria, and completed tens of thousands of joystick deflections in the task.

3.2 Results

Human RTs decreased in a similar way as in Experiment 1, with high RTs on initial trials sharply decreasing to asymptote (see the upper plot of Figure 8). Grand means of individual participants' mean RTs for fixed, overalternating, and random sequences were nearly identical (413 ms, 409 ms, and 416 ms, respectively). The change in mean RTs to fixed stimuli from first half of sequence exposures (100 sequence exposures was the approximate midpoint in this dataset) to the second half doubled from Experiment 1, to a 20 ms decrease (*SD* = 49). Changes in mean RTs across the two halve s of the experiment for overalternating stimuli (mean increase of 1 ms, *SD* = 51) and random stimuli (mean increase of 7 ms, *SD* = 35 ms) were similar to those of Experiment 1.

Unlike the RTs of human participants in Experiments 1 and 2, monkey RTs were volatile, even when aggregated across 1000 or more stimulus exposures. Figure 9 shows the five

Table 3. Training Summary

Monkey	Training Sessions	Experimental Stimulus Exposures			
		Fixed	Overalternating	Random	Total
Hank	25	18708	13888	29089	61685
Murph	8	31856	28204	16640	76700
Lou	6	22300	22886	24462	69648
Chewie	9	31573	34342	33903	99818
Obi	7	24106	20439	20088	64633

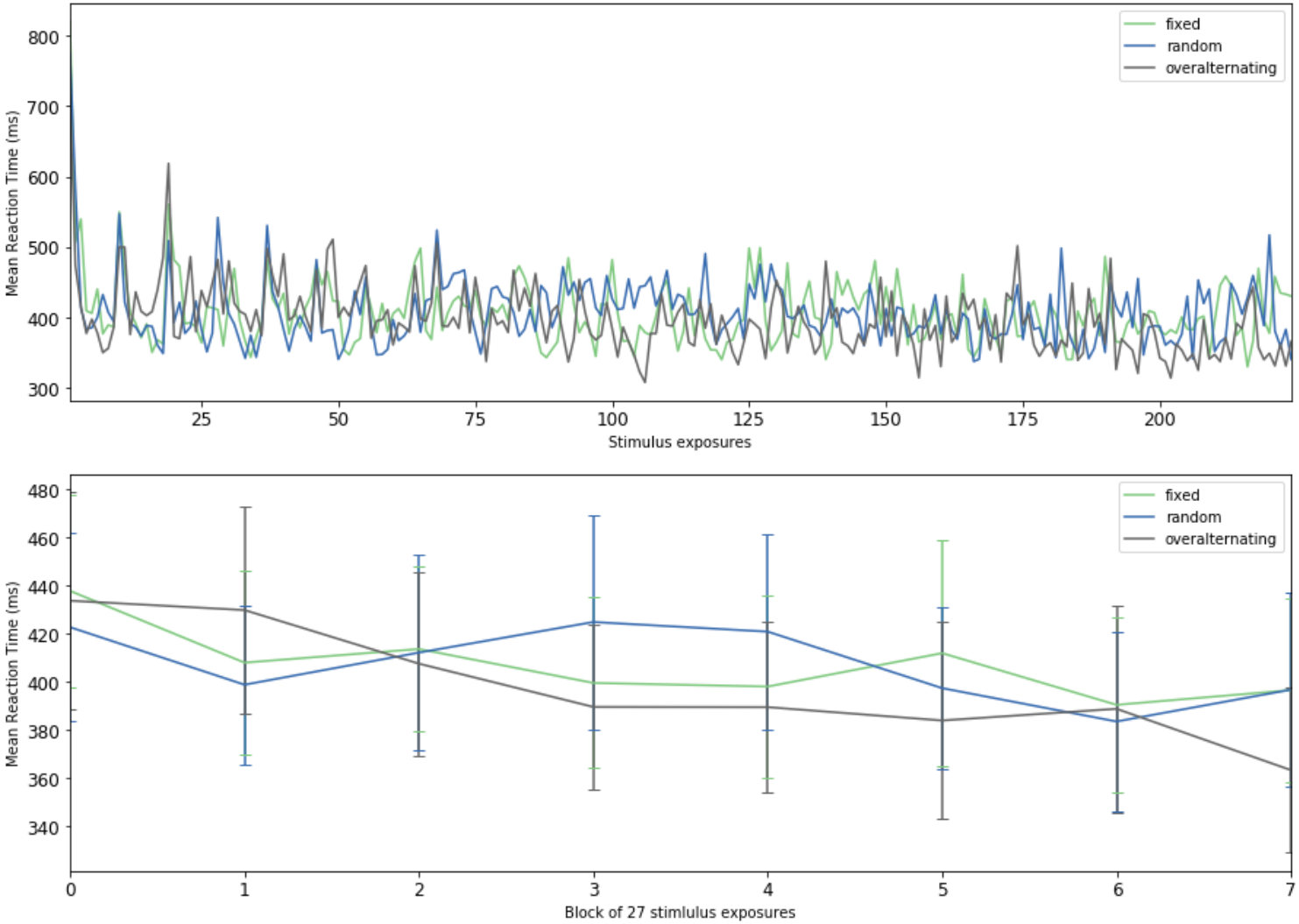


Figure 8. Mean RTs aggregated across human participants. The lower plot aggregates sequence exposures into blocks of 27 and includes Standard Error bars. Note that due to participant agency over how many sequence exposures they received, not all participants are included in each aggregation.

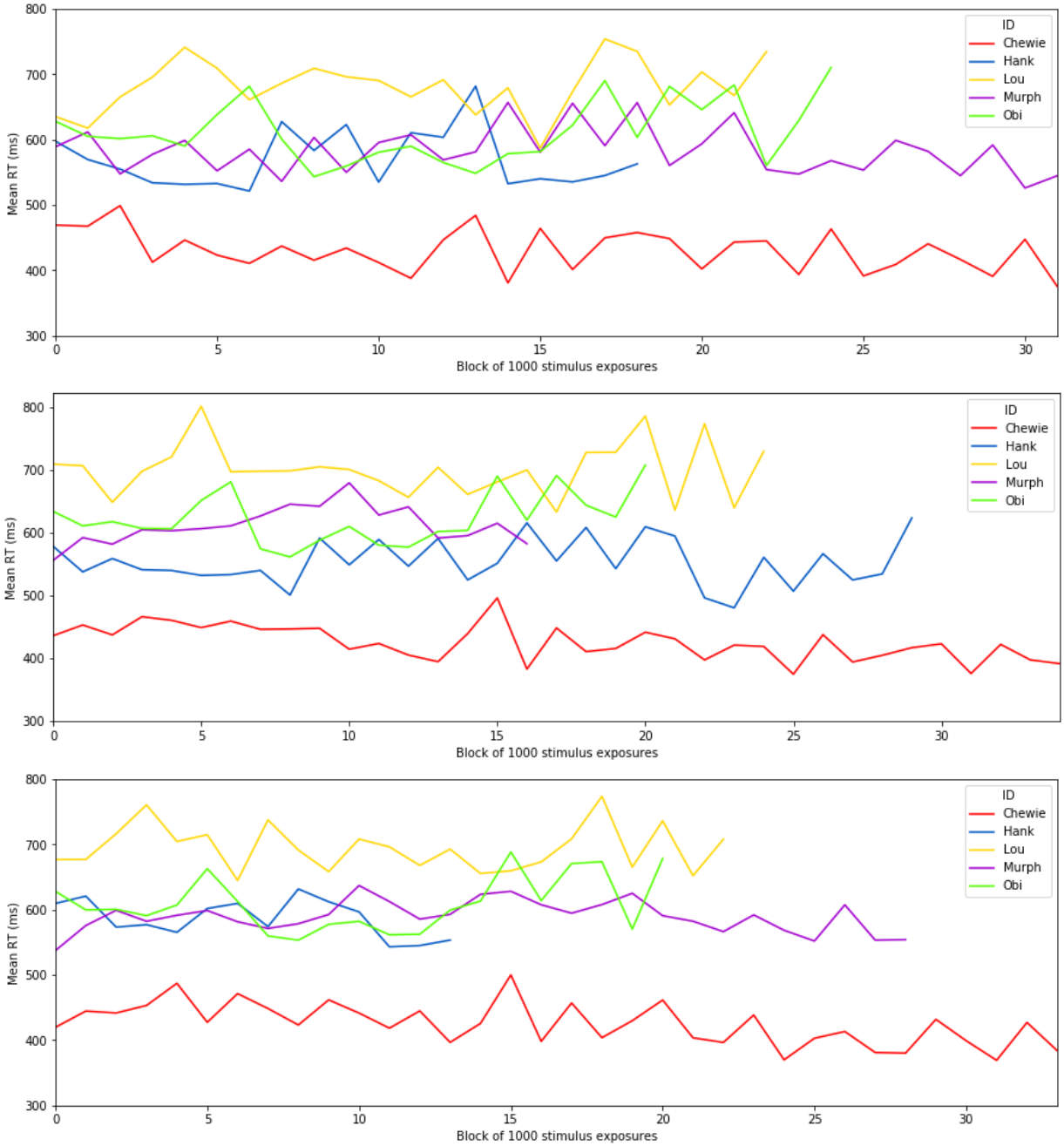


Figure 9. Mean RTs to stimuli from fixed (top), random (middle), and overalternating (bottom) sequence types for each monkey. The lower plot aggregates sequence exposures into blocks of 1000 stimulus exposures.

monkeys' mean RTs aggregated across 1000-trial exposures to stimuli from the fixed sequence (top plot), random sequence (middle plot), and overalternating sequence (bottom plot). From one block of 1000 stimulus exposures to another, monkeys' mean RTs were observed to change by as much as 100 ms. Surprisingly, monkeys' RTs were also much slower than those of human participants. With the exception of the monkey Chewie, the monkeys' RTs to the stimuli were nearly 200 ms slower than those of human participants, on average. Like those of human participants, the monkeys' RTs were relatively unaffected by the embedded structures of the fixed and overalternating sequences (Figure 10). Instead, RTs for the stimuli of the different sequences can largely be understood as deviations around the mean RT of an individual animal (for example, the relatively stable RT of approximately 450 ms of Chewie in the uppermost plot).

Figures 11 through 16 capture the choice behaviors of the human participants and individual monkeys in the SELECT task of Experiment 2. As in Experiment 1, stable sequence preferences were not observed in aggregated human data. Three out of five monkeys also did not exhibit stable changes in choice behavior. Monkey Hank (Figure 12) reliably chose the random sequence. Monkey Murph (Figure 13) chose the fixed sequence over the random sequence, the fixed sequence over the overalternating sequence, and the overalternating sequence over the random sequence.

3.2.1 SRT modeled.

The overall analysis plan remained the same as in Experiment 1. Two additional potential predictors were included in the linear mixed-effects model for SRT data. One was species, included as a fixed effect in the model. The other was a measure of overall experience with the task, as the monkeys' high trial numbers made such a predictor more informative. The results

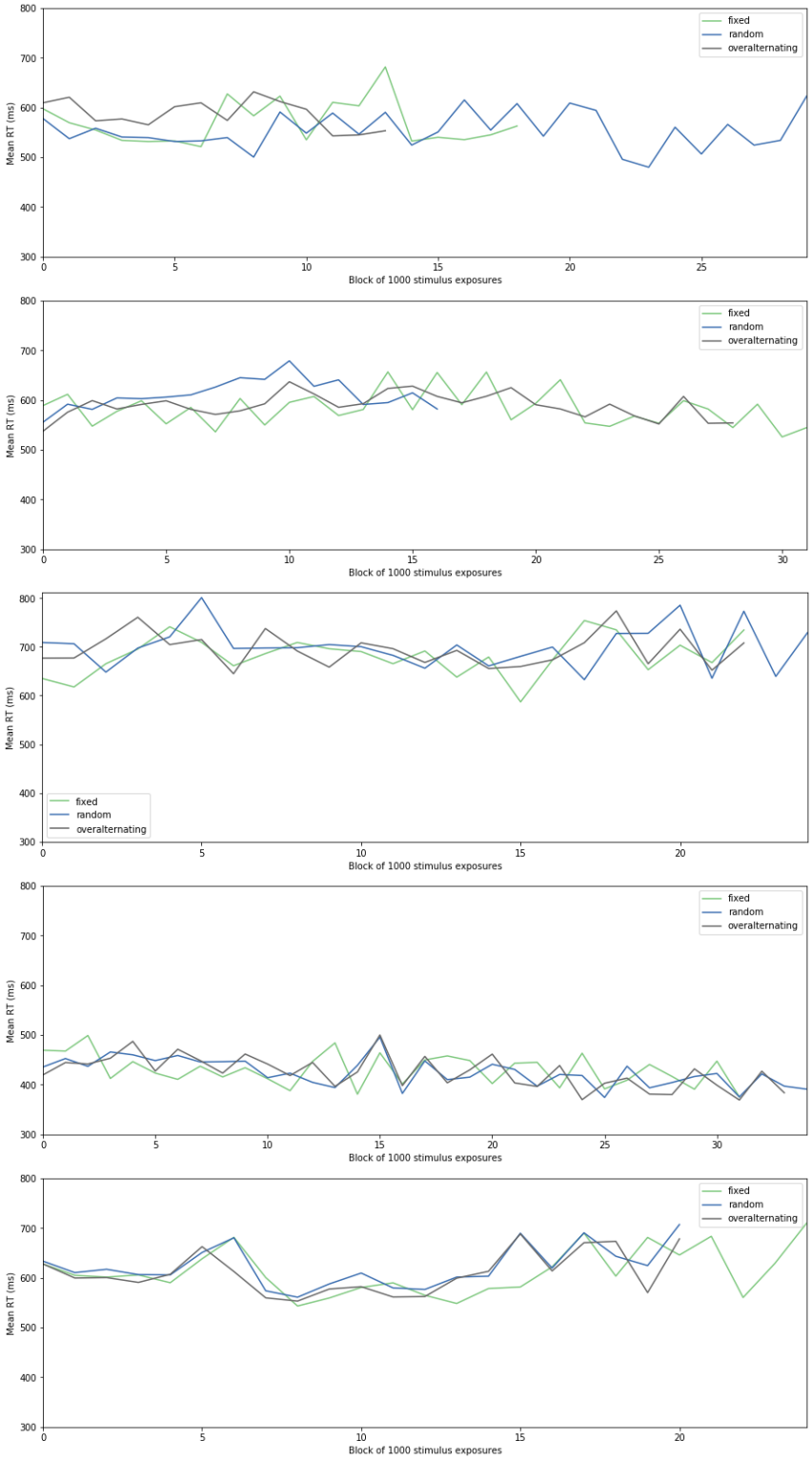


Figure 10. The mean RT plots from Figure 9, rearranged to capture effect of sequence type on RT for each monkey. From top: Hank, Murph, Lou, Chewie, Obi

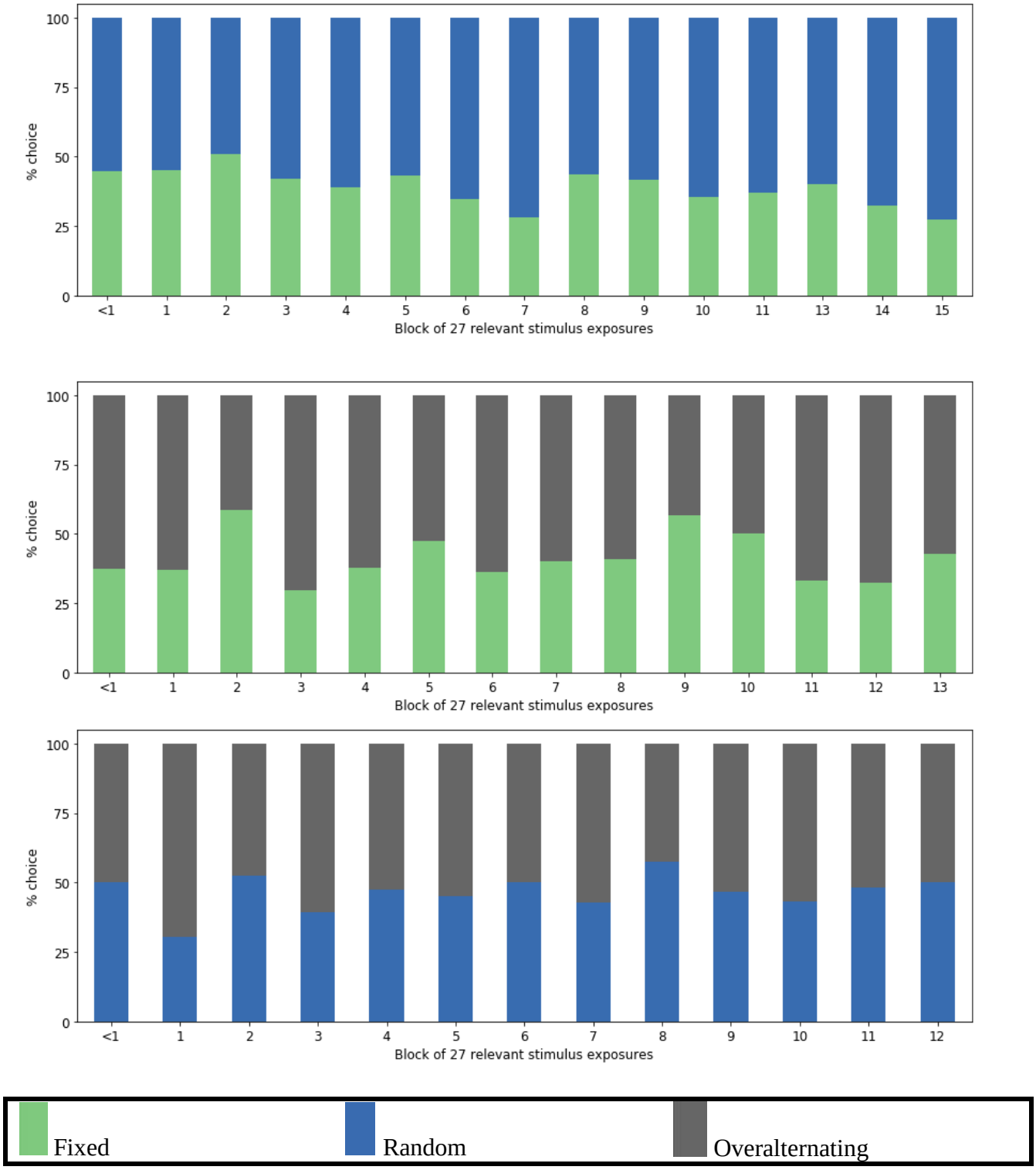


Figure 11. Aggregated human participant choices in different SELECT menus across different levels of relevant experience with the task. Blocks are of 27 stimulus exposures.

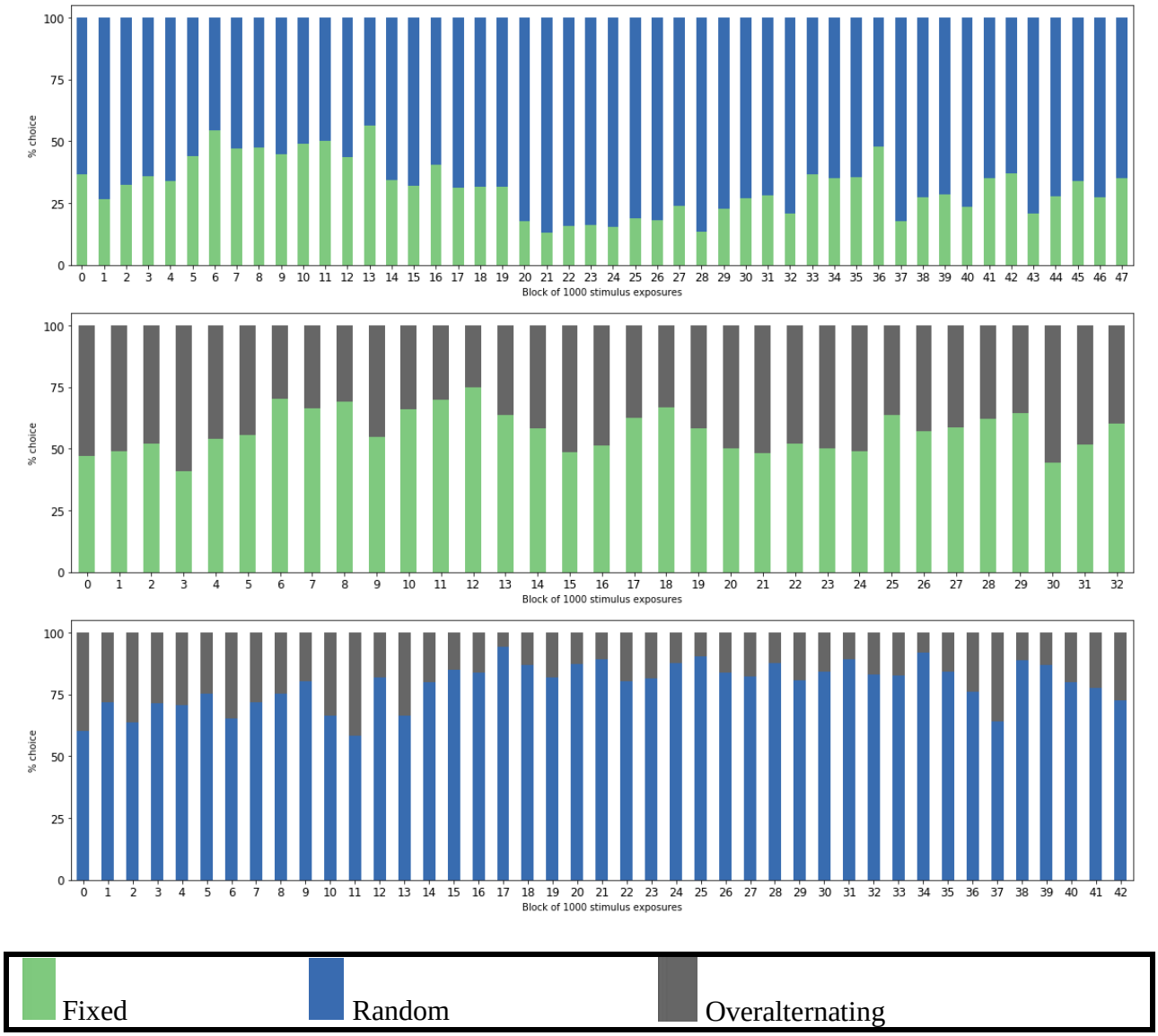


Figure 12. Aggregated choices in different SELECT menus across different levels of relevant experience with the task for monkey Hank. Blocks are of 1000 stimulus exposures.

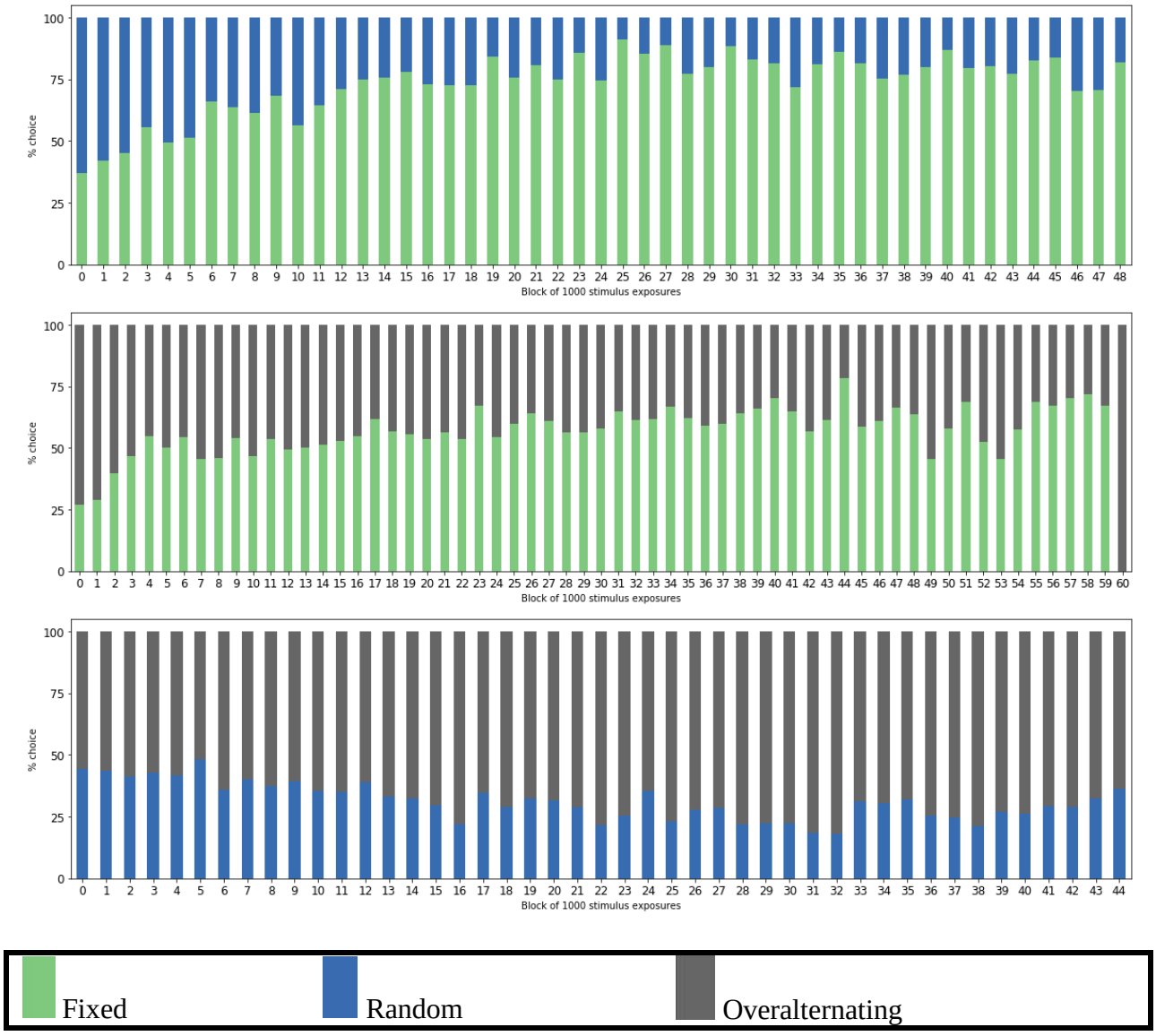


Figure 13. Aggregated choices in different SELECT menus across different levels of relevant experience with the task for monkey Murph. Blocks are of 1000 stimulus exposures.

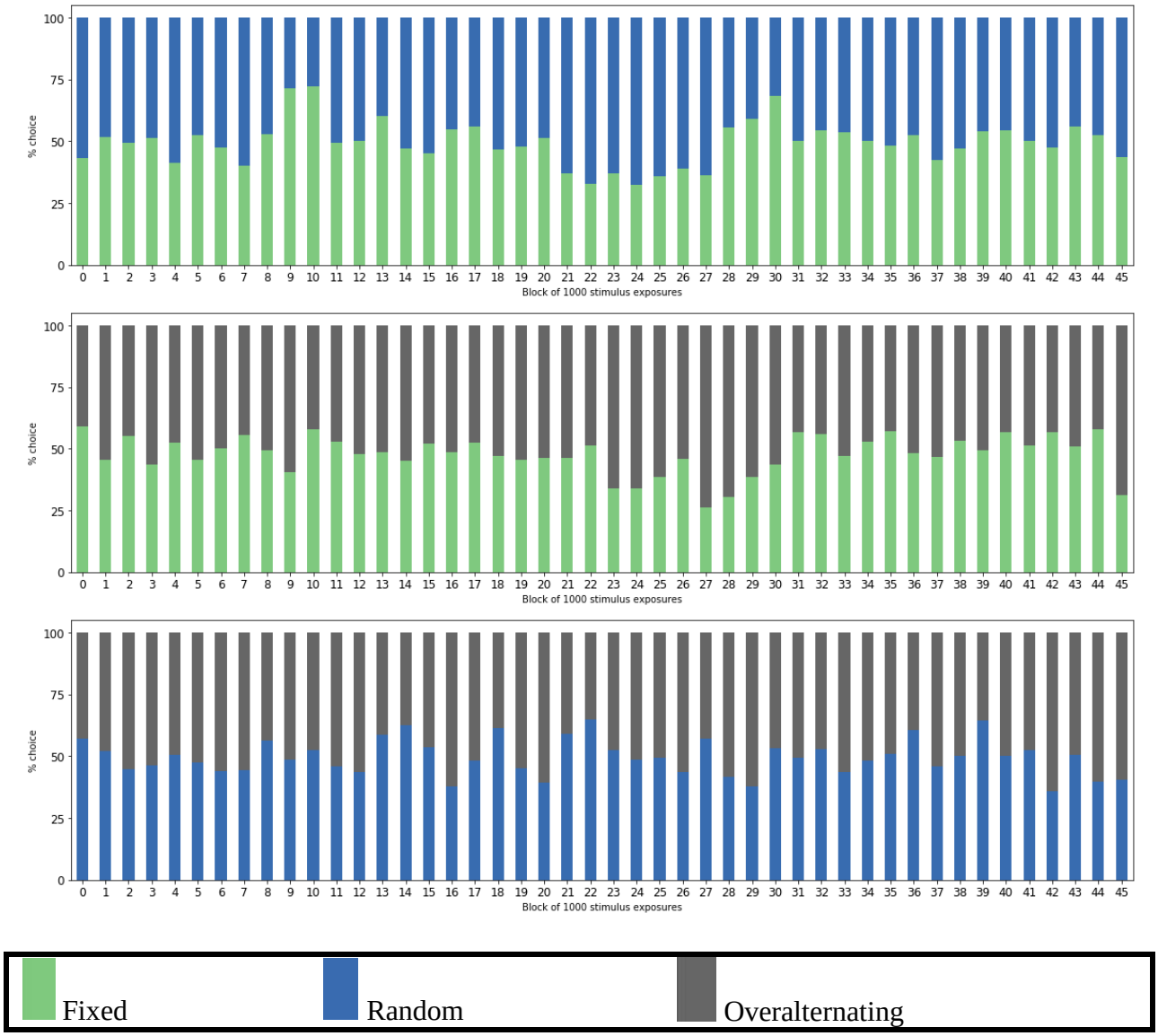


Figure 14. Aggregated choices in different SELECT menus across different levels of relevant experience with the task for monkey Lou. Blocks are of 1000 stimulus exposures.

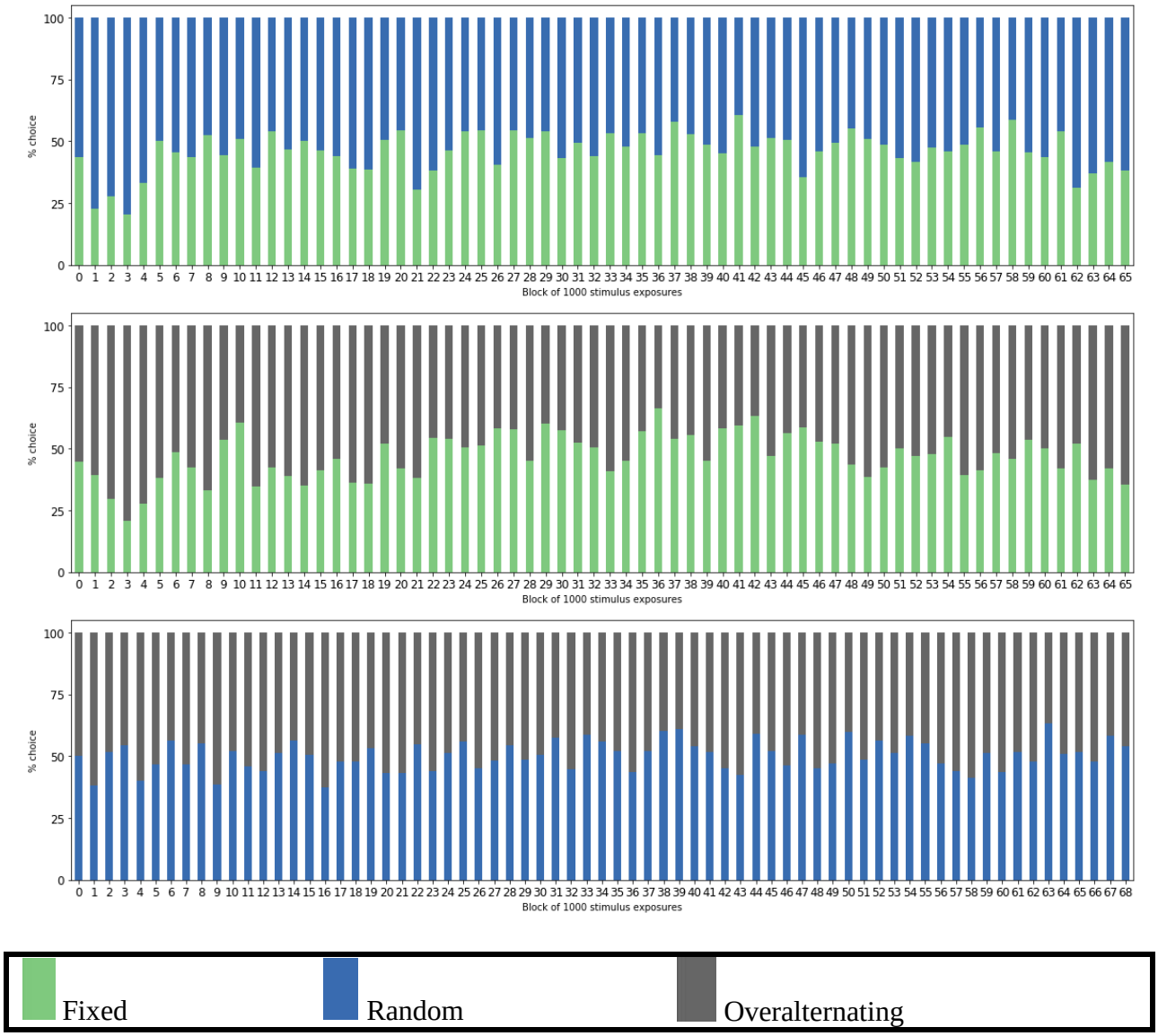


Figure 15. Aggregated choices in different SELECT menus across different levels of relevant experience with the task for monkey Chewie. Blocks are of 1000 stimulus exposures.

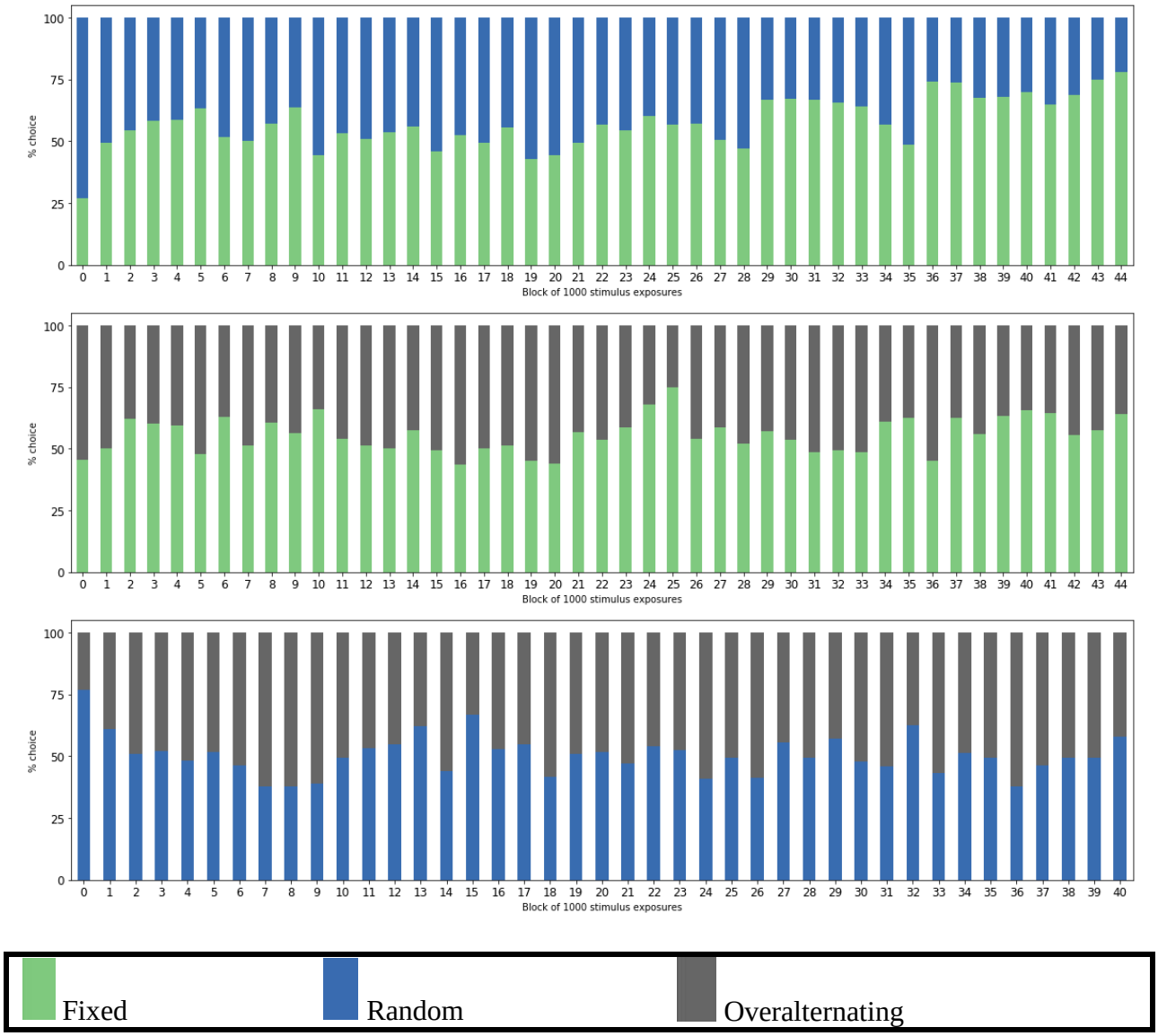


Figure 16. Aggregated choices in different SELECT menus across different levels of relevant experience with the task for monkey Obi. Blocks are of 1000 stimulus exposures.

of the linear mixed-effects model (*AIC* = 81790.641, *df* = 29) fit for the SRT data are presented in Table 4. The fixed effect of species was significant (β = -.18, *SE* = .030, $p < .001$), such that humans' log(RT/ms) were approximately 18% faster than monkeys (a difference of approximately 102 ms). The three-way interaction between species, the fixed trial type, and experience with the fixed trial type was significant (β = -.00009, *SE* = .00002, $p = .005$) such that human reaction times decreased in the fixed sequence with experience in the task whereas monkeys' did not (see Figures 17 and 18). In general, the model underestimated human RTs and overestimated monkey RTs. This is likely due to both the large imbalance in the number of trials monkeys completed relative to human participants and the difference in mean RTs for humans and monkeys. To better capture the joint effects of species and individual differences, aggregations of the estimated random effects parameters are presented in Figures 19 and 20. Certain parameters had a disproportionate effect on the RT of certain individuals (for example, the RTs of monkey Lou on the bottom plot of Figure 20 were best predicted by lower estimates for the coefficient of the effect of experience in the task). But in both species, individual parameter estimates were distributed around zero such that the coefficient of a predictor, for example, experience in the task, might be positive for one participant and negative for another. As in Experiment 1, genuine experimental effects may be obscured by conflict between learning and deleterious experience effects.

3.2.2 SELECT modeled.

The human and monkey data were modeled separately due to constraints associated with the great difference in experience with the menus for the monkeys and humans. Human choice behavior did not significantly change in any menu in interaction with increased experience relevant to that menu (fixed v random menu x experience: B = .0007, SE = .0005, $p = .176$; fixed

Table 4. Linear Mixed-Effects Model Fit

Parameter	β	*SE*	*p*
Intercept	6.138	.041	>.001*
Sequence Experience	-.001	.027	.970
Overall Experience	-.0002	.025	.992
Fixed	.013	.032	.693
Overalternating	-.001	.033	.965
Human	-.180	.041	<.001*
Fixed x Sequence Experience	-.0001	.00003	<.001*
Overalternating x Sequence Experience	-.00003	.00003	.210
Human x Sequence Experience	-.001	.027	.970
Fixed x Sequence Experience for Humans	-.0001	.00003	<.001*
Overalternating x Sequence Experience for Humans	-.00003	.00003	.213

*Notes.***p* < .05. The intercept is the grand mean for log(RT/ms) across sequence types. The *p* values are computed from Wald z-tests testing whether the model coefficient is significantly different from zero.

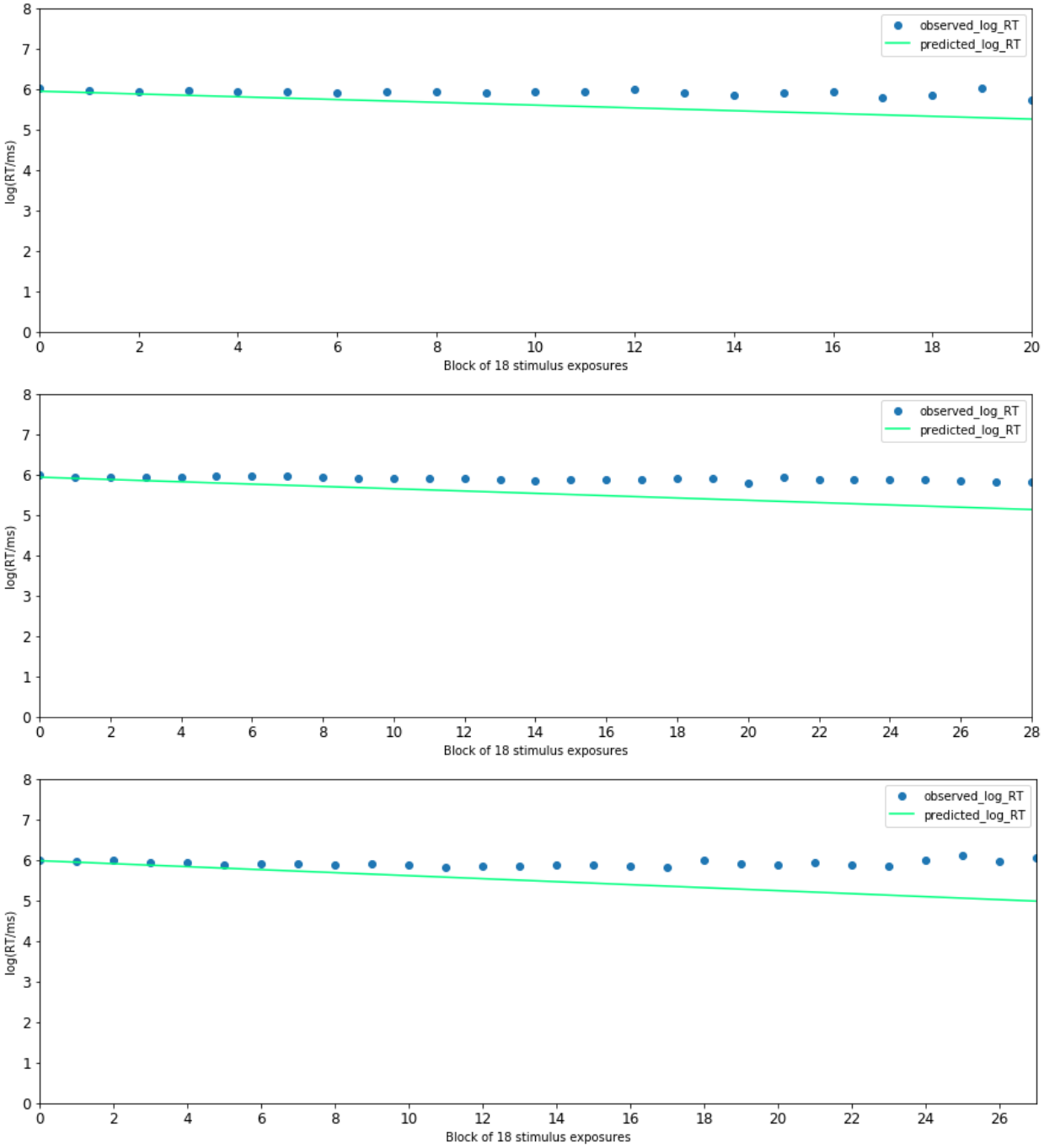


Figure 17. Averaged observed log(RT/ms) for humans across increasing blocks of experience with the fixed sequence (top), random sequence (middle), and overalternating sequence (bottom). Each block was equal to two full sequences, 18 stimulus exposures.

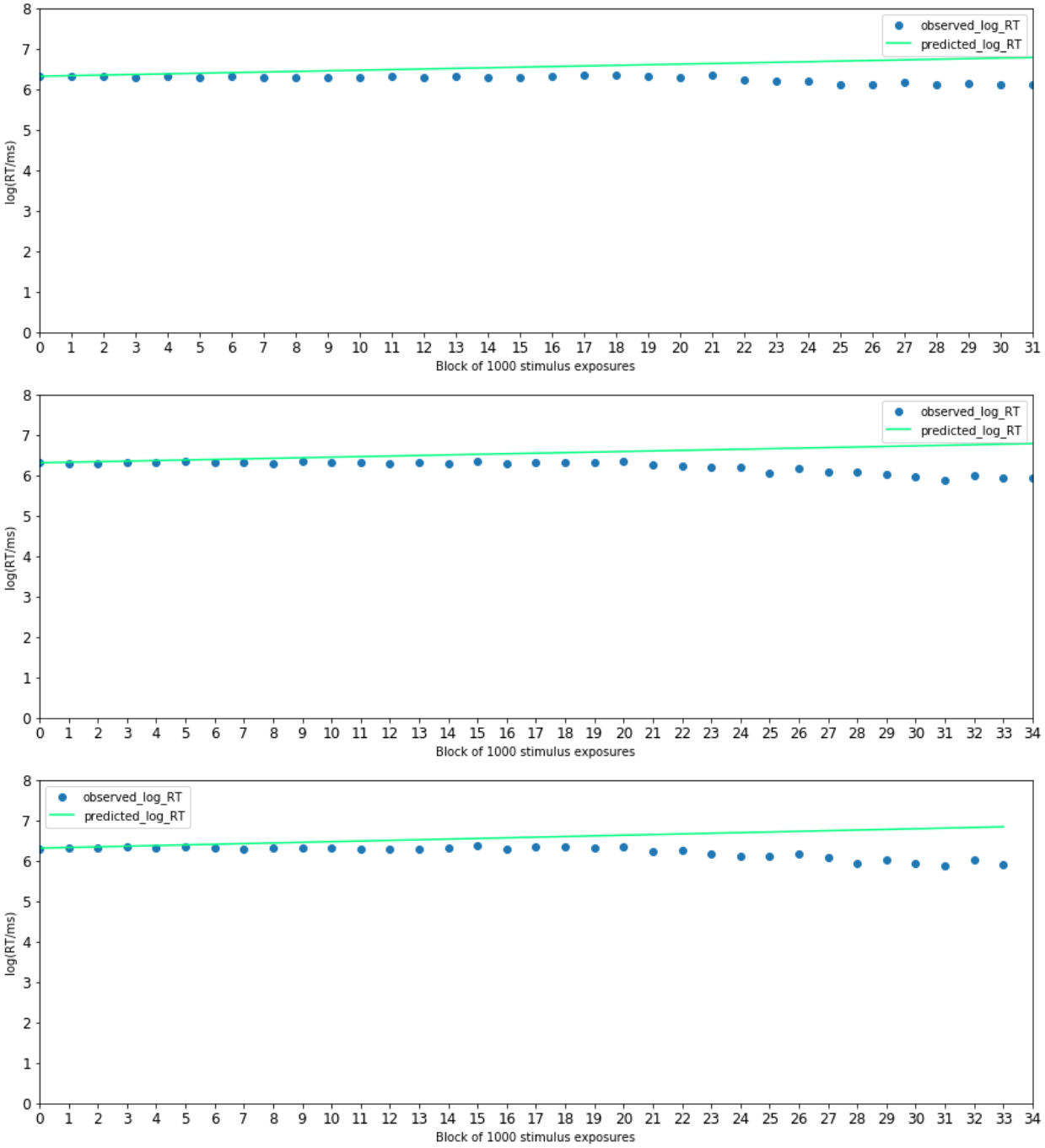


Figure 18. Averaged observed log(RT/ms) for monkeys across increasing blocks of experience with the fixed sequence (top), random sequence (middle), and overalternating sequence (bottom). Each block was equal to 50 full sequences, 300 stimulus exposures.

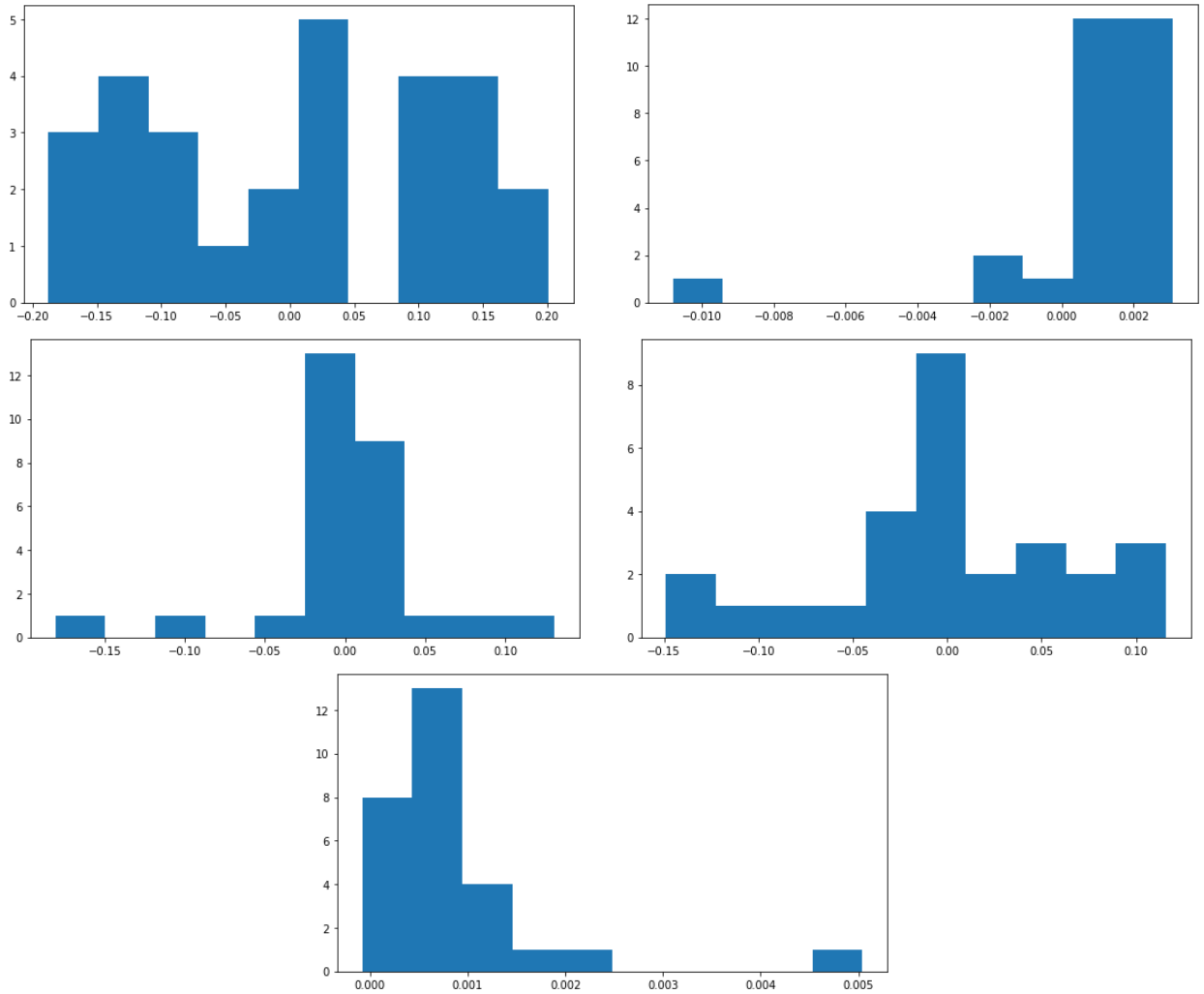


Figure 19. Histograms of frequency of deviations of individual human participants' parameter estimations from overall fixed effect parameter estimation for the grand mean log(RT/ms) (top-left), the effect of increasing experience with the different conditions (top-right), the effect of the sequence being fixed (middle-left), the effect of the sequence overalternating (middle-right), and the effect of increasing experience in the task across conditions (bottom).

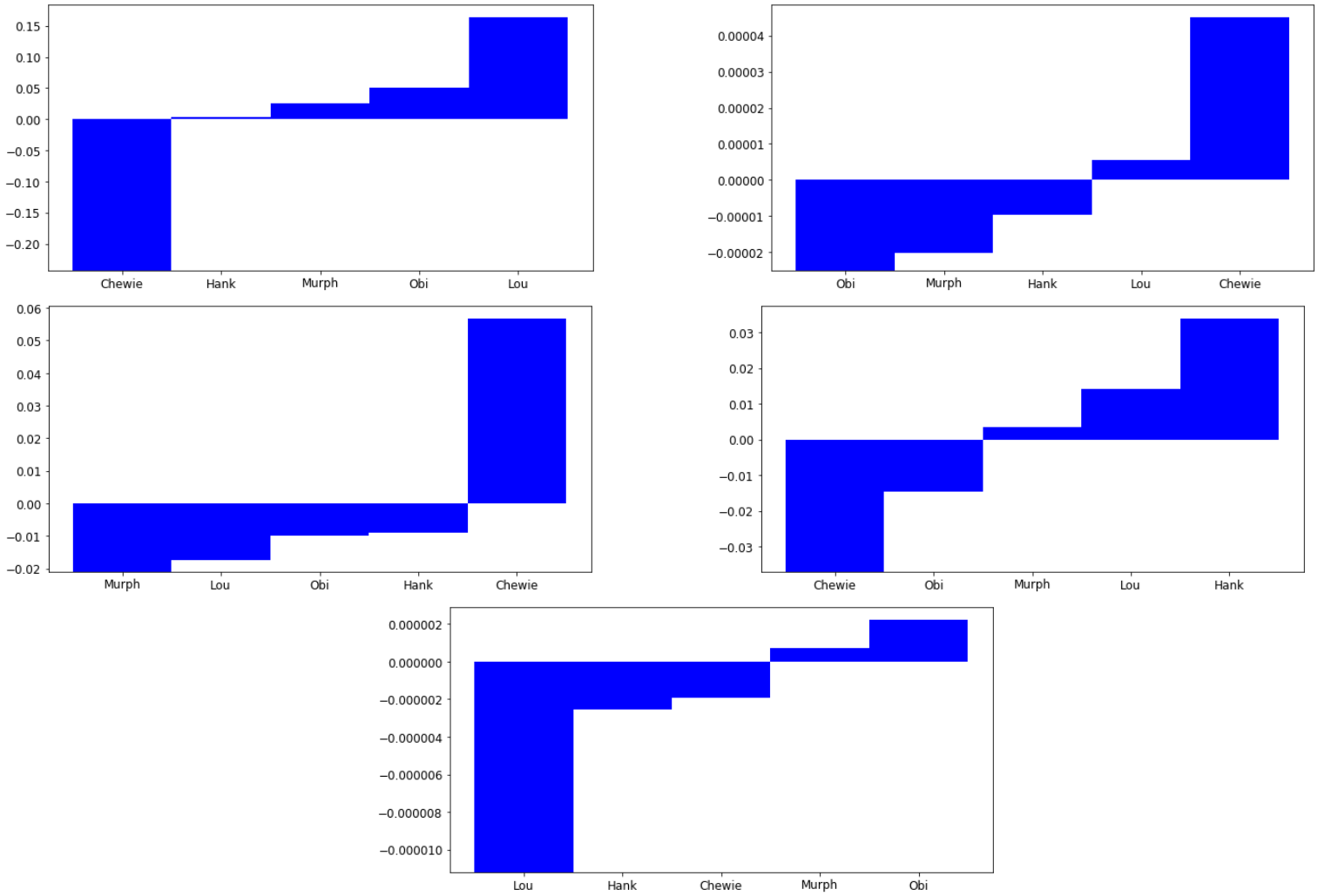


Figure 20. Individual monkeys' parameter estimations for the grand mean log(RT/ms) (top-left), the effect of increasing experience with the different conditions (top-right), the effect of the sequence being fixed (middle-left), the effect of the sequence overalternating (middle-right), and the effect of increasing experience in the task across conditions (bottom).

v overalternating menu x experience: B = -.0008, SE = .0005, p = .092). Monkey choice behavior significantly changed in the fixed v random menu (B = -.00004, SE = .000008, p < .001) but not in the fixed v overalternating menu (B = -.000003, SE = .000008, p = .668).

3.3 Discussion

As in Experiment 1, the data revealed few meaningful effects of learning in the SRT task or choice behavior in the SELECT task with increasing task experience. The monkeys, with hundreds of times the task experience of the human participants, failed to demonstrate either a meaningful shift in choice behaviors or mastery of the structured sequences of the SRT task. The methodological changes from Experiment 1 to Experiment 2 were effective in concentrating human participant's learning of the fixed sequence. Human participants RTs to fixed stimuli decreased across exposures to these stimuli. Yet choice behaviors were not observed to change in concert with decreasing RTs, and Experiment 2 failed to replicate the decrease in RTs for overalternating stimuli that was observed in Experiment 1.

4 GENERAL DISCUSSION

Across two experiments and two species, the changes in RTs to structured stimuli that were predicted in Hypothesis 1 were observed only inconsistently. Experiment 1 demonstrated a change in RT to overalternating stimuli only, whereas Experiment 2 demonstrated a change in RT to fixed stimuli only, and only in humans. This null result for Hypothesis 1 thwarts interpretations of the choice behaviors of the humans and monkeys that are relevant to Hypotheses 2, 3, and 4.

The inconsistency of significant change in RT to fixed or overalternating stimuli can be understood as a consequence of aspects of the experimental design. In order to maximize ability to interpret a change in choice behavior as a specific response to sequence properties, trial

durations were constant within sessions. This means that for human participants, task performance, whether in terms of time to complete a trial or in trial success or failure, was not incentivized. The fairly monotonous rhythm of the SRT task perhaps compounded motivational issues by making participants more keenly aware of the lack of incentives for participating actively. In order to preserve comparisons with a literature on randomness perception that has long focused on binary outcomes, a binary version of SRT task was used. But this use of only two responses might have made learning fixed and overalternating sequences more difficult for humans by burdening short-term and working memory systems with many associations linked to only two response items (i.e., a network of associative interference from the constant acquisition of more transitions associated with, say, the left response).

Observable effects of these variations to task parameters and incentive structure are perhaps evidenced in comparisons between the different experiments and species of this manuscript, and in comparisons with previous research. In Experiment 1, participants completed 21 item sequences approximately 62% of the time. In Experiment 2, participants completed 9 item sequences approximately 72% of the time. Non-completion of sequences should not have made a meaningful difference to data analyses, but may have contributed to participant's lack of learning of the transitional probabilities embedded in the sequences. For example, a participant who saw an average of three items in each sequence responded to a very different corpus of stimulus regularities than did a participant who responded to the full breadth of each sequence. Moreover, a lack of motivation to complete the task as designed may have incentivized exploring the task parameters in unknown ways. One observed example of this might be the unusual propensity of participants to "select" the blank space of the SELECT screen on forced trials.

Monkeys, however, received pellet rewards for successful task completion and were thus suitably incentivized on at least this dimension of task performance. With this incentive in place, the animals completed 88% of their 6-item sequences overall. However, the monkeys' mean RT of 572 ms was substantially slower than that of humans in both Experiment 1 (mean RT: 387 ms) and Experiment 2 (mean RT: 404 ms). Were the trial durations to be unlocked so that faster sequence completions increased the overall rate of reward, the animals might be suitably incentivized to demonstrate the effect of learning that is typically observed in SRT and statistical learning designs (e.g., Heimbauer et al., 2012a). Unlocking trial durations would at least allow for interpretation of Hypothesis 1, even if trends in the SELECT data would be much less valuable because they may be the result of task parameters other than the properties of the sequence (overall reward rate, latency between choice behavior and reward, etc.).

In the previous work of Heimbauer and colleagues (2012a), two of the monkeys from Experiment 2 (Obi and Lou) and one additional animal (Luke) demonstrated faster responding for sequences that were the same each time (as in the fixed sequences of this experiment) than for random sequences. The difference in results of those experiments and Experiment 2 of this manuscript is notable for a few reasons (see Figures 16, 17, and 18). In the Heimbauer et al. (2012a) experiments, ITIs were constant and did not titrate up and down in order to keep trial duration constant. This methodological difference, embedded in a largely equivalent SRT procedure with the same population of monkeys, acts as experimental evidence that the locked trial duration of the experiments presented here negated some part of the predicted effect of sequence learning. The mean RTs of the individual animals are also noteworthy. Despite these animals receiving fewer pellets for sequence completion and responding to stimuli in four directions (up, down, left, right) rather than the two directions (left, right) of this manuscript,

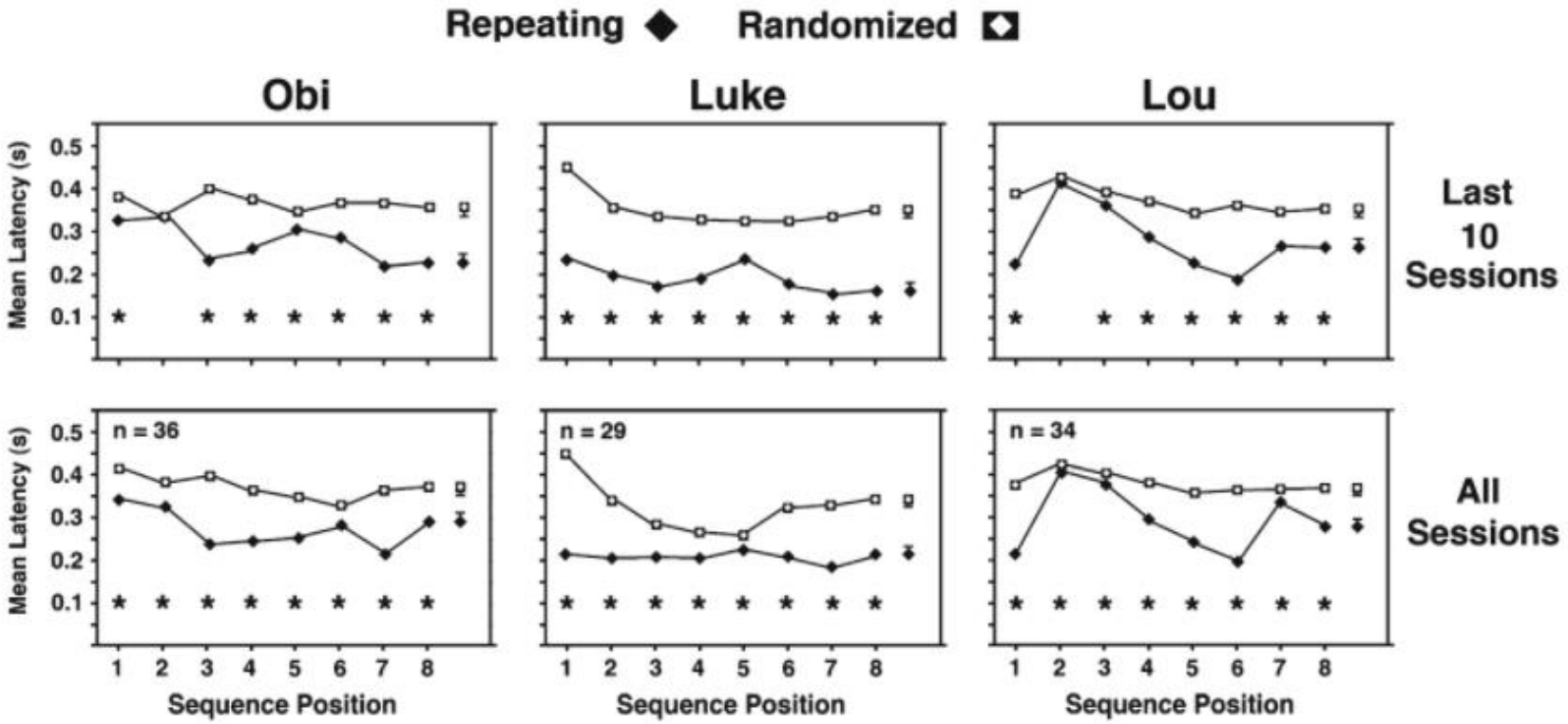


Figure 21. Mean RTs for two monkeys from this manuscript in a previous SRT task by Heimbauer et al. (2012a). Animals completed 8 item sequences of fixed (repeating) significantly more quickly than random (randomized) stimuli at each of the 8 stimulus positions. This effect was present in both the animals' final 10 sessions and in the experiment overall

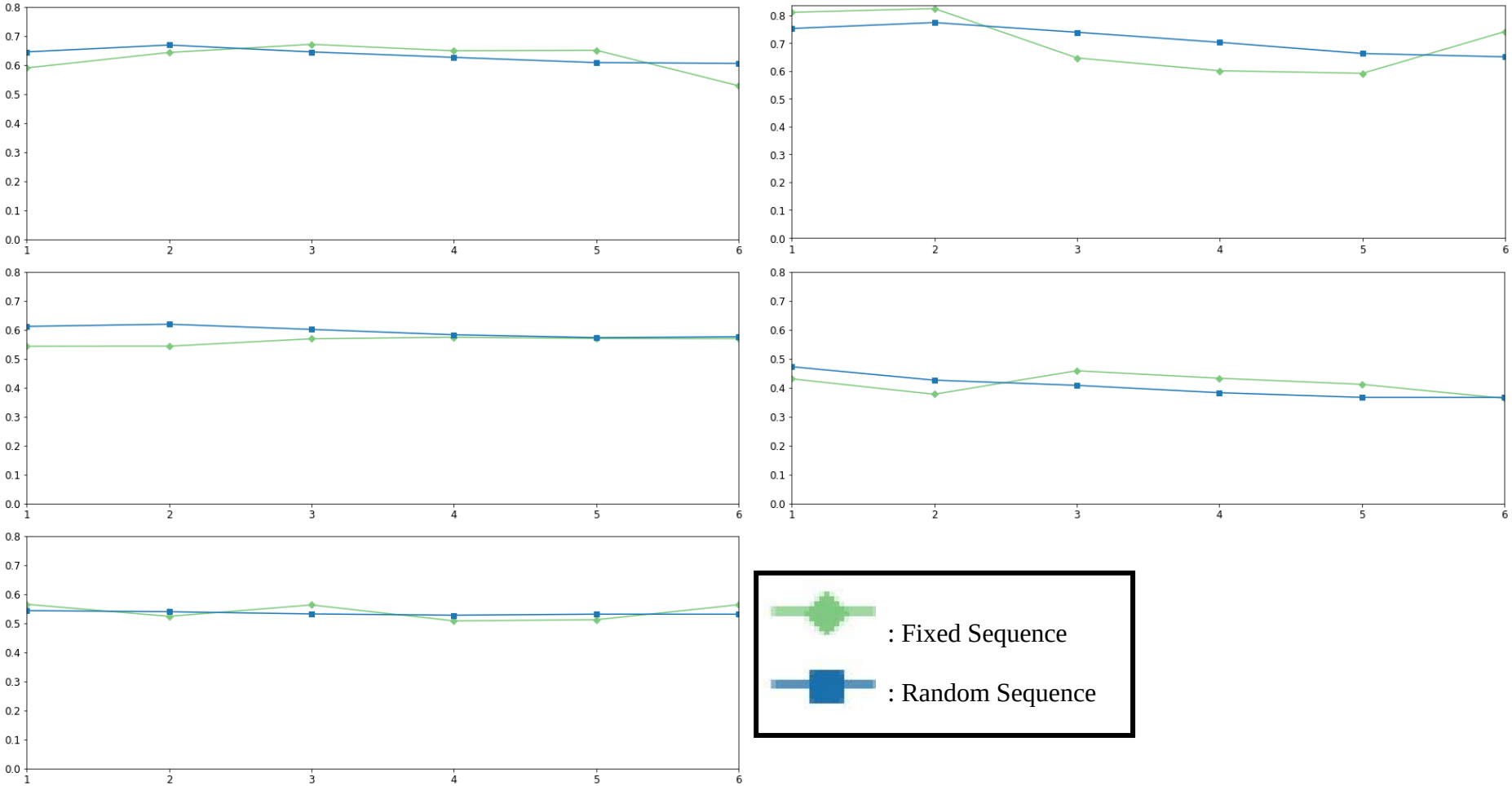


Figure 22. Mean RTs to different sequence positions for the monkeys from Experiment 2 in their final 5 sessions. Clockwise from top-left: Obi, Lou, Chewie, Hank, Murph

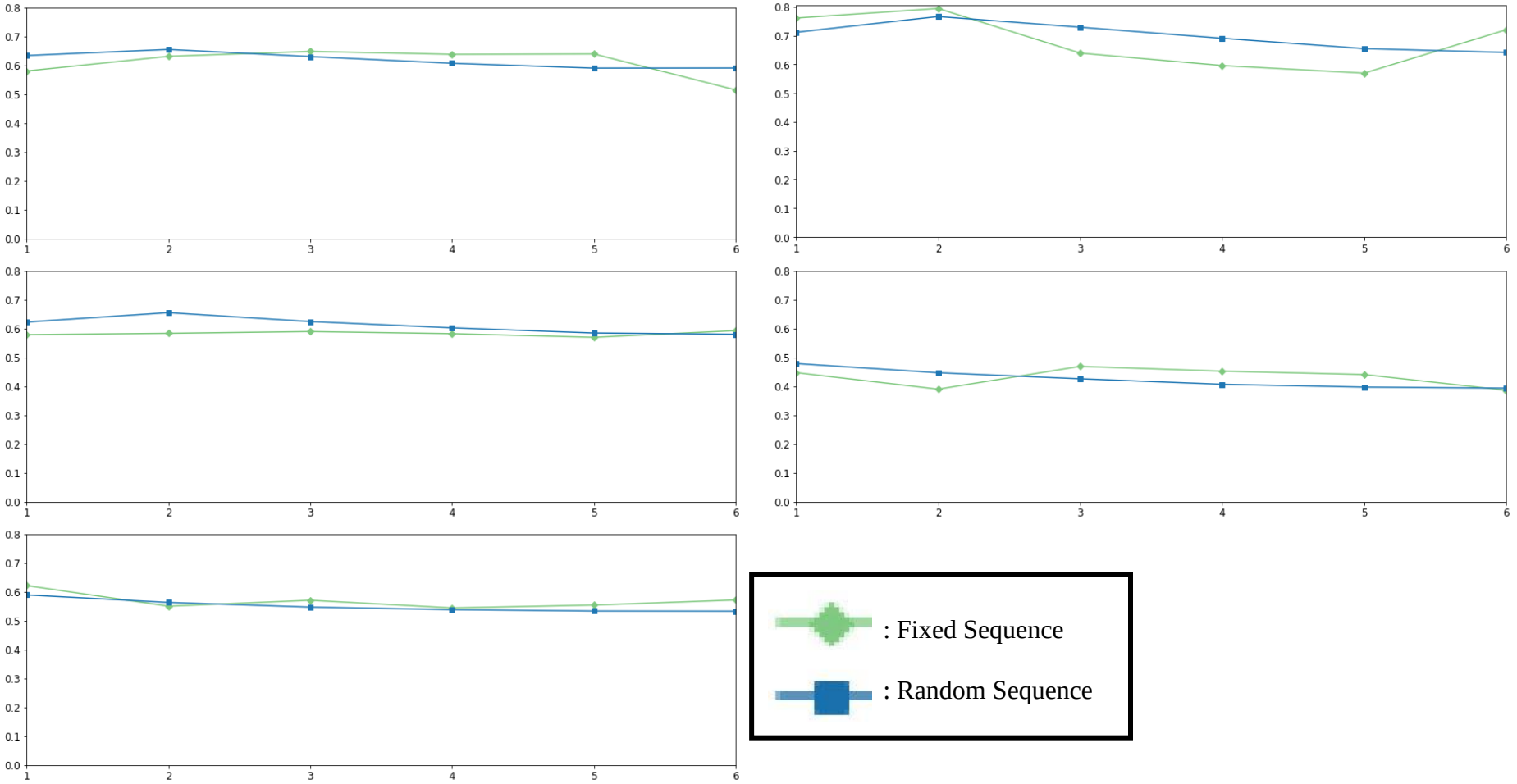


Figure 23. Mean RTs to different sequence positions for the monkeys from Experiment 2 across all sessions. Clockwise from top-left: Obi, Lou, Chewie, Hank, Murph

their mean RTs within each session were, without exception, below the mean RT for monkeys in Experiment 2.

In the works of Wilke and colleagues (2014) reviewed previously, human participants' and rhesus macaques' choice behaviors changed in response to experience with sequences. Specifically, participants preferentially chose sequences that alternated least and streaked most. The absence of this effect in both Experiments 1 and 2 is unexpected, and perhaps another byproduct from these experiments' constant trial durations. Without set trial durations participants could seek to maximize rewards through strategic decision-making on the SELECT task, and the choice behaviors observed by Wilke and colleagues would likely re-emerge. Nevertheless, a unique contribution of Experiments 1 and 2 is to nuance one potential conclusion of the work of Wilke and colleagues. One way of understanding the findings of those researchers is to conclude that specific sequence properties elicit specific decision behaviors. For example, mere observation of overalternating (clumpy) probabilistic environments elicit behaviors that seek more of such environments. But observation alone, without differential feedback or reward structures, was not sufficient to change choice behaviors of the participants of Experiments 1 and 2. Other associative cues and other learning is probably required for the previously reported shifts in choice behaviors.

Previous research has done much to explore the ways in which humans and animals interact with probabilistic information, make choices in probabilistic environments, and make explicit and implicit judgments about the probabilistic structures with which their environments are laden. The present pair of experiments was designed to try to unify two separate bodies of this research: that of how humans come to make judgments about the relative randomness and nonrandomness of sequences and that of the powerful statistical learning mechanism through

which sequence properties are implicitly learned. To make that unification, however, design elements were required that cannibalized the well-established effects that were to be paired. Most notably, the decision to make trial duration constant probably did much to limit any demonstration of learning or development of change in choice behaviors. Without constant trial durations, the stated interpretations of Hypotheses 2, 3, and 4 become fraught with alternative explanations. However, with constant trial durations, Hypotheses 2, 3, and 4 are evidently uninterpretable from the outset.

A parallel design could unlock trial durations, and reward humans for sequence completions with small monetary compensation. This could provide evidence for sequence learning in sequences of binary stimuli, a goal of Hypothesis 1. This design would also be akin to a replication and extension of the work of Wilke and colleagues, in that a correlation between sequence properties and choice behavior could be established (with the added component of having a measure of learning of sequence properties to accompany the correlation) (Blanchard, Wilke, & Hayden, 2014; Scheibehenne et al., 2011; Wilke et al., 2014). This design, perhaps with a sequence entropy measure to synthesize the information content a participant received more minutely, might offer a way forward for answering this theoretically interesting question of human and animal cognition.

REFERENCES

Azouvi, P., Jokic, C., Van Der Linden, M., Marlier, N., & Bussel, B. (1996). Working memory and supervisory control after severe closed-head injury: A study of dual task performance and random generation. *Journal of Clinical and Experimental Neuropsychology, 18*(3), 317–337.

Baddeley, A. D. (1966). The capacity for generating information by randomization. *Quarterly Journal of Experimental Psychology, 18*(2), 119–129.

Baddeley, A., Emslie, H., Kolodny, J., & Duncan, J. (1998). Random generation and the executive control of working memory. *The Quarterly Journal of Experimental Psychology, 51*(4), 819–852.

Bar-Hillel, M., & Wagenaar, W. A. (1991). The perception of randomness. *Advances in Applied Mathematics, 12*(4), 428–454.

Barraclough, D. J., Conroy, M. L., & Lee, D. (2004). Prefrontal cortex and decision making in a mixed-strategy game. *Nature Neuroscience, 7*(4), 404–410.

Blanchard, T. C., Wilke, A., & Hayden, B. Y. (2014). Hot-hand bias in rhesus monkeys. *Journal of Experimental Psychology: Animal Learning and Cognition, 40*(3), 280–286.

Brown, R. G., Soliveri, P., & Jahanshahi, M. (1998). Executive processes in Parkinson's - Disease random number generation and response suppression. *Neuropsychologia, 36*(12), 1355–1362.

Brugger, P. (1997). Variables that influence the generation of random sequences: an update. *Perceptual and Motor Skills, 84*(2), 627–661.

Brugger, P., Monsch, A., Salmon, D. , & Butters, N. (1996). Random number generation in dementia of the Alzheimer type: A test of frontal executive functions. *Neuropsychologia, 34*(2), 97–103.

Chapanis, A. (1953). Random-number guessing behavior. *American Psychologist, 8*, 332.

Chapanis, A. (1995). Human production of "random" numbers. *Perceptual and Motor Skills, 81*, 1347–1363.

Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences, 5*(12), 539–546.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 24–39.

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: pitting abstract against stimulus-specific representations. *Psychological Science, 17*(10), 905–12.

Diener, D. (1985). Recognizing randomness. *American Journal of Psychology, 98*(3), 433–447.

Falk, R. (1975). *Perception of randomness*. Hebrew University.

Falk, R. (2014). Monty's dilemma with no formulas. *Teaching Statistics, 36*(2), 58–61.

Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review, 104*(2), 301–318.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 458–467.

Goldman, A. (1986). *Epistemology and Cognition*. Cambridge, MA.: Harvard University Press.

Hale, D. J. (1969). Repetition and probability effects in a serial choice reaction task. *Acta Psychologica*, 29, 163–171.

Harris, E. H., Gulledge, J. P., Beran, M. J., & Washburn, D. A. (2010). What do Arabic numerals mean to macaques (Macaca mulatta)?. *Journal of Experimental Psychology: Animal Behavior Processes*, *36*(1), 66.

Heimbauer, L. A., Conway, C. M., Christiansen, M. H., Beran, M. J., & Owren, M. J. (2012a). A Serial Reaction Time (SRT) task with symmetrical joystick responding for nonhuman primates. *Behavioral Research*.

Heimbauer, L. A., Conway, C. M., Christiansen, M. H., Beran, M. J., & Owren, M. J. (2012b). Artificial Grammar Learning and Generalization by Rhesus Macaques (Macaca mulatta). *Unpublished Data*.

Kac, M. (1983). Marginalia: What is random? *American Scientist, 71*(4), 405–406.

Kareev, Y. (1995a). Positive bias in the perception of covariation. *Psychological Review, 102*(3), 490–502.

Kareev, Y. (1995b). Through a narrow window: working memory capacity and the detection of covariation. *Cognition, 56*(3), 263–269.

Kareev, Y. (2000). Seven (indeed, plus or minus two) and the detection of correlations. *Psychological Review, 107*(2), 397–402.

Kareev, Y., Lieberman, I., & Lev, M. (1997). Through a narrow window: Sample size and the perception of correlation. *Journal of Experimental Psychology: General, 126*(3), 278–287.

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology, 48*(1), 171–184.

Klein, E. D., Evans, T. A., Schultz, N. B., & Beran, M. J. (2013). Learning how to "make a deal": Human (*Homo sapiens*) and monkey (*Macaca mulatta*) performance when repeatedly faced with the Monty Hall Dilemma. *Journal of Comparative Psychology*, *127*(1), 103

Locurto, C., Dillon, L., Collins, M., Conway, M., & Cunningham, K. (2013). Implicit chaining in cotton-top tamarins (*Saguinus oedipus*) with elements equated for probability of reinforcement. *Animal Cognition, 16*(4), 611–625.

Locurto, C., Fox, M., & Mazzella, A. (2015). Implicit learning in cotton-top tamarins (*Saguinus oedipus*) and pigeons (*Columba livia*). *Learning & Behavior, 43*(2), 129–142.

Locurto, C., Gagne, M., & Nutile, L. (2010). Characteristics of implicit chaining in cotton-top tamarins (*Saguinus oedipus*). *Animal Cognition, 13*(4), 617–629.

Loetscher, T., & Brugger, P. (2009). Random number generation in neglect patients reveals enhanced response stereotypy, but no neglect in number space. *Neuropsychologia, 47*(1), 276–279.

Lopes, L. L., & Oden, G. C. (1987). Distinguishing between random and nonrandom events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(3), 392–400.

Lyons, J., Weeks, D. J., & Elliott, D. (2013). The gambler's fallacy: A basic inhibitory process? *Frontiers in Psychology, 4*(72), 1–22.

McKinney, W. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, 51-56.

Mookherjee, D., & Sopher, B. (1994). Learning behavior in an experimental matching pennies game. *Games and Economic Behavior*.

Neuringer, A. (1986). Can people behave "randomly?": The role of feedback. *Journal of Experimental Psychology: General*, *115*(1), 62–75.

Neuringer, A. (2002). Operant variability : Evidence , functions , and theory. *Psychonomic Bulletin & Review*, *9*(4), 672–705.

Neuringer, A., & Jensen, G. (2012). The predictably unpredictable operant. *Comparative Cognition & Behavior Reviews*, *7*, 55–84.

Neuringer, A., & Jensen, G. (2013). Operant variability. In *APA handbook of behavior analysis, Vol. 1: Methods and principles*. (Vol. 1, pp. 513–546). http://doi.org/10.1037/13937-022

Nickerson, R. S. (2002). The production and perception of randomness. *Psychological Review*, *109*(2), 330–357.

Nickerson, R. S. (2004). *Cognition and chance: The psychology of probabilistic reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Nickerson, R. S., & Butler, S. F. (2012). On producing random binary sequences. *The American Journal of Psychology*, *122*(2), 141–151.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive psychology*, *19*(1), 1-32.

Olivola, C. Y., & Oppenheimer, D. M. (2008). Randomness in retrospect: exploring the interactions between memory and randomness cognition. *Psychonomic Bulletin & Review*, *15*(5), 991–996.

Parrish, A. E., Brosnan, S. F., Wilson, B. J., & Beran, M. J. (2014). Differential responding by rhesus monkeys (*Macaca mulatta*) and humans (*Homo sapiens*) to variable outcomes in the assurance game. *Animal Behavior and Cognition*, *1*(3), 215.

Paulos, J. A. (1989). *Innumeracy: Mathematical illiteracy and its consequences*. New York, New York: Hill and Wang.

Perdue, B. M., Evans, T. A, Washburn, D. A, Rumbaugh, D. M., & Beran, M. J. (2014). Do monkeys choose to choose? *Learning & Behavior, 42*(2), 164–75.

Python Software Foundation. Python Language Reference, version 2.7. Available at http://www.python.org

Rapoport, A., & Budescu, D. V. (1992). Generation of random series in two-person strictly competitive games. *Journal of Experimental Psychology: General, 121*(3), 352–363.

Richardson, W. K., Washburn, D. A, Hopkins, W. D., Savage-Rumbaugh, E. S., & Rumbaugh, D. M. (1990). The NASA/LRC Computerized Test System. *Behavior Research Methods, 22*(2), 127–131.

Rosenberg, S., Weber, N., Crocq, M.-A., Duval, F., & Macher, J. P. (1990). Random number generation by normal, alcoholic, and schizophrenic subjects. *Psychological Medicine, 20*, 953–960.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month Old Infants. *Science*, 274, 1926–1928.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8*(2), 101–105.

Sanabria, F., & Thrailkill, E. (2009). Pigeons (*Columba livia*) approach Nash equilibrium in experimental Matching Pennies competitions. *Journal of the Experimental Analysis of Behavior, 91*(2), 169–183.

Seabold, S., & Perktold, J. (2010). *Statsmodels: Econometric and statistical modeling with python*. Proceedings of the 9th Python in Science Conference.

Scheibehenne, B., Todd, P. M., & Wilke, A. (2011). Expectations of clumpy resources influence predictions of sequential events. *Evolution and Human Behavior, 32*(5), 326–333.

Scholl, S. G., & Greifeneder, R. (2011). Disentangling the effects of alternation rate and maximum run length on judgments of randomness. *Judgment and Decision Making, 6*(6), 531–541.

Sun, Y., & Wang, H. (2010). Perception of randomness: On the time of streaks. *Cognitive Psychology, 61*(4), 333–342.

Tolman, E. C. (1938). The determiners of behavior at a choice point. *Psychological Review, 45*(1), 1–41.

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics, 67*(5), 867–75.

Tune, G. (1964). A brief survey of variables that influence random-generation. *Perceptual and Motor Skills, 18*, 705–710.

Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General, 134*(4), 552–64.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience, 21*(10), 1934–45.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty : Heuristics and biases. *Science, 185*(4157), 1124–1131.

Washburn, D. A, Hopkins, W. D., & Rumbaugh, D. M. (1991). Perceived control in rhesus monkeys (*Macaca mulatta*): Enhanced video-task performance. *Journal of Experimental Psychology: Animal Behavior Processes, 17*(2), 123–9.

Wilke, A., Scheibehenne, B., Gaissmaier, W., McCanney, P., & Barrett, H. C. (2014). Illusionary pattern detection in habitual gamblers. *Evolution and Human Behavior, 35*(4), 291–297.

Williams, M. A., Moss, S. A., Bradshaw, J. L., & Rinehart, N. J. (2002). Random number generation in autism. *Journal of Autism and Developmental Disorders, 32*(1), 43–47.

Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychological Science, 24*(5), 667–677.

Zhao, J., Hahn, U., & Osherson, D. (2014). Perception and identification of random events. *Journal of Experimental Psychology: Human Perception and Performance, 40*(4), 1358–1371.

APPENDICES

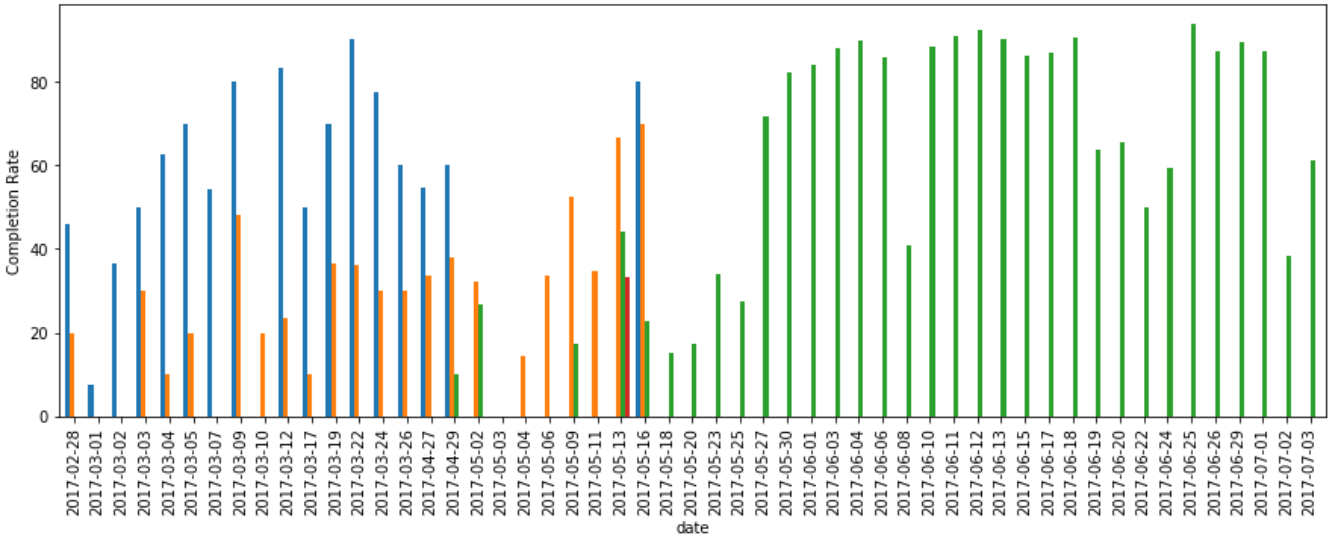
Appendix A: Sequence generation procedure for overalternating sequences

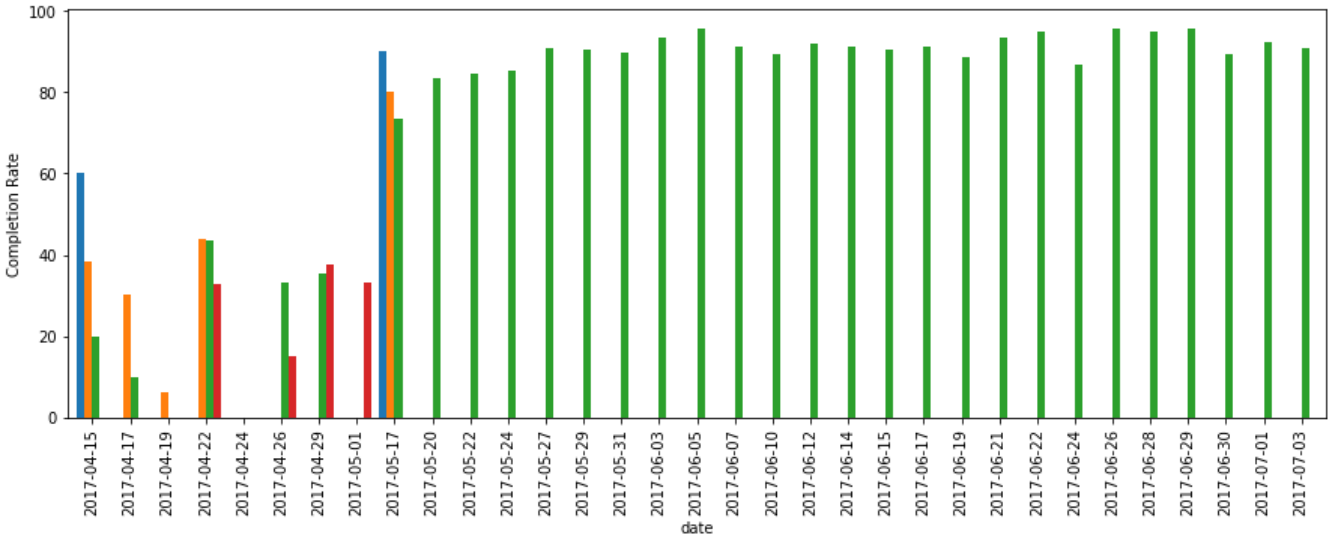
The overalternating sequences were designed to have a specified set of properties that serve both to fulfill previously-documented expectations that humans have of random sequences and to have a consistent internal structure:

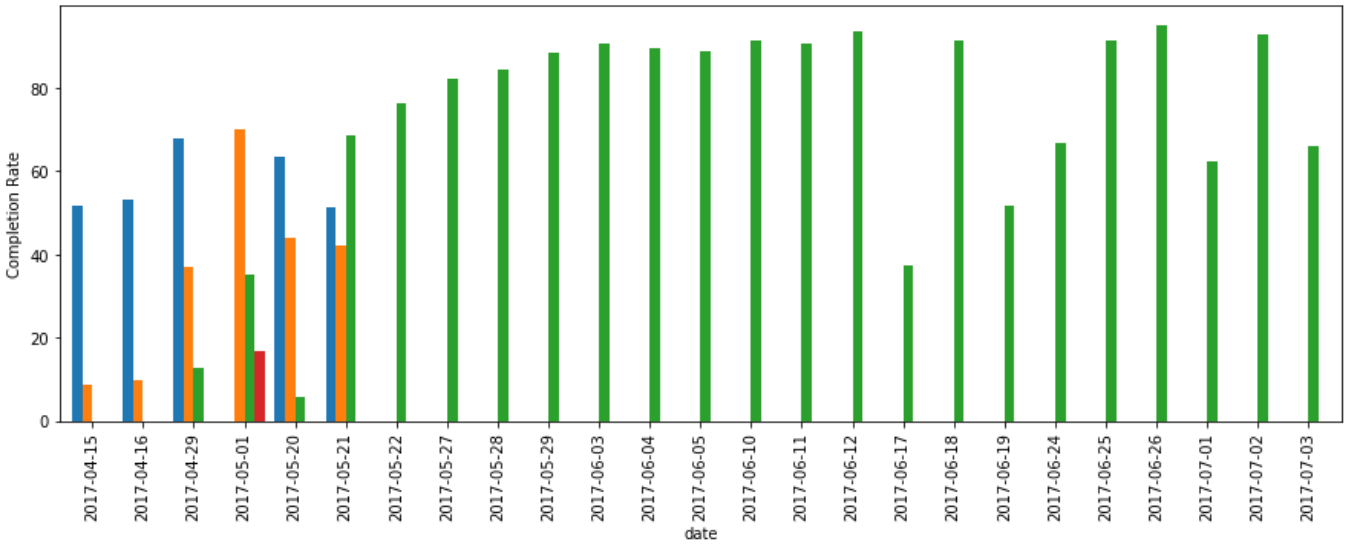
1. The first item in any sequence is [pseudo]randomly generated.
2. All four-item tuples are tested for three kinds of apparent serial dependencies:
    a. **All four items are identical**. In this case, the next item generated will always be different from the previous four.
    b. **Each item of the four is different from both the one preceding and succeeding it.** In this case, this apparent symmetry will be corrected by appending a fifth item that is identical to the fourth.
    c. **The four items are the concatenation of two runs of the same item twice.** In this case, the next item appended to the sequence will be the same as the fourth item of the tuple.
3. Any item that is not the first, and not following one of the corrected serial dependencies, alternates from the identity of the previous item 83% of the time. This number is used to preserve the global 65% alternation rate of the algorithm.

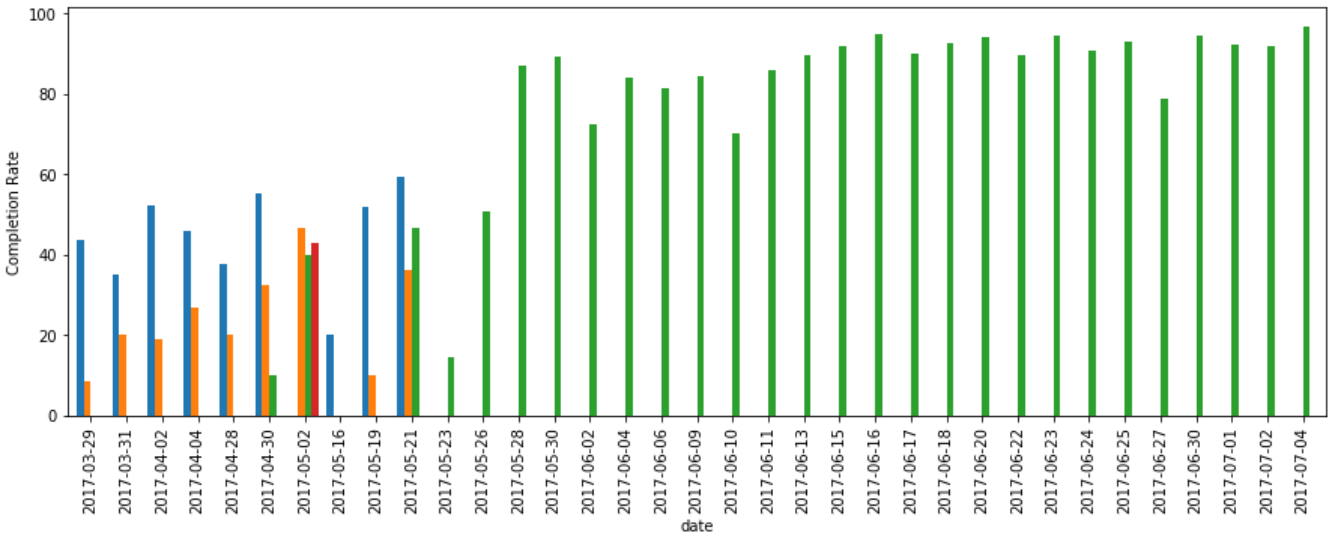
Appendix B: Training progress for each Rhesus macaque

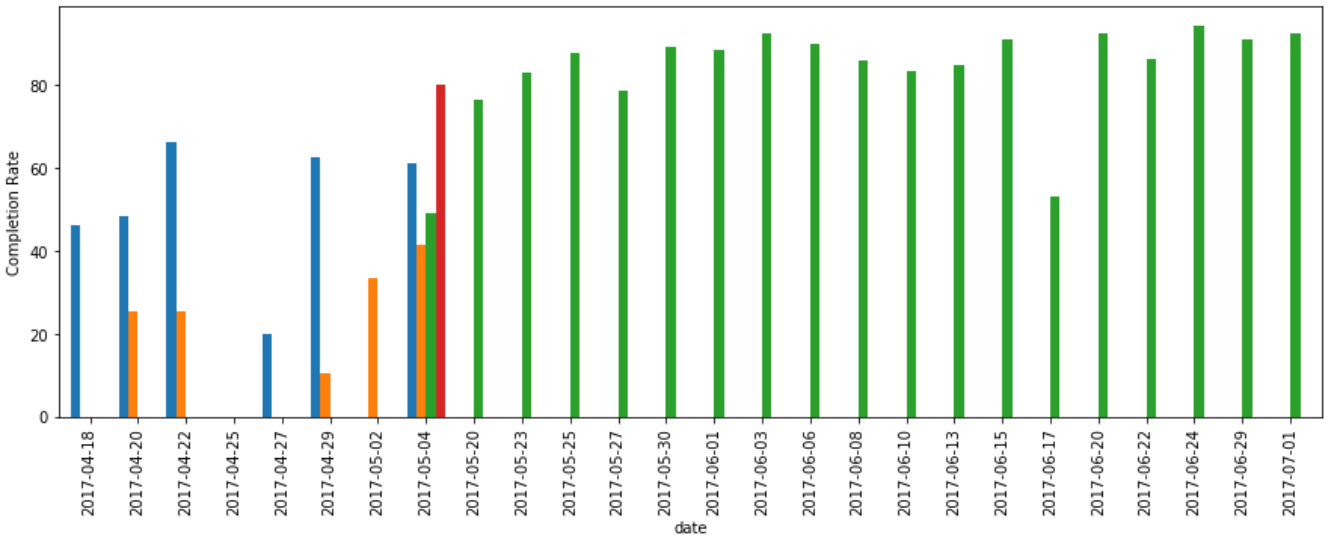
Plots are for Hank, Murph, Lou, Chewie, and Obi, respectively. Completion rates are the rates at which monkeys completed sequences that were 1 item long (blue bar), 3 items long (orange bar), 6 items long (green bar), or 9 items long (red bar).






Completion Rate
0
20
40
60
80
2017-04-15
2017-04-16
2017-04-29
2017-05-01
2017-05-20
2017-05-21
2017-05-22
2017-05-27
2017-05-28
2017-05-29
2017-06-03
2017-06-04
2017-06-05
2017-06-10
2017-06-11
2017-06-12
2017-06-17
2017-06-18
2017-06-19
2017-06-24
2017-06-25
2017-06-26
2017-07-01
2017-07-02
2017-07-03
date


Completion Rate
0
20
40
60
80
100
2017-03-29
2017-03-31
2017-04-02
2017-04-04
2017-04-28
2017-04-30
2017-05-02
2017-05-16
2017-05-19
2017-05-21
2017-05-23
2017-05-26
2017-05-28
2017-05-30
2017-06-02
2017-06-04
2017-06-06
2017-06-09
2017-06-10
2017-06-11
2017-06-13
2017-06-15
2017-06-16
2017-06-17
2017-06-18
2017-06-20
2017-06-22
2017-06-23
2017-06-24
2017-06-25
2017-06-27
2017-06-30
2017-07-01
2017-07-02
2017-07-04
date


Completion Rate
0
20
40
60
80
2017-04-18
2017-04-20
2017-04-22
2017-04-25
2017-04-27
2017-04-29
2017-05-02
2017-05-04
2017-05-20
2017-05-23
2017-05-25
2017-05-27
2017-05-30
2017-06-01
2017-06-03
2017-06-06
2017-06-08
2017-06-10
2017-06-13
2017-06-15
2017-06-17
2017-06-20
2017-06-22
2017-06-24
2017-06-29
2017-07-01
date