Beyond Dichotomy: Dynamics of Choice in Compositional Space

Greg Jensen

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2014

ABSTRACT

Beyond Dichotomy: Dynamics of Choice in Compositional Space

Greg Jensen

The quantitative study of choice under conditions of uncertainty dates back to the earliest applications of probability to games of chance. Over time, theories of choice have transitioned away from the 'oughts' of rational econometrics toward more face-valid descriptions of observed behaviors. Throughout this period, the problem of subjective probability has posed a consistent difficulty for theories of choice. The most successful approach for modeling these distortions is use of 'log-odds,' which provides a powerful description of two-alternative choice as a power law function of relative outcome probability. The log-odds approach can be generalized using the framework of 'compositional analysis.' The core statistical methodology of this framework is introduced and described, with an eye towards developing models of choice across any number of alternatives. The viability of these models is demonstrated on several previously published datasets.

A series of experiments with rats explored the effect of changing the number of alternatives. Power-law models continued to provide an effective description of behavior, but subjective probabilities were also found to be less distorted when subjects made choices among a larger number of alternatives (eight at once) than among smaller numbers (four or six). This effect was robust against controls for age, order of experience, chamber configuration, and schedule richness. A working hypothesis is put forward based on an analysis of responses as a dynamical process: Subjects succeed at complex tasks by limiting their transitions between response alternatives to a highly stereotyped 'default transition matrix,' making only slight deviations in order to adapt to changing task demands. This strategy is computationally efficient. However, severe mismatches between the schedule and a subject's default transition matrix are much more likely to occur when fewer alternatives are available, and behavior under such conditions is necessarily insensitive. Implications for other choice models are considered.

Table of Contents

List of Figures

List of Tables

Acknowledgments

Too many people deserve to be acknowledge for me to properly thank them all. We who have travelled far have done so carrying one another.

Thanks to Peter Balsam and Herb Terrace for their patience, encouragement, and mentorship throughout the years of graduate school. Thanks also to Alina Bica-Huiu, Francois-Charles Briands, Charles Gallistel, and Allen Neuringer for their ongoing interest and input throughout the course of developing this set of projects, and to Rotem Rusak and Jenny Porter for their marathons of editorial assistance.

Thanks to Brian Christian, Dave Freestone, Peter Killeen, and Ursula Whitcher for their comments my initial treatment of compositional analysis as it applied to the analysis of choice.

Thanks to Niall Bolger and Daniel Fürth for their early feedback on the development of the CPR algorithm.

Thanks to Kathleen Taylor for her oversight of the collection of animal data, and to Drew Altschul, Erin Danly, Jessica Joiner, Julia Kahn, Chris Mezias, Cait Williamson, and Elizabeth Yohe Moore for their work as lab technicians and managers.

Thanks to my fellow graduate students over the years, with a special emphasis on the members of my graduate cohort (James Cornwell, Jeff Craw, Juliet Davidow, Bruce Doré, Katherine Thompson Fox-Glassman, Barbie Huelser, Mariana Cunha Martins, Travis Riddle, Christine Webb, and Lisa Zaval) and lab partners (Billur Avlar and Matt Bailey). I can think of no more excellent band alongside whom to explore the edge of the known.

Thanks to my always-supportive parents, Mark & Agnes, to whom I owe my rapacious curiosity and to whom I am indebted for their tolerance of countless incoherent explanations of my research.

And of course, additional thanks to Peter Balsam, Niall Bolger, Charles Gallistel, Hakwan Lau, and Herb Terrace for agreeing to act as my dissertation committee.

To Annabelle, Hal, Jacqueline, and Paul

Chapter 1

Introduction

1.1 A Summary of Contents

This dissertation examines a range of theoretical propositions, analytic methods, and experimental results that inform our understanding of what it means to make repeated choices from among many alternatives. The general aim is to provide a comprehensive account of these topics, but this has resulted in a somewhat sprawling document. In order to help readers approach this material in an efficient manner, the following roadmap outlines the dissertation's contents, and each section's function in service of the document as a whole. The reader is encouraged to use this summary to roam a bit more freely in the text.

Throughout the text, a large number of equations are presented, with a correspondingly broad range of notation. Many of these, particularly those associated with compositional analysis, will be entirely unfamiliar to most readers. To facilitate comprehension, Appendix A provides a glossary of important mathematical notation and reprints every numbered equation, along with a brief description and the page number of its original appearance. Readers who read the dissertation in an order of their choosing are encouraged to use this appendix liberally to keep from losing the mathematical thread.

1.1.1 Introducing Choice

This introduction begins with a brief description of 'compositions,' a fundamental mathematical object central to the problem of choice. Although this same information is provided in greater detail at a later point, it is introduced from the outset to provide a lens through which to consider the many quantitative models proposed in the literature.

Subsequently, the introduction provides a historical backdrop to the quantitative study of choice and preference. Two major threads are considered, neither of which has interacted with the other until very recently. The first of these is the comparative work of learning theorists, such as the radical behaviorists. These models were developed primarily using data collected from animals in controlled experiments, in the hope of identifying the 'universal laws' of learning and behavior. The other literature is that of human decision making, developed in the traditions of econometrics and game theory. Much of the foundational work in economics relies more on mathematical axioms than on experimental evidence. These two threads come together in the modern study of 'behavioral economics.'

The experimental methods and results presented later in the dissertation do not depend directly on this historical backdrop. Nevertheless, understanding this historical context allows the experiments to be examined with an eye toward the often incompatible assumptions that have informed past models, and their resulting limitations.

1.1.2 Compositional Analysis

Part I of the dissertation is entitled 'Analytic Methodology' and provides a comprehensive overview of the basics of compositional analysis. 'Compositions' (briefly introduced in the following section of this introduction) are mathematical objects that share properties with both vectors and fractions, and they are the backbone of all subsequent analysis performed in this dissertation.

Several key points are meant to be taken from the chapters in Part I. The first is that compositions are necessarily constrained to a Non-Euclidean sample space, a fact with subtle but damning implications for many quantitative models of choice and preference. The second is that, despite being unfamiliar, this sample space is completely consistent as a vector space and host to a set of mathematical operations that make its analysis tractable. The third is that, using a generalization of the log-odd transformation, data of a compositional nature may be converted into a Euclidean vector space with no loss of information, permitting the full range of traditional statistics to be implemented.

Because compositional analysis is not limited to any particular discipline or topic, the chapters in Part I are written in abstract terms, without extensive application to the topics of choice, preference, and subjective value. A reader with a strong applied focus may find it more compelling to skip Part I initially, returning once the experiments and their analyses begin to make the nature of the problem more concrete.

1.1.3 Experiments & Results

Part II makes use of the tools of compositional analysis to study behavior when many choice alternatives are available. This is done using data collected from animals in laboratory contexts, both in previously published data and in experiments performed for this dissertation. The data from many of these experiments could not have been rigorously analyzed using earlier models, and the compositional paradigm provides a first look at how behavior adapts to changing task complexity.

Although compositions make the analysis of these data possible, the initial results are highly surprising and are not easily explained by extant theories of choice. A working hypothesis is put forward to explain these results, and its development requires the introduction of two other quantitative tools: change-point analysis (used to study the behavior as a time series) and information theory (used to measure the efficiency of behavioral strategies). These are introduced as the need arises in the Results section of Chapter 9. Because the change-point algorithm employed is somewhat involved, it is described at greater length in Appendix B.

Through the study of behavior as a process, Part II assembles these data into a working hypothesis regarding the way in which response structure and task structure interact. This yields a heuristic that scales very efficiently as the complexity of the task increases. Not only does this provide a plausible strategy used by subjects to perform complex tasks, but also makes a coherent prediction for what circumstances are necessary for that strategy to fail.

1.1.4 General Discussion

Part III examines the implications of the working hypothesis put forward to explain the experimental data. Although a summary of the results is provided, a full appreciation of the proposed mechanism depends on a clear image of the way in which the conditional probabilities of responding (as characterized by Markov chains) relate to the information-theoretic measures described in Part

II (particularly the 'divergence rate,' described as Equation 9.7).

A detailed consideration of the implications of compositional analysis and responding as a process are presented as they relate to existing models of decision making. Particular attention is given to finding a happy medium between radical behaviorism (in which the environment is dominant and the organism nearly vanishes) and representational cognition (which places the complete burden of simulating the world on the organism). This dissertation's working hypothesis is that although subjects initially make use of computationally costly strategies, these quickly crystalize when the layout of the environment is static, allowing successful behavior to persist and adapt to changing schedules of reward with minimal cognitive costs.

Several lines of research are proposed to flesh out the working hypothesis. In addition to further validation of the basic claim by studying animal behavior, experiments are also proposed to examine the question in human decision making (using simple games) and in neuroscience (by enhancing existing accumulator models).

1.2 Compositions: A Rigorous Approach to Relative Value

The proposal that 'value is relative,' in both figurative and mathematical terms, is hundreds of years old. Inspired by the theory of measurement put forward by Stevens (1946), many modern social scientists might frame the proposal as 'value is measured on a ratio scale.' However, Stevens' different scales of data (nominal, ordinal, scalar, and ratio) have been controversial among statisticians from the outset (Velleman & Wilkinson, 1993), and many rigorous quantitative problems do not conform to any of the classical categories (Chrisman, 1998). Far from being exotic or specialized, many familiar varieties of data fail to fit any of Stevens' categories.

Perhaps the most widely studied variety of non-standard data are 'proportions,' which include probabilities. Although 'relative odds' are often considered, suggesting that a ratio scale may be appropriate, proportions have both a true zero *and* a true maximum, since no single item can represent more that 100% of the total. As a result, proportions are fundamentally *constrained* data, giving them characteristics not commonly attributed to ratio scales. Understanding the nature of this constraint is essential to the sensible development of both analysis and theory.

We commonly represent ratios of two alternatives as fractions, and because the properties of these objects are taught to us in elementary school, they seem natural and obvious. We understand, for example, that $\frac{1}{4}$ and $\frac{10}{40}$ represent identical values. How one generalizes this principle of 'relative value' to sets of more than two measurements is not obvious, however, and most analysts are unfamiliar with a class of mathematical objects that might correspond to a 'many-sided fraction.' Such an object would automatically encode that the ratio $(1 : 5 : 14)$ is identical in relative terms to the probabilities $[0.05, 0.25, 0.7]$.

This object is the *composition*, which consists of some set of D alternatives that are only meaningful relative to one another[1]. This strictly relative property arises as a consequence of the *closure constraint* (denoted by $\mathcal{C}(\cdot)$), which converts a vector of observations $\mathbf{B} = [B_1, \cdots, B_n]$ into a set of relative frequencies:

$$\mathcal{C}(\mathbf{B}) = \left[\frac{B_1}{\sum_{j=1}^{n} B_j}, \cdots, \frac{B_n}{\sum_{j=1}^{n} B_j}\right] \tag{1.1}$$

The closure constraint results in several counterintuitive properties, because it confines the data to a finite sample space whose geometry is non-Euclidean. This can be conveyed visually for three alternatives using a 'ternary diagram,' as depicted in Figure 1.1. Here, the three vertices of the equilateral triangle, p_1, p_2, and p_3, each correspond to 100% of one of three alternatives, and any point within this triangle represents some other set of proportions. The gray point in the center, for example, corresponds to the proportions $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$, which is the 'barycenter' of the triangle. Drawn in 3D Euclidean space, this triangle would lie diagonally with respect to the axes, connecting the points (0,0,1), (0,1,0), and (1,0,0). This leads to the two most obvious properties of compositional data: Points are confined to a finite domain, and that domain has one fewer dimensions than the number of alternatives (in this case, a 2D triangle embedded in 3D space). This finite domain is called a 'simplex' (the n-dimensional generalization of an equilateral triangle).

A less obvious characteristic of the closure constraint is that the space within the simplex is non-Euclidean. Since everything within the simplicial geometry is relative, it is multiplication (rather than addition) that acts as the basic arithmetic operation (see Equation 3.3). When working with compositions, although multiplication and division are consistent mathematical operations,

[1] Compositions are very closely related to, but distinct from, multinomial vectors. For example, the multinomial vector (1, 5, 14) encodes both relative *and* absolute information. The vector that encodes the parameters of the multinomial distribution, however, qualifies as a composition, and is the form that analysts are most likely to have encountered compositions in the past.

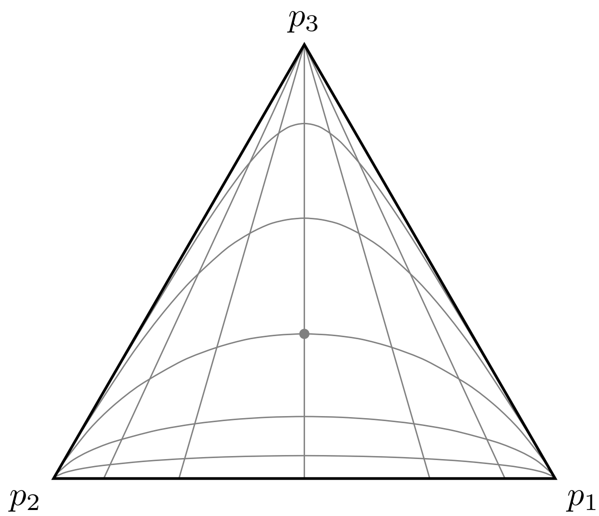

Figure 1.1: A ternary plot comparing three proportions p_1, p_2, and p_3 (whose sum is 1.0, due to closure). The gridlines represent two compositionally orthogonal sets of compositionally parallel lines, representing a possible orthonormal contrast. No single vertex has priority over the others, and the non-Euclidean geometry is fully symmetric.

addition and subtraction are *not*. It is for this reason that the compositional sample space is inconsistent with Stevens' 'ratio scale,' as the latter permits both addition and multiplication. The inconsistency of simple addition in compositional data has profound implications for hypothesis testing, because the arithmetic mean and its corresponding variance also become inconsistent.

The result of these constraints is an unfamiliar (but internally coherent) form of vector geometry. For example, Figure 1.1 depicts two sets of 'compositional lines,' defined according to a non-Euclidean vector space within the simplex. Although some lines appear curved while others do not (according to a Euclidean definition of curvature), every one of them is linear with respect to the (non-Euclidean) geometry of the simplex. Furthermore, each of the depicted lines is either 'compositionally parallel' or 'compositionally perpendicular' to every other line drawn. The lines belong to two groups, each following the contours of the contrast between p_3 and a composite of p_1 and p_2 (in the case of the vertical lines) and the contrast of p_1 vs. p_2 independent of p_3 (in the case of the horizontal lines). Because this geometry is fully symmetrical, no single vertex has priority, any other orthonormal contrast not depicted could in principle be drawn.

Another non-obvious property of closure-transformed data is that even when three independent random variables A, B, and C are unrelated, the covariance of $\mathcal{C}([A, B, C])$ will automatically be negative, because more of one alternative necessarily means less of the others. At the same time, the covariance matrix will be singular (as a result of the insufficient degrees of freedom in the geometry). This can render the behavior of compositional data perplexing in the event that its compositional character is not recognized.

Thanks to techniques pioneered by Aitchison (1986) and subsequently refined by Egozcue & Pawlowsky-Glahn (2006), this unfamiliar geometry can be transformed into a sample space that has all the properties traditionally associated with a Euclidean sample space. Using non-standard mathematical operators, the contents of Figure 1.1 may be represented as a consistent vector space[2], within which familiar hypothesis testing tools may be applied.

Interpreting the relative selection of alternatives using compositional methods (reviewed in Pawlowsky-Glahn & Buccianti, 2011) offers substantive advantages. From a theoretical perspective, the most important of these is the recursive character of the simplex: Each (n)-simplex is composed of a set of $(n-1)$-simplexes, which ensures that the relationships between alternatives are undistorted by arbitrary analytic decisions. For example, in a symmetrical model, the parameters estimated for each (n)-simplex will be consistent with parameters estimated from data that only include a subset of the alternatives. This is particularly important because the analyst may not know what the organism considers to be its alternatives (for example, 'go to sleep' may be one of the alternatives from which a rat chooses). A model that yields consistent parameters given subsets of the data can also be expected to be consistent if the analyst lacks sufficient information to encode every possible alternative. If, on the other hand, behavior displays asymmetries, these can be studied in terms of compositional covariance, which relaxes the assumption that every alternative B_i must be entirely discriminable from every other alternative B_j. A case in which this solution is effective is described in Chapter 8.

From a practical perspective, compositional transformations permit proportional data to be examined in Euclidean terms, enabling the full range of parametric statistics to be brought to bear on problems of choice.

[2]This domain may also be transformed into a Euclidean space with no loss of information using either the 'centered log-ratio' transformation (Equation 4.3) or the 'isometric log-ratio transformation' (Equation 4.5), which will be discussed in detail henceforth (see also Egozcue et al., 2003; Egozcue & Pawlowsky-Glahn, 2006).

1.3 Classical Accounts of Choice

Prior to the 17th century, choice was a predominantly philosophical topic, limited to considerations of ethics and dogma. However, the foundational work by Pascal, Fermat, and Huygens in the study of probability (Ore, 1960) hinged on a dilemma about which traditional ethics appeared mute: If a game of chance was believed to be governed by random events, and the game ended early (that is, while money was still on the table), what would be the 'fair' way to distribute the money to the participants? Framed in broader terms: Given uncertainty about future outcomes, how should some resource (such as money or effort) be distributed in order to represent those outcomes in a fair way? This question led to the formulation of *expected value* as a 'mathematically fair' means of understanding games of chance. In the context of a lottery A with D different outcomes, the expected value L_A is defined as the sum of products between the magnitudes M_i of each outcome i, and that outcome's probability p_i:

$$L_A = \sum_{i \in D} p_{i,A} \cdot M_{i,A} \tag{1.2}$$

This very basic formulation carries both a scalar and an ordinal implication. From a scalar perspective, L_A is a value in the same monetary units as each of the outcomes $M_{i,A}$, such that expected values and concrete monetary amounts are treated interchangeably. From an ordinal perspective, expected values may be used to specify preference orderings[3], for example $A \succeq B$ if $L_A \geq L_B$. Notice that both implications fail to take the uncertainties of L_A and L_B into consideration.

1.3.1 Expected Utility Theory

In its historical context, *expected utility* provides an elegant solution to what was once considered a thorny problem: Why won't people pay an exorbitant amount to play a game whose expected value is infinite? More generally, why do some gamblers favor risk, while others avoid it? Bernoulli

[3] The symbols $\succ$ and $\prec$ correspond to 'succeeds' and 'precedes,' respectively. Rather than comparing two relative measures (as the $>$ and $<$ symbols do), successions and precession are prescriptive. The statement $X \succ Y$ can be read as, "X should always be favored over Y." Analogously, $\succeq$ and $\preceq$ indicate the possibility of indifference, such that $X \succeq Y$ can be read as, "X may be favored over Y, or the two may be considered equivalent." Rounding out this notation, $X \sim Y$ denotes that X and Y are similar under transformation, $X \simeq Y$ denotes that X and Y are either similar or exactly equivalent, and $X \equiv Y$ denotes that X and Y are identical.

(1738|1954) proposed that, rather than using expected value, gamblers might instead experience some different subjective function of utility, which could capture the experience of diminishing returns as a function of wealth. Expressed as a power law, this takes the form:

$$EU\left(L_A|\beta\right) = \sum_{i \in D} p_{i,A} \cdot M_{i,A}^{\beta} \tag{1.3}$$

Here, the exponent β distorts the magnitude of each outcome in the lottery, and these distorted values are used in computing the utility. Note that the expected utility is *not* presented in the original units (e.g. dollars), but rather in some subjective units, due to the transformation imposed by β. These subjective units are typically denoted by u (for 'utility') and are sometimes called 'utils.'

Bernoulli's proposal made an early distinction between *value* and *odds of occurence*, noting the potentially subjective quality of the former. Although the power-law formulation would be revived as the Cobb-Douglas production function (Cobb & Douglas, 1928), the full mathematical implications of expected utility would not be realized until the codification of the von Neumann-Morgenstern utility theorem (von Neumann & Morgenstern, 1947). This theorem laid out four axioms that were necessary and sufficient for 'subjectively rational' behavior under the non-linear functions of value. This transformative work led to an explosion of theorizing arising from the principles of game theory. However, it was observed almost immediately that participants routinely violate these rationality axioms. This was most famously argued by Allais (1953) in what is now termed the 'Allais paradox.' Such paradoxes arise whenever two lotteries with identical expected values yield different preferences. For example, being given a guaranteed \$50 will be seen as favorable relative to a gamble that has a 0.75 probability of paying \$60 by some participants, an econometrically reasonable outcome because $\$50 > 0.75 \cdot \$60$. However, of those participants, a further subset will prefer a 0.75 probability of winning \$60 to a 0.25 probability of winning \$200, even though $0.75 \cdot \$60 < 0.25 \cdot \$200$. Participants whose preference flip in this way are, according to Allais, considered to violate the axioms of rationality, and hence be 'paradoxical.'

The implied limits of these rationality axioms have been widely misunderstood. It is far from clear that von Neumann and Morgenstern intended for their axioms to have a prescriptive quality (i.e. that people ''ought' to be rational); instead, their original theorem merely states that *if* a decider were 'subjectively rational' (according to the classical sense of the word rational), then their behavior should conform to the axioms (Fishburn, 1989).

In their original treatment, Von Neumann and Morgenstern speculate that a *subjective probability function* could be introduced to the model, but they do not rigorously derive the set of theorems that would result from such a model. The first such subjective function was specified by Pfanzagl (1967), but was limited to two alternatives. The first complete generalization to arbitrarily large sets of outcomes was provided by Luce & Krantz (1971) as 'conditional expected utility,' also termed 'subjective expected utility' (Luce, 1969).

The proposal that subjective probability may be represented as a conditional probability distribution carried with it the implication of the closure constraint. Simply put, transformed probabilities must still sum to 1.0 or else they fail to be probabilities in the first place. Although this is not given prominent treatment (presumably because it was taken to be a trivial consequence of working with probability), closure is built into the axioms specified by Luce & Krantz. A simple implementation of this is 'subjective weighted expected utility' (*SWEU*), which explicitly invokes closure when transforming the probabilities:

$$SWEU\left(L_A|\alpha,\beta\right)=\sum_{i\in D}\frac{p_{i,A}^{\alpha}}{\sum_{j\in D}p_{j,A}^{\alpha}}\cdot M_{i,A}^{\beta} \tag{1.4}$$

Here, the parameter β retains its original role as an exponent, distorting the magnitude of the reward outcomes. The new parameter, α, provides a corresponding exponential distortion to the subjective probabilities. As in expected utility, the subjective expected utility is measured in some subjective units u rather than in the objective units originally specified by the problem. Equation 1.4 can be attributed to Karmarkar (1978).

In both Equations 1.3 and 1.4, interpreting the resulting 'utility' measures is made difficult by the subjective units u. Returning the results to their original units simply required reversing the magnitude operation, which yields a *subjective expected value*, denoted by $\mathcal{L}_A$:

$$\mathcal{L}_A=\left(\sum_{i\in D}\frac{p_{i,A}^{\alpha}}{\sum_{j\in D}p_{j,A}^{\alpha}}\cdot M_{i,A}^{\beta}\right)^{\frac{1}{\beta}} \tag{1.5}$$

Note that this construction includes sure-thing payouts, which take the form of 'lotteries' with a single outcome whose probability is 1.0: since β and $\frac{1}{\beta}$ cancel each other out, a no-risk objective outcome of $10 has a subjective value of $10, without needing to make reference to a participant's (presumably private and implicit) units of utility u. Consequently, other risky-choice scenarios may be evaluated in terms of the original 'objective' units, substantially facilitating interpretation.

To understand what it would mean for both value and probability to be subject to transformations, consider the visualized example in Figure 1.2. In each case, the same two lotteries A and B are presented, and the total area in gray for a lottery is proportional to its experienced value. The top pairing presents the expected values, as specified by Equation 1.2. In strictly objective terms, these two lotteries should both be valued[4] at \$5.2.

In the second row, the expected utilities for each lottery are presented, assuming the power law in Equation 1.3 and a coefficient of $\beta = 0.5$. This corresponds to a somewhat risk-averse individual: Not only is lottery A (\$4.59) preferred over lottery B (\$4.16), but a flat payoff of \$5.2 would be favored over both lotteries[5] . Another way to consider this hypothetical individual is that they are relatively *insensitive* to the contrasts between the outcomes, such that larger dollar amounts seem less different from smaller amounts than is objectively true.

In the third row, the subjective expected utilities (per Equation 1.4) are presented with an additional subjective probability parameter of $\alpha = 2$. Note that $\beta = 0.5$, as in the middle row. Despite this, the preference reverses to favoring lottery B (\$5.76) over lottery A (\$3.78). Such an individual would be insensitive to outcome magnitudes but *hypersensitive* to outcome probabilities. This yields the somewhat surprising behavior that lottery B is favored[6] over both lottery A *and* a flat payout of \$5.2.

Although Equation 1.4 follows naturally from a simple implementation of operations proposed by Luce & Krantz, almost all subsequent modeling of choice and preference has marginalized or ignored the closure constraint on probabilities. Among this dissertation's objectives is to show that formally acknowledging closure yields intuitive models whose parameters are straightforward to interpret. At the same time, a failure to accommodate closure (particularly when attempting to simulate its characteristics indirectly) can lead to considerable theoretical confusion.

---

[4] Or, more formally, if $L_A = L_B = \$5.2$, then the 'objectively rational' preference should be that $\mathcal{L}_A \sim \mathcal{L}_B \sim \$5.2$.

[5] Or, $\$5.2 \succ \mathcal{L}_A \succ \mathcal{L}_B$.

[6] Or, $\mathcal{L}_B \succ \$5.2 \succ \mathcal{L}_A$.

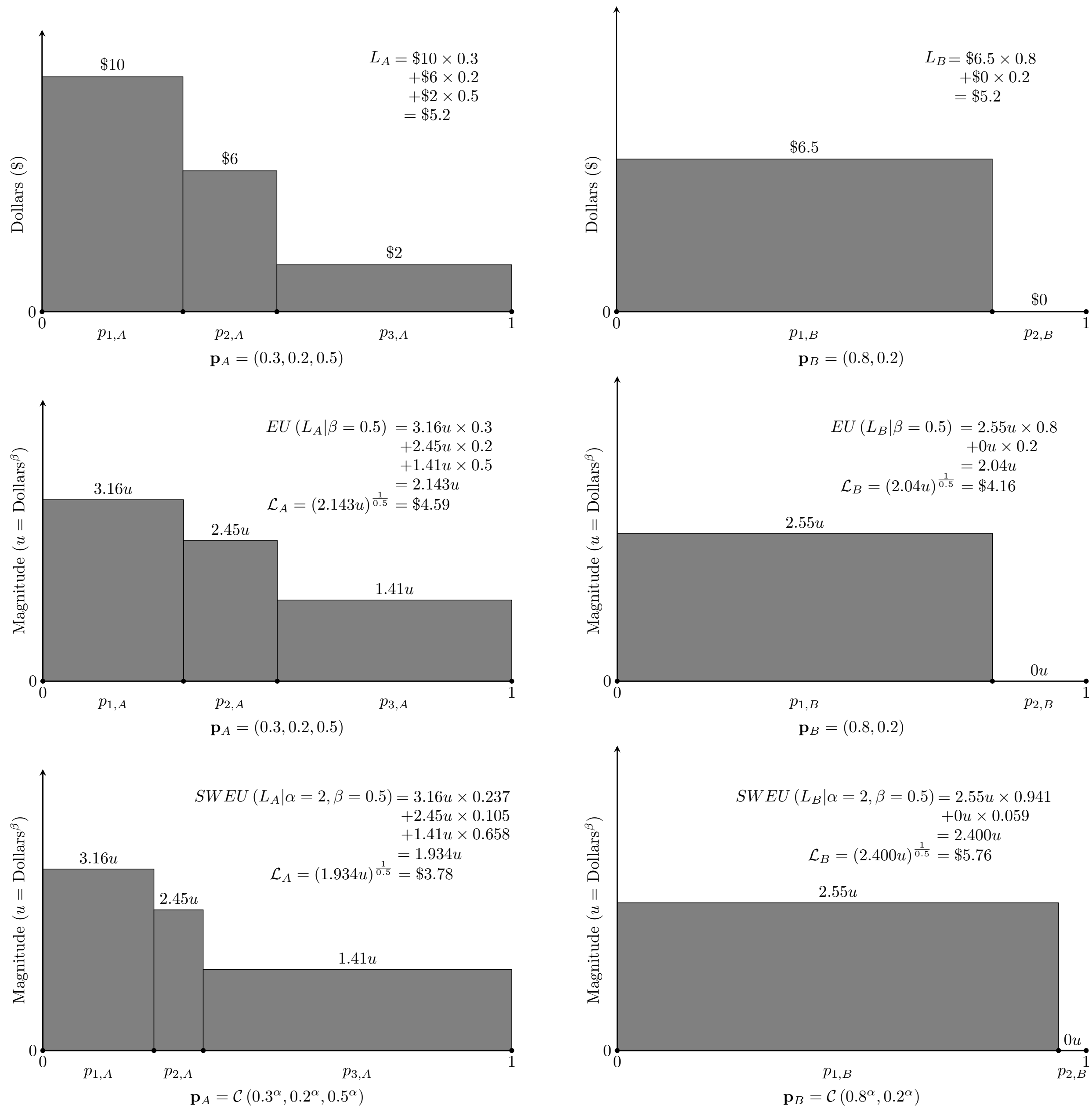

Figure 1.2: Visualization of expected value (Equation 1.2, top), expected utility (Equation 1.3, middle), and subjective expected utility (Equation 1.4, bottom) between two lotteries, A and B. The two lotteries have identical expected values $\mathcal{L}_A = \mathcal{L}_B = \$5.2$, but given a subjective magnitude parameters of $\beta = 0.5$, the expected utility of lottery A exceeds that of lottery B. However, if a subjective probability parameter $\alpha = 2$ is also applied, the resulting subjective expected utility for lottery B exceeds that of lottery A. Notice that in all cases, the probabilities $\mathbf{p}_x$ must sum to 1.0, which is ensured by the closure operation $\mathcal{C}(\mathbf{p}_x)$.

1.4 Choice and Learning Theory

1.4.1 Generalized Matching

The behaviorist tradition has a long history of studying choice. Its most influential formal theory is undoubtedly Herrnstein's matching law (reviewed in Herrnstein, 1997), which takes two closely related forms:

$$\frac{B_1}{B_1 + B_2 + \cdots + B_n} = \frac{R_1}{R_1 + R_2 + \cdots + R_n} \tag{1.6}$$

$$\frac{B_i}{t} = \frac{k_i}{t} \cdot \frac{R_i}{R_i + r_e} \tag{1.7}$$

Equation 1.6 is Herrnstein's original matching law (Herrnstein, 1961). It is sometimes simply called 'strict matching' because it predicts a one-to-one relationship between 'response strength' B_i and the corresponding frequency of reinforcement R_i. Equation 1.7 is sometimes called 'Herrnstein's hyperbola' and predicts response rate (i.e. responses B_i per unit time t) as a function of the asymptotic response strength k and competing reinforcement r_e (de Villiers & Herrnstein, 1976). These two formulations are arithmetically equivalent when all k_i are equal. The implicitly compositional character of Equation 1.6 is obvious, as it is equivalent to the statement $\mathcal{C}(\mathbf{B}) = \mathcal{C}(\mathbf{R})$.

Although Herrnstein's matching law remains influential in a wide range of fields, (e.g. Sugrue et al., 2004; Sakai & Fukai, 2008), and although Equation 1.7 fits certain kinds of data exceptionally well, the general premise of strict matching has been invalidated by extensive experimentation (reviewed in Davison & McCarthy, 1988). Most notably, consistent deviations from strict matching are observed in schedules with two concurrently available alternatives whose outcomes are probabilistic.

Most of strict matching's failures to explain empirical results (at least in molar terms) can be addressed by the 'generalized matching law,' first proposed[7] by Baum (1974) as a description of asymptotic choice behavior:

$$\frac{B_1}{B_2} = \frac{k_1}{k_2} \cdot \left(\frac{R_1}{R_2}\right)^s \tag{1.8}$$

Here, the ratio of responses B to two alternatives 1 and 2 (typically levers in rodent experiment or wall-mounted keys in pigeon experiments) are predicted as a function of the ratio of reinforcers R earned from each. This reinforcer ratio is raised to an exponent s (the 'sensitivity') and multiplied

[7]Remarkably, the original treatise on expected utility by Bernoulli (1738|1954) predicted that models similar to strict matching were likely to fail, and its model of expected utility (Equation 1.3) is almost identical to Equation 1.8.

by the ratio $\frac{k_1}{k_2}$ (the 'bias,' often simply identified as k in the literature). Since B and R belong to ratios in which units cancel out, this equation ignores the absolute frequencies of occurrence and makes a prediction solely on their relative rates.

The variable k_i plays a similar role in Equation 1.7 to that of the ratio $\frac{k_1}{k_2}$ in Equation 1.8, in that both represent unitless measures of 'value.' The bias ratio introduces a systematic skew to the subjective value of obtained rewards. If, for example, a subject was presented with a schedule in which $R_X = R_Y$, then any deviation from equal response allocation is explained in terms of the bias ratio skewing the subject's preference. The broader implication is that the multiplicative nature of bias should be consistent across reinforcer ratios.

It is sensitivity, however, that sets Equation 1.8 apart from preceding models. This exponential parameter describes, in terms of curvature, how the contrast between the two outcomes influences their relative frequency of selection. If $s = 0.0$, then the subject is indifferent to the relative frequency of payoffs, and bias will be the only factor predicting how often an alternative is selected. If, on the other hand, $s = 1.0$, then behavior will adjust on a 1:1 basis relative to the ratio of rewards, and if $s > 1.0$, shifts in reward will result in exaggerated shifts in responding, skewing responding toward exclusive selection of the 'subjectively better' alternative. As a shorthand, $s \approx 1.0$ is called 'matching,' $s < 1.0$ is called 'undermatching,' and $s > 1.0$ is called 'overmatching.' Importantly, there does not appear to be a 'universal' sensitivity parameter: Human participants show considerable individual differences (Kollins et al., 1997), and different species show dramatic difference in their range of likely values (Baum, 1979).

Note that, strictly speaking, neither bias nor sensitivity constitutes a subject's *preference*. Instead, preference is a function of both parameters. A major motivation for the form given by Equation 1.8 is that it can easily be linearized using a log-transformation:

$$\log\left(\frac{B_1}{B_2}\right) = \log\left(\frac{k_1}{k_2}\right) + s \cdot \log\left(\frac{R_1}{R_2}\right) \tag{1.9}$$

Provided the bias ratio is treated as a single parameter, estimates can easily be obtained using a linear regression[8], where the bias corresponds to the intercept and the sensitivity corresponds to the slope on log-log coordinates.

[8]Note that, in practice, regressions of log-transformed relative proportions display considerable heteroscedasticity. As a result, OLS parameter estimates should be interpreted with caution, and robust methods, such as weighted least squares (Jacquez et al., 1968) should be favored when applicable.

1.4.2 Compositional Geometry and the Barycentric Matching Model

On the one hand, the hypothesis that probabilities are perceived as log-odds ratios has received considerable support in recent years, with some going so far as to describe the relationship as 'ubiquitous' (H. Zhang & Maloney, 2012) because such ratios reliably describe a wide variety of psychological phenomena. On the other hand, although strict matching is demonstrably false, it retains one clear advantage over Equation 1.8, which is the ability to make a prediction for more than two alternatives. In order for a matching model to be fully generalized, the log-odds approach should make reasonable predictions given an arbitrary number of alternatives. Unfortunately, modifying Equation 1.9 to accommodate additional alternatives is an ambiguous proposition with multiple solutions.

The most obvious generalization is accomplished by adding a constant to both sides of the equation (Goldstein & Einhorn, 1987). Unfortunately, because of the widespread use of k instead of $\frac{k_1}{k_2}$, the arithmetic generalization is typically presented as:

$$\frac{B_i}{\sum_{j=1}^{n} B_j} = \frac{k \cdot (R_i)^s}{k \cdot (R_i)^s + \left[\sum_{j \neq i} (R_j)^s\right]} \tag{1.10}$$

This formulation is, in a limited sense, identical to Equation 1.8, but only if k is divorced from its functional role as a ratio of relative bias. Indeed, when more than two alternatives are included in the equation, the values of k and s will change as a function of which alternative i is the 'reference item' (i.e. singled out by the numerator) for any value of s other than one. This inconsistency is the result of a phenomenon called 'symmetric approximation by leading term' (or SALT) (Natapoff, 1970). Consequently, although best-fitting parameters for k and s can be obtained numerically for any given pre-determined reference item, the resulting parameters are neither consistent nor interpretable unless all bias values $k_i \approx 1$. This parametric inconsistency is sufficient grounds to reject Equation 1.10 as being malformed. Despite this, Equation 1.10 is surprisingly widespread (Lattimore et al., 1992; Tversky & Fox, 1995; Gonzalez & Wu, 1999; Trepel et al., 2005; Abdellaoui et al., 2005).

In order to preserve both the theoretical interpretability of the parameters and their internal symmetry, consistent generalization benefits from representing a set of options using vector notation, such that $\mathbf{B} = [B_1, \cdots, B_n]$. Given this notation, a more effective general form of the matching law invokes the centered log-ratio (or CLR) transformation, which relates each alternative to the geometric mean of all the alternatives:

$$\mathbf{B}^{\circ} = \mathrm{clr}\left(\mathbf{B}\right) = \left[\log\left(\frac{B_1}{\sqrt[n]{\prod_{j=1}^{n} B_j}}\right), \cdots, \log\left(\frac{B_n}{\sqrt[n]{\prod_{j=1}^{n} B_j}}\right)\right] \tag{4.3}$$

Since $\mathbf{B}^{\circ}$ denotes a set of log-ratios, this notation can be used to describe a highly general form of the matching law in the following succinct way:

$$\mathbf{B}^{\circ} = \mathbf{k}^{\circ} + s \cdot \mathbf{R}^{\circ} \tag{1.11}$$

Here, each alternative B_i is considered as a relative value (as in Equation 1.8), but uses the geometric mean as the 'barycenter' of the observations. These are then interpreted on a log scale (as in Equation 1.9). Originally proposed by Aitchison (1986), the CLR transformation is one of the backbones of 'compositional' data analysis, which considers the statistical properties of the relative proportions (rather than absolute values) of related sets. The first empirical demonstration linking the CLR transformation to the generalized matching law was provided in principle by S. M. Schneider & Davison (2005), in a study of response sequences. It was subsequently framed explicitly as the 'barycentric matching model' by Jensen & Neuringer (2009), with the following formulation:

$$\frac{k_1 \cdot (R_1)^s}{B_1} = \cdots = \frac{k_n \cdot (R_n)^s}{B_n} \tag{1.12}$$

This formulation is arithmetically identical to Equation 1.11, but conveys the intuitive relationship between responses B and outcomes R in a straightforward fashion. Unlike Equation 1.10, this formulation retains the compositional symmetry of the model, giving no special preference to one alternative over the others.

1.4.3 Dynamics of Choice

Although Equation 1.11 provides an effective description of stable asymptotic distributions of behavior (i.e. behavior at the 'molar' level of analysis), it provides no information about either the learning process or about response-by-response dynamic processes (i.e. the 'molecular' level of analysis). In general, analysts whose theories emphasize molar features have a competitive relationship with those interested in 'molecular' processes, and the integration of the two has been limited until relatively recently (Hineline, 2001).

In some cases, molar analyses may provide insight into the sequential regularities in behavior. For example, S. M. Schneider & Davison (2005) studied two-alternative choice in terms of the four possible two-item pairings. By embracing a full compositional approach, their two-operandum data can be represented as a four-alternative scenario. This can be represented in hierarchical terms, such that the ratio $\frac{\Pr(B_1)}{\Pr(B_2)}$ can be considered simultaneously with the conditional probability ratios $\frac{\Pr(B_1|B_1)}{\Pr(B_2|B_1)}$ and $\frac{\Pr(B_1|B_2)}{\Pr(B_2|B_2)}$, each corresponding to a different orthogonal factor in a compositional transformation. Such an approach can also be used to examine the degree to which behavior acutely changes in response to individual outcomes (Baum & Davison, 2004).

An important question, however, is whether the steady-state equilibrium of stable behavior differs in kind from the process by which subjects adapt to substantial shifts in schedules. Recent studies have eroded the foundation of 'reinforcement learning' models, both by demonstrating precise modifications of behavior in the absence of detectable differences of reinforcement (Kheifets & Gallistel, 2012) and by bringing into question reinforcement as a theoretical construct (Baum, 2012). Furthermore, the shortcomings of contiguity-based models of learning, relative to contingency-based models, are becoming increasingly evident in neuroscience (Loewenstein & Seung, 2009; Gallistel & King, 2010). In light of these developments, a promising alternative approach is to use information theory to study how subjects discover contingent relationships between outcomes and behavior (Jensen et al., 2013; Ward et al., 2013).

Given these developments, it is increasingly clear that characterizing how patterns of choice adjust in light of feedback will depend on developing a more sophisticated understanding of how choice scenarios are represented by organisms. When outcomes are uncertain and depend in part on the distribution and timing of responses, organisms must effectively make inferences about conditional probabilities.

One promising possibility is to model the subjective accumulation of evidence and subsequent prediction in terms of Bayesian updating (Vilares & Kording, 2011). The Bayesian approach is particularly attractive given the compositional formulation of outcome frequency and value, because these can be linked the Bayesian study of psychophysical functions (e.g. Sun et al., 2012). The proposal that subjective probability and value might be best understood in terms of psychophysical functions is not a new idea (Berlyne, 1970; Kahneman & Tversky, 1984), but the Bayesian hypothesis provides a much more sophisticated approach to the problem of how 'experienced value' might be updated as a function of experience. Provided the relative value of alternatives can be assessed in compositional terms, decision trees based on information criteria can be used to implement Bayesian updating both for gradual and sudden shifts in behavior (Gallistel et al., 2014).

This is not to say that organisms have objectively accurate machinery for computing posterior odds. There is substantial controversy over whether brains literally engage in Bayesian computations, or whether those models instead provide a better approximation than traditional approaches to cognition or association; dissociating between these two accounts is exceptionally difficult given our current tools for understanding neural circuits (Knill & Pouget, 2004).

1.4.4 The Scheduling of Outcomes

Another prominent feature of the animal literature is a focus on the behaviors that arise under different *schedules*. When an organism represents its expected outcome in a complex task, that representation must take into consideration the interaction between the responder and the environment. How organisms come to represent these interactions is of central importance to the topic of choice, because decisions are never made in the absence of some assumptions about the outcome.

Consider, for example, the 'gambler's fallacy.' If one gamble has a fixed 10% probability of yielding a reward, while another has a fixed probability of 20% of yielding an identical outcome, a 'rational' player should favor the 20% option exclusively, all else being equal. However, when many people are exposed to games of chance, they have a strong sense that the future odds depend in some way on past events, such that rare outcomes that have not appeared recently might feel as though they are 'due to appear any time now.' This subjective experience is obviously fallacious when applied to a die roll, which is necessarily 'memoryless' and thus independent of past rolls.

This same feeling is not inappropriate in many real-world contexts. For example, if someone waiting for a bus has learned from experience that busses typically arrive every ten minutes, they will begin to feel that the bus is 'due' as the ten-minute mark approaches. Whether this feeling is inappropriate or not depends entirely on the distribution of intervals between busses. In practice, this expectation may be called a fallacy *only* if the interval between busses is randomly drawn from an exponential distribution; otherwise, the expected wait time changes as a function of time elapsed. If the interval between busses is Gaussian, for example, then a bus genuinely *will* be due after a lot of time has passed. Consequently, representing the distribution of intervals between events is essential to making choices, and there may be mechanisms that, by default, treat uncertain circumstances as more closely resembling the bus scenario than a gambling scenario.

The distribution of expected intervals often depends not merely on the passage of time, but also on behavior. Consider, for example, the contrast between 'variable ratio' (VR) schedules and 'variable interval' (VI) schedules. Both schedules operate by setting some precondition on which a reward is contingent; in the VR case, an uncertain number of responses is required before the next response is rewarded, whereas in the VI case, an uncertain interval must elapse before a response is rewarded. In both cases, a response must be made to earn a reward. However, in the VI case, *a single response* is sufficient, provided it occurs after the interval has elapsed. Effectively, this means that the VI schedule has a memory (for time elapsed), whereas the VR schedule is memoryless. Consequently, although superficially similar, these two schedules have fundamentally different hazard functions: The gambler's fallacy remains a fallacy in the VR case, but is *not* a fallacy in the VI case. Indeed, unless a schedule is entirely memoryless, it ceases to be a fallacy altogether and becomes an effective strategy (Fiorina, 1971). Rather than a 'gambler's fallacy,' the feeling of being due for a reward is sometimes an 'actuary's wisdom.'

This contrast gives rise to very different optimal strategies. In the VR case, the best approach is to respond exclusively to the next alternative, whereas in the VI case, an effective approach is to treat Equation 1.11 as prescriptive and 'match' proportions of responses to proportions of rewards. Most organisms (including humans) do not assume by default that schedules are memoryless. Instead, extensive training on VR schedules is required before near-exclusive behavior is observed in humans (Vulkan, 2000) and other species (Jensen & Neuringer, 2008).

Schedules in which the probability of a reward are fixed for every response (e.g., typical VR schedules) are routinely described as being qualitatively different from those schedules whose odds of rewarding a response to a particular alternative grow over time (e.g., typical VI schedules). However, these two categories are merely the end points of a continuum. Just as intermediary schedules between these extremes may be engineered, functionally appropriate intermediary behaviors are exhibited by subjects under those conditions (for more detail, see the discussion of 'reinforcer hold' in Jensen & Neuringer, 2008). Baum & Aparicio (1999) also report an example in which two concurrent schedules display very different hazard functions. Given these results, it is reasonable that a general framework for predicting behavior can be constructed that applies across all scheduling paradigms whose uncertainty can be formally specified.

1.5 Human Decision Making

> It is immediately obvious to the most casual observer with the meanest intelligence that people do not think at the margin. Nobody goes to a grocery store and thinks, "I'm going to buy an orange. I'm going to buy another orange. I'm going to buy another orange." But if people don't think at the margin, and if, as Mankiw says, rational people *do* think at the margin, we are led to a most unhappy conclusion.
>
> –Yoram Bauman, presenting at AAAS, February 16, 2007. Adapted from Bauman (2003).

Although some of the early pioneers in human decision making, such as Luce and Herrnstein, had their work informed by the study of animal models, the vast majority of recent work on the topic has been pursued by economists and cognitive psychologists who are unfamiliar with the non-human literature. For example, because non-human subjects cannot be given verbal instructions, all schedules of reward must instead be learned through experience; human studies, on the other hand, have chiefly consisted of giving participants explicit instructions about the odds of various outcomes. Only recently have studies of human decision making struck upon the possibility that knowledge of reward schedules learned through feedback may differ from those learned through verbal instruction, yielding a 'description-experience gap' (Hertwig & Erev, 2009). Studies of this gap in humans routinely make no reference to the animal literature, and as such do not benefit from that literature's rich history of comparing different reward schedules. In general, normative accounts drawn from classical economics (particularly 'rational choice theory,' a prescriptive variation on expected utility) provide a poor description of actual decision making (Herrnstein, 1990), and the development of more accurate descriptive and predictive models has been an area of active development only since the 1970s.

1.5.1 Prospect Theory: A Break From Rationality

One of the major triumphs of cognitive theory over classical economics is the demonstration that participants violate even the 'subjectively rational' utility functions implied by Equation 1.4 when questions are framed in different ways. For example, framing otherwise mathematically identical lotteries in terms of gains or losses can reliably engender reversals in preference. The first adequate mathematical model of this behavior was achieved by Kahneman & Tversky (1979), who dubbed it 'prospect theory.'

The premise that gains and losses needed to be treated differently is not original to Kahneman & Tversky. This was a dominant theme in the 'theory of the firm' put forward by Shackle (1949, 1970), whose schematic representation of a concave utility function anticipates that of prospect theory (see also Shackle, 1958). However, it is instructive to consider why mainstream economics did not take the gains/losses contrast seriously:

> The exposition [of Shackle's theory] is greatly complicated by his insistence on differentiating between gains and losses. It is completely unclear to me what the meaning of the zero-point would be in a general theory; after all, costs are usually defined on an opportunity basis only.
>
> –Arrow (1951, p. 432)

Despite this mainstream skepticism, Shackle's justification for treating losses differently is entirely reasonable. Putting decisions into their proper context (where the consequences of current decisions impact the availability of future decisions), the long-term consequences of gains and losses are not symmetrical. A firm incurring a crippling loss, for example, enjoys fewer subsequent opportunities than one that enjoys a corresponding gain, and loss aversion can be seen under these conditions as a reasonabe strategy for increasing marginal profit over the long term.

Refinements were made to prospect theory over the following decade, with models shifting to a rank-dependent framework (discussed below). These efforts culminated in a general form described by Luce & Fishburn (1991), which was subsequently rebranded as 'cumulative prospect theory' by Tversky & Kahneman (1992). Strictly speaking, however, neither consists of a specific model of choice; instead, a placeholder function is specified for subjective magnitude, denoted by $\mathcal{V}(\cdot)$, and subjective probability, denoted by $\mathcal{W}(\cdot)$. Thus, a lottery (termed a 'prospect' by Tversky & Kahneman) is divided into its gains $A+$ and its losses $A-$, and the subsequent subjective expected utility is determined by their corresponding sum:

$$\begin{aligned} SWEU(L_A) &= SWEU(L_{A^+}) + SWEU(L_{A^-}) \\ &= \left(\sum_{i \in D} \mathcal{W}^+(\mathbf{p}_{A^+}) \cdot \mathcal{V}(\mathbf{M}_{A^+}) \right) + \left(\sum_{i \in D} \mathcal{W}^-(\mathbf{p}_{A^-}) \cdot \mathcal{V}(\mathbf{M}_{A^-}) \right) \end{aligned} \tag{1.13}$$

Here, $\mathbf{p}_{A^+}$ and $\mathbf{p}_{A^-}$ correspond to the probabilities associated with positive and negative outcomes, respectively, such that the probabilities associated with gains and losses may be transformed according to different parameters.

Prospect theory's standing in economics remains contentious in part because of disagreement about how $\mathcal{W}(\cdot)$ and $\mathcal{V}(\cdot)$ should be defined. According to models building on the foundations of classical economics, well-documented problems for expected utility theory can be explained by concavity in one or both of these functions (Starmer, 2000). The contrasting 'non-conventional' view put forward by prospect theory is that an abrupt discontinuity exists demarcating gains from losses, and that the position of this discontinuity can be influenced by framing effects.

One of the consequences of this segregation is that rather than representing the uncertainty of outcomes as subjective probabilities (which are subject to closure), they are instead represented as 'weights' that are not subject to closure. Curiously, Tversky & Kahneman acknowledge the efficacy of closure briefly:

> The weighting scheme used in the original version of prospect theory and in other models is a monotonic transformation of outcome probabilities. This scheme encounters two problems. First, it does not always satisfy stochastic dominance, an assumption that many theorists are reluctant to give up. Second, it is not readily extended to prospects with a large number of outcomes. These problems can be handled by assuming that transparently dominated prospects are eliminated in the editing phase, and by normalizing the weights so that they add to unity. Alternatively, both problems can be solved by the rank-dependent or cumulative functional, first proposed by Quiggin (1982) for decision under risk and by Schmeidler (1989) for decision under uncertainty. Instead of transforming each probability separately, this model transforms the entire cumulative distribution function. The present theory applies the cumulative functional separately to gains and to losses.
>
> –Tversky & Kahneman (1992, p. 299)

Here, “normalizing the weights so they add to unity” plainly denotes applying the closure constraint, as presumably can be attributed to Karmarkar (1978). Despite this admission, no further exploration of a compositional approach is considered, because rank-dependent theories are effectively ordinal. In a treatment of the problem by Luce (1988), whose rank-dependent function under uncertainty predates that of Schmeidler (see also Luce & Narens, 1985), the need to transition to a rank-dependent approach is motivated by the need to assemble complex and multi-stage lotteries from their component parts in a fashion that does not result in trivial intransitivity. These concerns can be successfully accommodated in the compositional framework because, as will be shown in Part I, the compositional sample space has a recursive structure that permits de- and recomposition without loss of information.

1.5.2 Prospect Theory, or Prospect Theme?

The formal flexibility of cumulative prospect theory is at once the key to its enduring popularity and its greatest theoretical shortcoming. Although Tversky & Kahneman (1992) argue strongly that framing has the power to dramatically impact choice, they propose no rigorous model for framing, which is presumed to unfold during their unspecified ‘editing’ phase. Furthermore, although they specify forms for $\mathcal{W}(\cdot)$ and $\mathcal{V}(\cdot)$, they do so in an ad-hoc manner based on the convenient features of those functions, and they make no commitment to any given form:

> The estimation of a complex choice model, such as cumulative prospect theory, is problematic. If the functions associated with the theory are not constrained, the number of estimated parameters for each subject is too large. To reduce this number, it is common to assume a parametric form (e.g., a power utility function), but this approach confounds the general test of the theory with that of the specific parametric form. For this reason, we focused here on the qualitative properties of the data rather than on parameter estimates and measures of fit.
>
> –Tversky & Kahneman (1992, p. 311)

Consequently, cumulative prospect theory is only a theory in a loose sense. Although it provides an explanation for many violations of rationality axioms, it is not sufficiently precise to make novel predictions that could differentiate it from other, similar proposals. Instead, cumulative prospect theory has been credited for redescriptions of the original empirical results using a variety of placeholder functions. This 'anything-goes' approach has resulted in an explosion of functional forms. In a review of published variations, Stott (2006) studied eight functions for $\mathcal{W}(\cdot)$ and eight for $\mathcal{V}(\cdot)$, as well as four 'decision functions' $\mathcal{D}(\cdot)$ (considered in the following section), which could be combined to form 256 distinct variations[9] of the model. Most of these forms (either in their original published form or in the resulting recombination) have little or no justification beyond either fitting the data or satisfying rationality axioms that are assumed *a priori*. Unsurprisingly, this functional menagerie collectively yielded substantially divergent predictions.

Additionally, most attempts to identify the 'functional form' for cumulative prospect theory make the assumption that all individuals should be described by the same parameters. Consequently, parameters for this zoo of functional forms are routinely fit to data pooled across groups of participants, without much interest in determining what the population range for the parameters might be. This assumption of parametric uniformity within participant pools is particularly strange given the applied focus of human decision making research: Presumably, if different groups of people make decisions differently, those differences would be of interest to scientists trying to shape and predict those decisions.

Despite these ambiguities, cumulative prospect theory does place a number of specific and testable constraints on the forms that the data can take. As previously noted, one of these is the preservation of stochastic dominance (for which ample evidence to the contrary is available). More interesting is the selection of whether $\mathcal{W}(\cdot)$, $\mathcal{V}(\cdot)$, or both should be parametrically discontinuous with respect to gains and losses. In principle, because $\mathcal{W}(\cdot)$ and $\mathcal{V}(\cdot)$ necessarily take on different functional forms (if only because probabilities and outcome magnitudes reside in sample spaces with different properties), the choice of which approach to take should yield divergent predictions. However, because the choice to make $\mathcal{W}(\cdot)$ discontinuous is largely a consequence of preserving stochastic dominance, identifying the optimal functional form on the basis of evidence remains an open question.

[9] Additionally, four other forms for $\mathcal{V}(\cdot)$ and three other forms for $\mathcal{W}(\cdot)$ are cited but not considered, increasing the range of *potential* models to 528.

1.5.3 Stochastic Dominance: When Axioms Attack

In a previously quoted passage, Tversky & Kahneman raise the issue of stochastic dominance, and this is a topic of sufficient importance to econometric models of decision-making that it merits special consideration. A core axiom of 'rational choice' theories of decision making is that the subjective utility (Equation 1.4) must rise monotonically as a function of the expected value (Equation 1.2) (Levy, 1992). That is, if $L_A > L_B$ then any rational model must also conclude that $\mathcal{L}_A > \mathcal{L}_B$. Whereas a specific case in which $A \succeq B$ may be interpreted as 'A dominates B,' the more general case in which $A \succeq B$ *regardless of the free parameters in the model* may be read as 'A stochastically dominates B.' To the degree that a model violates this assumption, it is fundamentally incompatible with a generalization of the von Neumann-Morgenstern axioms, and thus is not 'rational' even in a weak sense. Because of this, a sizable body of economic theorizing remains ideologically wedded to the axiom (e.g. Blavatskyy, 2011). In fact, the assumption is so strongly held that it provides the basis for formal statistical testing in applied contexts (e.g. in the definition and measurement of poverty by Anderson, 1996).

Despite this, there is considerable experimental evidence that participants routinely violate stochastic dominance, often to a considerable degree (Birnbaum & Navarrette, 1998; Birnbaum et al., 1999; Charness et al., 2007). The most obvious manifestation of this violation is that risky choices engender not only different behaviors than riskless ones, but also different preference orderings, collectively called 'gambling effects' (Deicidue et al., 2004). Another is the 'endowment effect' whereby ownership of an object distorts the subjective utility function, thereby reversing which of two items is valued more greatly than the other (Knetsch, 1989). These patterns of violation are sufficiently consistent that they cannot be attributed to 'errors arising from stochastic decision rules' (Birnbaum, 2008). Given that stochastic dominance does not appear to be a consistent feature of human decision making, it is unclear why it should be a constraint on psychological models of risky choice, beyond some assumption of 'rationality' made prior to an examination of the evidence.

In order to reconcile the reactionary opposition to models that violate stochastic dominance with the evidence that such violations occur routinely, it is important to draw a distinction between the kinds of *phenomenological* violations that participants routinely make, and those *theoretical* violations that only arise mechanically from deeper assumptions in the model. In practice, these

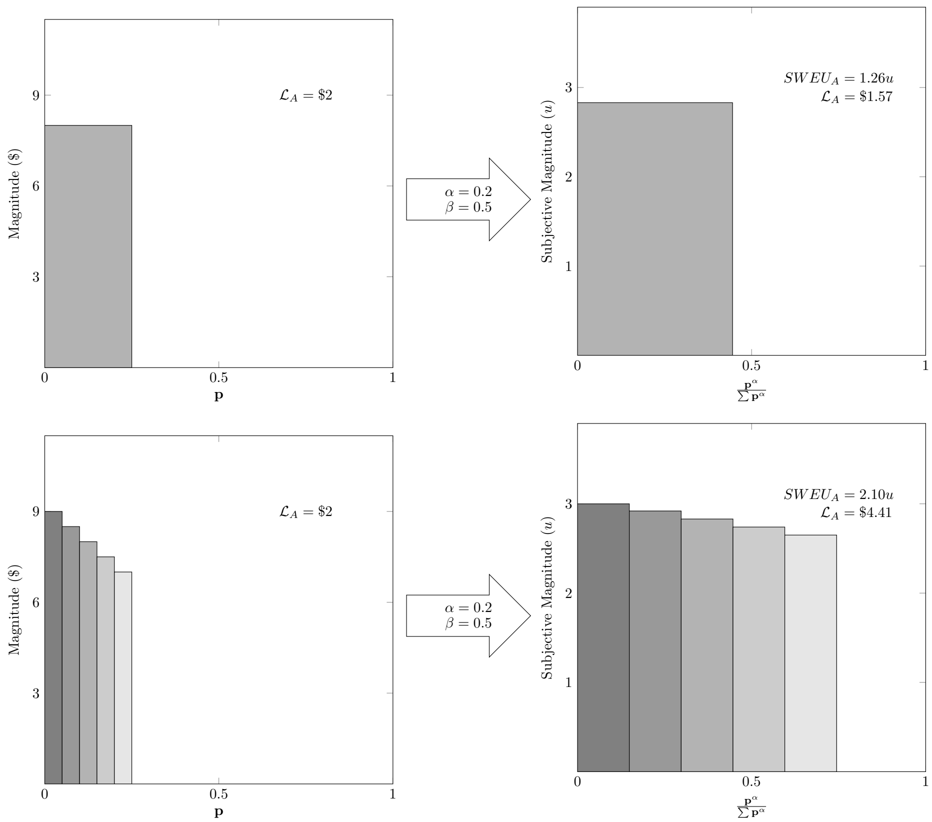

Figure 1.3: Two lotteries, each of which have an identical expected value, are depicted on the left. When a single 0.25 probability of \$8 (top) is compared to five 0.05 probabilities averaging \$8 (bottom), applying Equation 1.4 with the parameters $\{\alpha = 0.2, \beta = 0.5\}$ results in wildly different expected values. Given a sufficiently low value for α, subdividing a low-probability event will consistently swamp the much more likely outcome of receiving \$0.

theoretical violations of stochastic dominance can be severe, and a viable model should mitigate their effects in such a way that does not preclude the milder violations for which there is empirical evidence.

Crucially, participants must somehow represent the lotteries, and these representations are unlikely to be entirely discrete. For example, consider the three lotteries in Figure 1.3. In the first two, each of the outcomes is defined in entirely discrete terms. Consequently, when the probabilities are distorted according to Equation 1.4 (given an extreme α of 0.2), the six-outcome lottery (bottom) becomes dominated by the non-zero payoffs, leading to a much larger expectation than in the single-outcome case. These two cases assume, however, that the participant is representing each of the outcomes with no ambiguity, as well as treating them as qualitatively distinct (rather than displaying a degree of exchangeability).

Because individuals are not routinely cheated out of vast sums of money by grifters who present them with subdivided lotteries, it is reasonable to conclude that the behavior predicted by Figure 1.3 is a poor model. However, the unreasonableness of this prediction arises from its assumptions of exact precision and nominal segregation of outcomes, *not* from its violation of stochastic dominance. Participants who do not use mental arithmetic for a living (as well as non-human subjects) routinely violate stochastic dominance without engaging in pathological financial behavior. A much more reasonable approach to the problem, then, is to try to model the fuzziness of mental representation, rather than insisting that the absence of pathological behavior on behalf of participants somehow proves the face-validity of the von Neumann-Morgenstern axiom.

In the end, the need for a more psychologically valid approach to decision making is best demonstrated by the behavior of economists themselves. In an encounter described by Weber & Camerer (1987), economists Jimmie Savage and Maurice Allais (of the 'Allais paradox' described above) argued over lunch about whether a theory of economics could be a valid reflection of the real world if it required assuming that economic agents were strictly rational. Savage pressed Allais with a number of hypothetical lotteries, and showed in short order that, rather than being rational, Allais was himself a 'paradoxical' individual who violated the prescriptions laid down by the axioms. If ostensibly irrational preference reversals can be expected of even the mathematical sophisticates at the apex of economic thought, it follows that a better model of human decision making awaits.

Part I

Analytic Methodology

Chapter 2

Introduction to Compositions

Compositions lie at the heart of every aspect of this dissertation. Consequently, an extensive overview of their basic properties and formal applications is provided in the following four chapters. Some material is presented redundantly with the brief summary provided in the introduction in order to ensure that readers are given a complete treatment of the main topics.

In Chapter 3, the *simplex* is described as the fundamental sample space for compositional data. Naïve treatment of compositional data is routinely suboptimal because of a failure to understand this "simplicial" geometry, which is both bounded and non-Euclidean in nature. Chapter 3 also describes operators in the simplex that mirror arithmetic operations used in Euclidean vector space.

In Chapter 4, the geometry of the simplex is linked with Euclidean vector space using transformations that overcome troublesome properties of simplicial space. These transformations are lossless and reversible, allowing data to migrate from one sample space to another at will.

Chapter 5 applies the techniques described in the preceding two chapters to specify statistical models with practical applications, while Chapter 6 addresses how to deal with problematic data.

These chapters introduce an extensive array of specialized symbols and operators. Additionally, they define a wide range of variables whose interpretations are used consistently in subsequent sections. A glossary of all notation used in the dissertation is provided in Appendix A.

Although these chapters are concerned with compositions as abstract objects, their abstract properties both inform and constrain theoretical models. Consequently, attention is periodically drawn to topics with empirical relevance. These asides are not intended to be exhaustive, and their implications are examined in greater detail in Part III.

Chapter 3

Basic Operations and Geometry of Compositional Space

Compositional analysis, first systematically described by Aitchison (1986), considers datasets in which observations are represented as ratios of other observations, without regard for their absolute values. In other words, a compositional analysis is an analysis of *relative proportions*, as opposed to an analysis of absolute values. Correspondingly, any data in which observations must sum to some fixed total (1.0 for proportions, 100.0 for percentages, and so forth) can be considered compositional. Data properly represented on a ratio scale may also be interpreted as having compositional properties, under certain circumstances.

Compositional data can be found in every scientific discipline. Examples include:

- **Astrophysics**: Proportions of regular and dark matter in galaxies.
- **Biology**: Allele frequency in gene pools.
- **Economics**: Distributions of discretionary consumer purchasing.
- **Geology**: Compositions of geological strata.
- **Neuroscience**: Receptor over- and under-expression as a function of psychopathology.
- **Political Science**: Electoral voting patterns.
- **Psychology**: Relative psychophysical intensity of stimuli.

- **Sociology**: Population shifts in age, gender, and other demographic variables.

Alongside this wide range of measures, distributions of choice (long analyzed in relative terms) are also prime candidates for compositional analysis.

Despite this ubiquity, compositional data are challenging to analyze using traditional statistics. Compositional data reside in an unusual sample space governed by non-standard mathematical operators. When analyzed with respect to the wrong sample space, the data appear to violate the assumption of statistical independence, because the value of every variable is scaled relative to the value of all the others. Below, the basic properties of this sample space are outlined.

3.1 Compositions and Closure

Suppose a vector $\mathbf{x}$ consists of D elements, all of which have positive values. If an analysis is concerned with only the *relative* values of these elements, as opposed to the absolute values, the values of each element in the vector can be rescaled relative to a constant $\eth$, using an operation called *closure*, denoted by $\mathcal{C}(\cdot)$:

$$\mathcal{C}(\mathbf{x}) = \left(\frac{\eth \cdot x_1}{\sum_{i=1}^{D} x_i}, \ldots, \frac{\eth \cdot x_D}{\sum_{i=1}^{D} x_i} \right) \tag{1.1}$$

Each component x_i of $\mathbf{x}$ is divided by the sum of the vector, then multiplied by $\eth$. Values for $\eth$ are arbitrary, and common values include 1.0 (for proportions) and 100.0 (for percentages); all subsequent discussion will assume that $\eth = 1$.

Any vector that has been scaled in this way is a *composition*, which by definition has a fixed sum, and as such conveys only the relative information about the *components* that make it up:

$$\mathbf{x} = (x_1, \ldots, x_D) \Big| x_i > 0, \sum_{i=1}^{D} x_i = \eth > 0 \tag{3.1}$$

In addition to being the operation used to convert a vector to a composition, closure should be understood as a constraint on the sample space of the composition. As such, the values of the components can approach either zero or $\eth$, but may never move beyond them.

Because compositions must always be rescaled to the closure constant $\eth$, simple vector concatenation causes the absolute values of all components to be rescaled. For example:

$$\mathbf{z} = \mathcal{C}\left(\eth_x \cdot \mathcal{C}(\mathbf{x}), \eth_y \cdot \mathcal{C}(\mathbf{y})\right) \tag{3.2}$$

Here, a composition **z** is assembled from the vectors **x** (of length i) and **y** (of length j). Closure is applied to both vectors, which are scaled by factors of $\eth_x$ and $\eth_y$. Because a composition encodes only *relative* information, this concatenation results in no loss of compositional information. For example, z_1 and x_1 both correspond to the first component of **x**, and z_{i+1} and y_1 both correspond to the first item in **y**, despite the fact that, in absolute terms, $z_1 \neq x_1$ and $z_{i+1} \neq y_1$. Since only the relative values of each item to every other item are encoded, ratios of values such as $\frac{x_1}{x_2} = \frac{z_1}{z_2}$ and $\frac{y_1}{y_2} = \frac{z_{i+1}}{z_{i+2}}$ are preserved.

Subdivision of a composition (that is, *decomposition*) may also be performed within the constraints of closure. Since $\mathbf{x} = \mathcal{C}(z_1, \ldots, z_i)$, the original compositions that went into the concatenation in Equation 3.2 can easily be restored. Thus, **x** is considered a *subcomposition* of **z**, defined as a subset of the components of **z** that contain the compositional information about **x**.

The closure constraint renders the standard operation of arithmetic addition inconsistent. For example, if **x**, **y**, and **z** are compositions (and thus subject to closure), then $\mathcal{C}(\mathbf{z} + \mathcal{C}(\mathbf{x} + \mathbf{y})) \neq \mathcal{C}(\mathbf{x} + \mathcal{C}(\mathbf{y} + \mathbf{z}))$ unless $\mathbf{x} = \mathbf{z}$. Even more dramatically, scalar multiplication ceases to have any meaning, as $\mathcal{C}(\alpha \cdot \mathbf{x}) = \mathbf{x}$ for all positive values of α, and is undefined if $\alpha \leq 0$. This undermines the interpretability of basic descriptive statistics like the arithmetic mean, and thus all inferences built upon the properties of those statistics.

Untransformed compositional data also give rise to illusory statistical relationships. Consider a dataset in which three variables are produced by a random number generator, but are recorded as *relative proportions* rather than as absolute values. A correlation matrix calculated from untransformed proportions will show negative correlation among the items (because gains in one proportion must be balanced by losses in the others), despite there being no true correlation in their origin.

All of the problems described above arise from mistakenly treating a composition as a Euclidean vector. Compositions do reside in a vector space, but one with a different geometry. Building up from the fundamentals of that geometry permits a direct connection to be drawn between compositional operations and their Euclidean counterparts.

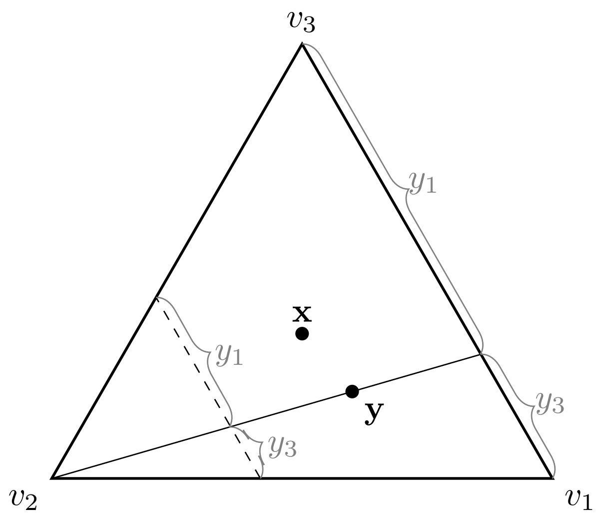

Figure 3.1: A ternary plot depicting the coordinates of the simplicial geometry $\mathcal{S}^3$ as a two-dimensional triangle. Here, the composition $\mathbf{x}$ is located at $\mathcal{C}(1,1,1)$, which is the *barycenter* of the simplex. The composition $\mathbf{y}$ is located at $(0.5, 0.3, 0.2)$, assuming the closure constant $\eth = 1$. The line connecting vertex v_2 with $\mathbf{y}$ shows that, for any point lying on the line, the ratio $\frac{y_1}{y_3}$ remains unchanged.

3.2 The Simplex

In order to develop intuitions about compositional data, it is essential to understand that the geometry of the compositional sample space is *simplicial*; that is, it resides within a simplex. A simplex is a d-dimensional polytope in which D vertices are all equidistance from one another, where $d = D - 1$. Hereafter, the internal geometry of a simplex defined by D vertices will be denoted by $\mathcal{S}^D$. Any simplex $\mathcal{S}^D$ will have D *facets*, each of which will be a simplex with geometry $\mathcal{S}^d$. A 1-simplex with geometry $\mathcal{S}^2$ is called an *edge*.

The triangle depicted in Figure 3.1 is a *ternary plot* where each vertex represents one of three alternatives, v_1, v_2, and v_3. If a composition $\mathbf{x} = \mathcal{C}(x_1, x_2, x_3)$, then every possible vector $\mathbf{x}$ must correspond uniquely to a coordinate within this ternary plot, that is, $\mathbf{x} \in \mathcal{S}^3$. In Figure 3.1, $x_1 = x_2 = x_3 = \frac{1}{3}$, positioning $\mathbf{x}$ at the barycenter of the triangle, whereas $\mathbf{y}$ lies at an offset position at $(0.5, 0.3, 0.2)$.

Each edge in Figure 3.1 is itself a simplex with geometry $\mathcal{S}^2$, containing only the information relating the two connected vertices. Consider the line passing through the vertex v_2 and $\mathbf{y}$. The

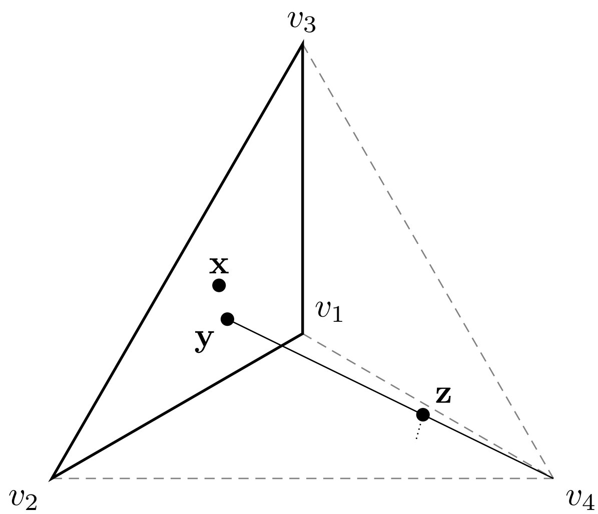

Figure 3.2: The geometry $\mathcal{S}^4$ plotted as a tetrahedron. The facet (v_1, v_2, v_3), denoted with solid edge lines, is identical to the simplex in Figure 3.1, including the positions of the compositions $\mathbf{x}$ and $\mathbf{y}$. The composition $\mathbf{z}$ is located at (0.2, 0.12, 0.08, 0.6), and intersects the line connecting $\mathbf{y}$ and v_4.

point where this line intersects the edge (v_1, v_3) divides that edge into two segments, whose relative lengths are specified by $\mathcal{C}\,(y_1, y_3)$. So long as $\mathbf{y}$ lies on any point along this line, then $\frac{y_1}{y_3} = \frac{0.5}{0.2}$, regardless of the value of y_2.

This logic extends arbitrarily to higher dimensions. Figure 3.2 depicts a $\mathcal{S}^4$ simplex (i.e. a tetrahedron) that introduces a fourth alternative, the vertex v_4. Implied by this $\mathcal{S}^4$ geometry are four facets, each of which is a simplex with geometry $\mathcal{S}^3$. In this case, the facet (v_1, v_2, v_3) in Figure 3.2 is identical to the simplex in Figure 3.1 (including the positions of $\mathbf{x}$ and $\mathbf{y}$). Further, if $\mathbf{z} = \mathcal{C}\,(0.4 \cdot \mathbf{y}, 0.6)$, then the position of $\mathbf{z}$ must lie along the line projected from the vertex v_4 to the opposing facet (v_1, v_2, v_3), and the point $\mathbf{z}$ will subdivide this line into segments whose lengths conform to the ratio $\frac{0.4}{0.6}$. Because of this, $\mathcal{C}\,(z_i, z_4) = (0.4, 0.6)$ for all $(i \neq 4)$.

3.2.1 Perturbation and Powering

Because of the closure constraint, the arithmetic operations of addition and multiplication are not well-behaved in simplicial geometry. Instead, an operation called *perturbation* is analogous to addition, while *powering* is analogous to item-wise multiplication. These operations, introduced by Aitchison (1986) and described at length by Egozcue et al. (2011), are the basic operators of simplicial geometry.

Perturbation, denoted by the $\oplus$ operator, consists of item-wise multiplication between two compositions, followed by closure:

$$\mathbf{x} \oplus \mathbf{y} = \mathcal{C}\left(x_1 \cdot y_1, x_2 \cdot y_2, \ldots, x_D \cdot y_D\right) \tag{3.3}$$

Powering, denoted by the $\odot$ operator, raises all components of a composition by an exponent, followed by closure:

$$\alpha \odot \mathbf{x} = \mathcal{C}\left(x_1^{\alpha}, x_2^{\alpha}, \ldots, x_D^{\alpha}\right) \tag{3.4}$$

In the interest of notational efficiency, *anti-perturbation* is denoted with the $\ominus$ operation, which reverses perturbation:

$$\mathbf{x} \ominus \mathbf{y} = \mathbf{x} \oplus \left((-1) \odot \mathbf{y}\right) \tag{3.5}$$

Perturbation and powering allow the simplex to be navigated as a vector space because these operations have the appropriate properties to do so:

1. (associative) $\mathbf{x} \oplus (\mathbf{y} \oplus \mathbf{z}) = (\mathbf{x} \oplus \mathbf{y}) \oplus \mathbf{z}$

2. (commutative) $\mathbf{x} \oplus \mathbf{y} = \mathbf{y} \oplus \mathbf{x}$

3. (distributive) $(\alpha \odot \mathbf{x}) \oplus (\alpha \odot \mathbf{y}) = \alpha \odot (\mathbf{x} \oplus \mathbf{y})$

The full properties of the operators are described by Egozcue et al. (2011).

3.2.2 Basic Metrics

In addition to basic geometric operations, an analysis of the simplex requires some basic metrics. These metrics, also introduced by Aitchison (1986), describe analogs to the inner product and to distance. In his honor, these are called the *Aitchison inner product* and *Aitchison distance*, respectively, and are indicated with a subscript a.

The Aitchison inner product, denoted by $\langle \cdot, \cdot \rangle_a$ is specified in terms of the Euclidean properties of the CLR transformation.

$$\begin{aligned} \langle \mathbf{x}, \mathbf{y} \rangle_a &= \sum_{i=1}^{D} \log \frac{x_i}{\mathfrak{g}\left(\mathbf{x}\right)} \cdot \log \frac{y_i}{\mathfrak{g}\left(\mathbf{y}\right)} \\ &= \langle \mathrm{clr}\left(\mathbf{x}\right), \mathrm{clr}\left(\mathbf{y}\right) \rangle = \langle \mathbf{x}^{\circ}, \mathbf{y}^{\circ} \rangle \end{aligned} \tag{3.6}$$

The Aitchison inner product is equal to the Euclidean inner product $\langle\cdot,\cdot\rangle$ of CLR-transformed compositions. This leads naturally to defining the Aitchison norm as follows:

$$\|\mathbf{x}\|_a = \langle \mathbf{x}, \mathbf{x} \rangle_a \tag{3.7}$$

The subscript a distinguishes this from the Euclidean norm $\|\cdot\|$.

Aitchison distance, denoted by $\mathfrak{d}_a(\cdot,\cdot)$, arises naturally from the anti-perturbation operator specified in Equation 3.5.

$$\begin{aligned} \mathfrak{d}_a(\mathbf{x},\mathbf{y}) &= \|\mathbf{x} \ominus \mathbf{y}\|_a \\ &= \mathfrak{d}(\operatorname{clr}(\mathbf{x}), \operatorname{clr}(\mathbf{y})) = \mathfrak{d}(\mathbf{x}^\circ, \mathbf{y}^\circ) \end{aligned} \tag{3.8}$$

Note that the Aitchison distance between compositions is equal to the Euclidean distance $\mathfrak{d}(\cdot,\cdot)$ between the CLR transformations of those compositions.

The Aitchison inner product (Equation 3.6) and the perturbation operation (Equation 3.3) are sufficient to interpret the geometry of the simplex as being an internally consistent Hilbert space whose properties are isomorphic to those of a Euclidean space consisting of real (i.e. non-compositional) coordinates (Aitchison, 1986).

3.2.3 Simplicial Lines and Orthogonality

A *simplicial line* (also called a compositional line) is defined by the following equation:

$$\mathbf{y} = \boldsymbol{\kappa} \oplus (\alpha \odot \mathbf{x}) \tag{3.9}$$

Here, the composition $\boldsymbol{\kappa}$ acts as the intercept, while the parameter α plays the role of the slope. Four compositional lines are depicted in Figure 3.3.

Compositional lines are, strictly speaking, geodesics confined to the geometry of the simplex. For practical purposes, they behave exactly like straight lines, albeit in terms of simplicial geometry. These lines may be parallel or orthogonal with respect to the Aitchison distance (Equation 3.8) and the Aitchison inner product (Equation 3.6), just as lines plotted in real geometries are considered parallel or orthogonal with respect to Euclidean distance and inner products.

On occasion, there is need to perform a *perturbation-linear combination*, which is denoted by

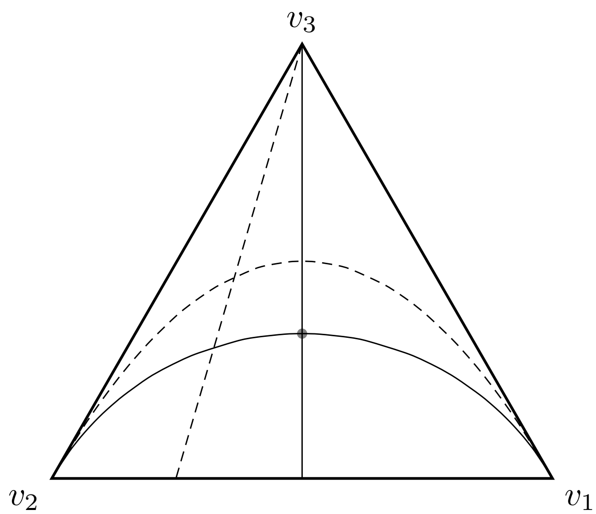

Figure 3.3: A ternary plot of the geometry $\mathcal{S}^3$, with the barycenter indicated by a gray point. Four compositional lines are depicted. The two solid lines in this figure are compositionally orthogonal from one another, as are the two dashed lines, as determined using the Aitchison inner product (Equation 3.6). Additionally, the two lines originating at v_3 are considered compositionally parallel, as are the two lines connecting v_1 and v_2.

the $\bigoplus$ operator:

$$\begin{aligned}\mathbf{y} &= \bigoplus_{j=1}^{D} (\alpha_j \odot \mathbf{x}_j) \\ &= (\alpha_1 \odot \mathbf{x}_1) \oplus (\alpha_2 \odot \mathbf{x}_2) \oplus \\ &\quad \cdots \oplus (\alpha_D \odot \mathbf{x}_D)\end{aligned} \tag{3.10}$$

This operator is a compositional equivalent of linear combination as it traditionally is performed in a real geometry (Egozcue & Pawlowsky-Glahn, 2006).

Chapter 4

Transformation Between Compositional and Real Geometries

Although the sample space within which compositions reside is a fully consistent vector space with its own fundamental operations and measures, it is beyond the means or interest of most analysts to rebuild the tool of statistics using the these operations. As a vector space, it necessarily follows that compositions within the simplex can be 'made Euclidean' by transformation. Three such transformations are described below. Once an analyst becomes comfortable moving from one transformation to another, standard statistical tools may be used to interrogate the data with all the depth and sophistication normally associated with scalar data.

It may not be clear to the reader why understanding three different transformations are necessary. After all, if a set of compositions may be transformed with no loss of information from one vector space to another, what difference does it make which coordinates are used? The answer is that although the *coordinates* translate from one condition to the next without a loss of information, each transformation carries with it different assumptions about the *uncertainty* of these measurements. Thus, although a purely descriptive comparison of the data in each sample space is isomorphic, the appropriate calculation of test statistics and confidence intervals may not be possible in all cases. In each case, the limitations of each transformation is described in explicit terms. Of the three, the ILR transformation (p. 43) permits the most sophisticated statistics, limited only insofar as it is more abstruse than the others, and thus somewhat more difficult to discuss.

4.1 Compositions and Log Odds

Log-transformation is central in all of the operations described below. Because perturbation is a form of multiplication, and powering a form of exponentiation, log transformation converts these operations to ordinary addition and multiplication, making the linear analysis of the transformed data straightforward.

All of the transformations below can be seen as variations of *log odds* transformations. However, while traditional log-odds analyses compare only two values (typically $\frac{p}{1-p}$ given some probability p), compositional transformations are general to any number of alternatives. Consequently, in an important way, all transformations below provide ways to generalize the idea of a log-odds analysis to more than two alternatives. That there are several different transformations should be taken as a clear sign that, when more than two alternatives are considered, an analyst must display some discernment in selecting a transformation that serves the specific statistical questions they wish to investigate.

4.2 The Additive Log-Ratio Transformation

One way to convert compositional data into more familiar terms is to use the *additive log-ratio* (or ALR) transformation, identified by Aitchison (1986) and denoted by $\text{alr}(\cdot)$:

$$\begin{aligned} \text{alr}(\mathbf{x}) &= \mathbf{x}^{@} \\ &= \log\left(\frac{x_1}{x_D}, \frac{x_2}{x_D}, \ldots, \frac{x_d}{x_D}\right) \end{aligned} \tag{4.1}$$

Here, a vector $\mathbf{x}$, consisting of D elements, is transformed into a vector consisting of length[1] $d = (D-1)$, summarized using the shorthand $\mathbf{x}^{@}$. The ALR transformation reveals that, when considering ratios of values, one fewer degree of freedom is needed to describe the data. Put another way, the ALR transformation permits information about the D items contained in $\mathbf{x}$ to be encoded (with no loss of information) in the real geometry $\mathbb{R}^d$.

[1] We will assume $d = (D-1)$ in all subsequent calculations, borrowing this notation from Aitchison (1986).

Although the reduction in dimensionality brought about by the ALR transform is compelling, its use is fraught with difficulties because the resulting geometry is asymmetrical. The best-fitting parameters depend on the "reference category" x_D, which is given a privileged status in the denominator of Equation 4.1. The selection of this denominator is arbitrary in principle (Kwak & Clayton-Matthews, 2002), and in practice it influences obtained parameter estimates. The magnitude of these distortions becomes increasingly problematic as the dimensionality of the data grows (N. Changizi & Hamarneh, 2010). Despite decades of work considering the implications of this transformation, no standard rule for selecting a reference category has been agreed upon (Egozcue et al., 2012). Even under perfect circumstances, comparing parameters obtained across different studies is complicated by the ad-hoc selection of reference categories.

4.2.1 The Generalized Matching Law

Recasting the generalized matching law (Equation 1.8) in compositional terms using the ALR transformation (Equation 4.1) gives the following:

$$\begin{aligned} \operatorname{alr}(\mathbf{y}) &= \operatorname{alr}(\boldsymbol{\kappa}) + \alpha \cdot \operatorname{alr}(\mathbf{x}) \\ \mathbf{y}^{@} &= \boldsymbol{\kappa}^{@} + \alpha \cdot \mathbf{x}^{@} \end{aligned} \tag{4.2}$$

Although this notation has not been previously used in descriptions of generalized matching, it is nevertheless arithmetically identical to how matching analyses have almost always been performed (since, after all, $\operatorname{alr}(\mathbf{y}) = \log\left(\frac{y_1}{y_2}\right)$ in the two-alternative case). In addition to fully capturing the computation described in Equation 1.8, there is no requirement that Equation 4.2 be limited to compositions with two components: Given D different component behaviors, the vectors $\mathbf{x}^{@}$, $\mathbf{y}^{@}$, and $\boldsymbol{\kappa}^{@}$ will each consist of d log-ratios.

As noted above, although this form of matching is intuitively straightforward, it is statistically problematic for all cases besides the two-alternative case. It is presented here mainly to show the connection between the traditional two-alternative approach and the broader compositional paradigm.

4.3 The Centered Log-Ratio Transformation

Aitchison (1986) also identified the *centered log-ratio* (CLR) transformation, denoted by $\operatorname{clr}(\cdot)$:

$$\begin{aligned}\operatorname{clr}(\mathbf{x}) &= \mathbf{x}^{\circ} \\ &= \log\left(\frac{x_1}{\mathfrak{g}(\mathbf{x})}, \frac{x_2}{\mathfrak{g}(\mathbf{x})}, \ldots, \frac{x_D}{\mathfrak{g}(\mathbf{x})}\right)\end{aligned} \tag{4.3}$$

Here, $\mathfrak{g}(\mathbf{x})$ denotes the geometric mean of $\mathbf{x}$. Unlike $\operatorname{alr}(\cdot)$, the vector resulting from $\operatorname{clr}(\cdot)$ is symmetrical. However, both $\mathbf{x}$ and $\mathbf{x}^{\circ}$ have D elements.

The CLR transformation may be reversed using the closure operation:

$$\mathbf{x} = \mathcal{C}(\exp(\mathbf{x}^{\circ})) \tag{4.4}$$

Because the data in the composition $\mathbf{x}$ have d degrees of freedom, but the CLR transformation embeds the data in a geometry with D dimensions, the covariance matrix associated with $\mathbf{x}^{\circ}$ is singular, a substantial statistical shortcoming. This constraint is hereafter referred to as *equilibrium* and corresponds to the following relationship:

$$\text{For any composition } \mathbf{x}, \sum \mathbf{x}^{\circ} = 0$$

Consequently, although $\mathbf{x}^{\circ}$ is a vector whose coordinates may be any real number, their sum is constrained.

Like closure, equilibrium limits the permitted range of values that may appear within the vector. Conceptually, equilibrium forces CLR-transformed data to lie along a diagonal cross-section of the real geometry $\mathbb{R}^D$, such that the true sample space belongs to a subspace with the geometry $\mathbb{R}^d$. For example $\mathbf{x}^{\circ} = (1, -1)$ is a legal vector in CLR-transformed space, and corresponds approximately to $\mathbf{x} = (0.88, 0.12)$. However, the vector $\mathbf{x}^{\circ} = (1, 1)$ is *not* a legal vector in CLR transformed space, because it does not fall on the plane specified by equilibrium. This plane is responsible for the singular covariance matrix identified by Aitchison.

Although equilibrium and closure are both constraints on the legal values that data may take, they arise independently from one another. Equilibrium is a consequence of using the geometric mean $\mathfrak{g}(\cdot)$ to normalize the data in the CLR transformation, and it is expected to arise whether or not the transformed data were originally subject to closure.

4.4 The Isometric Log-Ratio Transformation

> This point comes up quite often vector analysis. The fundamental proofs of things require the elegant equations in general and make nice elegant proofs, but in making various calculations and analyses, it's always a good idea, and there's no harm in it, in taking axes in some convenient way... The point of writing the vector equations is to demonstrate in the beginning that the equations are independent of the coordinate system. And that *means* you're allowed to choose any coordinate system you want! The answer will be the same. So why bother with some complicated one where everything is at some complicated angle when you can choose a neat one for the particular problem. Provided that the equations are vector equations, that *means* they're independent of the coordinate system, and that *means* that you can choose any coordinate system you want. So by all means take advantage of it.
>
> –Richard Feynman, the Feynman Lectures on Physics, the California Institute of Technology, October 15, 1962.

Most published analyses of compositional data implicitly use either the ALR transformation or the CLR transformation without fully considering the impact doing so has on the sample space. Given the choice between these two admittedly imperfect transformations, Aitchison (1986) favored the ALR transformation because an asymmetrical sample space was less problematic for inferential statistics than a singular covariance matrix. More recently, however, Egozcue et al. (2003) developed a compromise between these two transformations called the *isometric log-ratio transformation*, denoted by $\mathrm{ilr}(\cdot)$. This transformation combines the reduced dimensionality of the ALR transformation with the symmetry of the CLR transformation. This approach not only puts many existing analyses on a more firm statistical footing, but it also creates a foundation for developing new models that allow easy resolution of many previously thorny problems.

The ILR transformation is achieved by applying a matrix rotation to data already transformed using the CLR transformation (Equations 4.3). Since the CLR transformation results in a symmetrical geometry $\mathbb{R}^d$ embedded at an angle in a higher space $\mathbb{R}^D$, rotating the data by any orthonormal basis $\mathbf{U}$ achieves the ILR transformation, which is denoted by $\mathbf{x}^*$:

$$\begin{aligned}\mathbf{x}^* &= [\mathbf{U}\,(\mathbf{x}^\circ)^\intercal]^\intercal \\ \mathbf{x}^\circ &= (\mathbf{x}^*)\,\mathbf{U}\end{aligned} \tag{4.5}$$

Here, $\intercal$ signifies matrix transposition. That is, the orthonormal basis $\mathbf{U}$ may be matrix multiplied with $\mathrm{ilr}\,(\mathbf{x})$ to produce $\mathrm{clr}\,(\mathbf{x})$, and the process may also be easily reversed. These operations allow lossless translation between the CLR and ILR transformations, so long as the orthonomal basis $\mathbf{U}$ is consistent.

Despite being a simple modification, this matrix rotation entirely resolves the tension between the ALR and CLR transformations. When compositional data are ILR-transformed, the explanatory factors being analyzed are symmetrical, and genuinely orthogonal effects may be identified as such through an analysis of covariance. The challenge is to approach selecting the orthonormal basis carefully. Doing so correctly has the potential to yield very powerful inferences.

4.4.1 The Orthonormal Basis

In standard parlance, an orthonormal basis $\mathbf{U}$ is a square matrix of size D in which each row $\mathbf{u_i}$ is a vector whose norm equals 1 (that is, $||\mathbf{u_i}|| = 1$ for all i), and in which the inner product of every pair of rows is zero (that is, $\langle \mathbf{u_i}, \mathbf{u_j} \rangle = 0, i \neq j$). As it relates to the ILR transformation, however, the term is used slightly differently, because we know in advance that any data $\mathbf{x}^\circ$ necessarily has one fewer dimensions than the space in which it is embedded, and thus has a singular covariance matrix. Consequently, $\mathbf{U}$ is hereafter defined as being a matrix of size $(d \times D)$, whose purpose is to remove the extraneous dimension of $\mathbf{x}^\circ$ from consideration:

$$\mathbf{U} = \begin{bmatrix} \mathbf{u_1} \\ \vdots \\ \mathbf{u_d} \end{bmatrix} = \begin{bmatrix} u_{1,1} & \cdots & u_{1,d} & u_{1,D} \\ \vdots & \ddots & \vdots & \vdots \\ u_{1,1} & \cdots & u_{1,d} & u_{1,D} \end{bmatrix} \quad \begin{aligned} ||\mathbf{u_1}|| = \cdots = ||\mathbf{u_d}|| = 1 \\ \langle \mathbf{u_i}, \mathbf{u_j} \rangle = 0, i \neq j \end{aligned} \tag{4.6}$$

Any orthonormal basis that conforms to this description is a viable candidate for the ILR transformation (Egozcue & Pawlowsky-Glahn, 2006). However, as we will see, different orthonormal bases can be used to ask different questions about the data. No single orthonormal basis is the 'canonically correct' basis, as the data are symmetrical and will tolerate any rotation that does not introduce distortion.

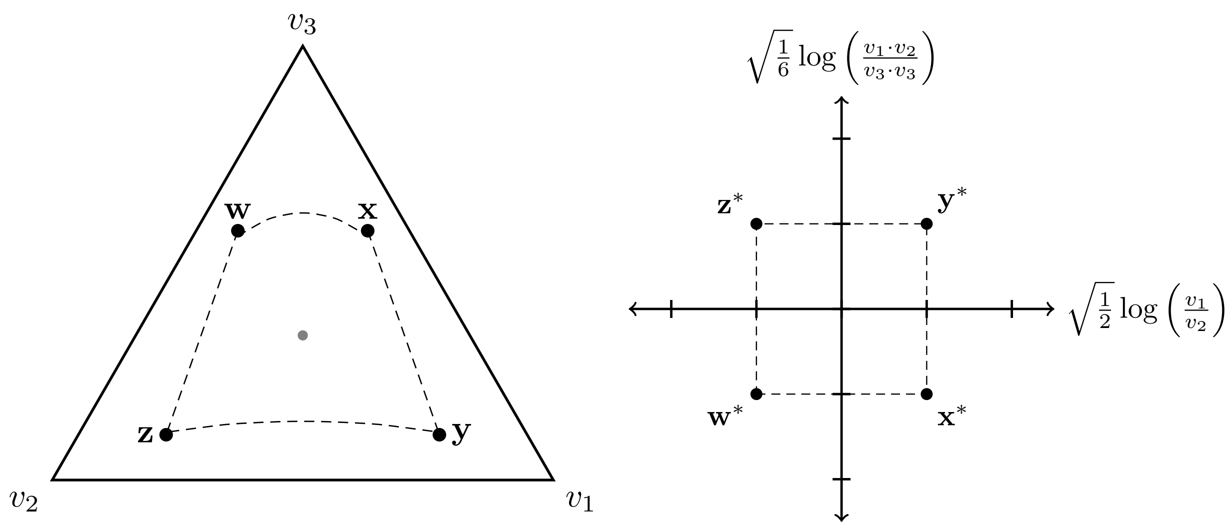

Figure 4.1: The geometry $\mathcal{S}^3$ on the left is ilr-transformed to the geometry $\mathbb{R}^2$ on the right. The points $\mathbf{w}^*$, $\mathbf{x}^*$, $\mathbf{y}^*$, and $\mathbf{z}^*$ lie at (-1,-1), (1,-1), (1,1), and (-1,1) respectively in $\mathbb{R}^2$. The barycenter of $\mathcal{S}^3$ is marked with a gray point, and corresponds to the origin in $\mathbb{R}^2$. The dashed lines have also been transformed.

A simple example in the three-alternative case is as follows:

$$\mathbf{U} = \begin{bmatrix} \mathbf{u_1} \\ \mathbf{u_2} \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} & 0 \\ \sqrt{\frac{1}{2 \cdot 3}} & \sqrt{\frac{1}{2 \cdot 3}} & -\sqrt{\frac{2}{3}} \end{bmatrix} \tag{4.7}$$

Here, the first row, $\mathbf{u_1}$, compares only alternatives 1 and 2, ignoring alternative 3. Then, $\mathbf{u_2}$ compares alternative 3 to a composite of alternatives 1 and 2. Each comparison is orthogonal from the other, permitting them to be interpreted as lying on independent axes in two-dimensional space. Figure 4.1 depicts the relationship between four compositions and the four ILR-transformed points that uniquely correspond to them using the basis $\mathbf{U}$ defined in Equation 4.7.

In many empirical applications, being able to specify particular contrasts (e.g., alternative 1 vs. alternative 2) is necessary to isolate interesting effects, so being able to specify a set of contrasts and construct the appropriate basis $\mathbf{U}$ is important.

4.4.2 Orthonormal Basis Specification Using The Gram-Schmidt Process

In their original specification of the ILR transformation, Egozcue et al. (2003) present a 'default' orthonormal basis, obtained through the Gram-Schmidt process for orthonormalizing vectors (Hefferon, 2013). The general form for this default matrix is as follows:

$$\mathbf{U} = \begin{bmatrix} \mathbf{u_1} \\ \mathbf{u_2} \\ \mathbf{u_3} \\ \vdots \\ \mathbf{u_d} \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} & 0 & 0 & \cdots & 0 & 0 \\ \sqrt{\frac{1}{2\cdot 3}} & \sqrt{\frac{1}{2\cdot 3}} & -\sqrt{\frac{2}{3}} & 0 & \cdots & 0 & 0 \\ \sqrt{\frac{1}{3\cdot 4}} & \sqrt{\frac{1}{3\cdot 4}} & \sqrt{\frac{1}{3\cdot 4}} & -\sqrt{\frac{3}{4}} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \sqrt{\frac{1}{d\cdot D}} & \sqrt{\frac{1}{d\cdot D}} & \sqrt{\frac{1}{d\cdot D}} & \sqrt{\frac{1}{d\cdot D}} & \cdots & \sqrt{\frac{1}{d\cdot D}} & -\sqrt{\frac{d}{D}} \end{bmatrix} \tag{4.8}$$

Here, each row $\mathbf{u_i}$ conforms to the requirements laid out above in Equation 4.6, such that each corresponds to a vector that is orthogonal from the rest, and has a unit length of 1.0. This is accomplished by setting the first row as a symmetrical contrast between u_1 and u_2. Next, the second row merges u_1 and u_2, setting both in contrast to u_3. The third row pits the first three elements against u_4, and so on.

The resulting orthonormal basis is entirely arbitrary, merely representing a convenient operation for rotating CLR-transformed data in such a way as to eliminate the extraneous dimension of the data. For models wherein the data are expected to display sphericity with respect to the various response alternatives, any $\mathbf{U}$ is as good as any other. However, if sphericity is not an assumption one wishes to make (for example, because some choice alternatives are more easily confused than others), then careful selection of a particular orthonormal basis permits specific comparisons to be made, as described below.

4.4.3 Orthonormal Basis Specification Using A Bifurcation Matrix

If an analyst has a particular set of contrasts that they wish to examine, they may build the corresponding basis manually. First, a bifurcation matrix $\mathbf{B}$ (Egozcue & Pawlowsky-Glahn, 2005) must be specified. In a bifurcation matrix, each element may have a 1, 0, or -1. Elements omitted from a particular contrast are assigned values of zero, and the rest are grouped into either those coded as positive or negative. For example, the basis in Equation 4.7 can be obtained by beginning with the following bifurcation matrix:

$$\mathbf{B} = \begin{bmatrix} \mathbf{b_1} \\ \mathbf{b_2} \end{bmatrix} = \begin{bmatrix} +1 & -1 & 0 \\ +1 & +1 & -1 \end{bmatrix}$$

Here, the first bifurcation splits vertex v_1 from vertex v_2 to form $\mathbf{b_1}$, whereas the second dimension $\mathbf{b_2}$ sets the vertex v_3 apart from the other two.

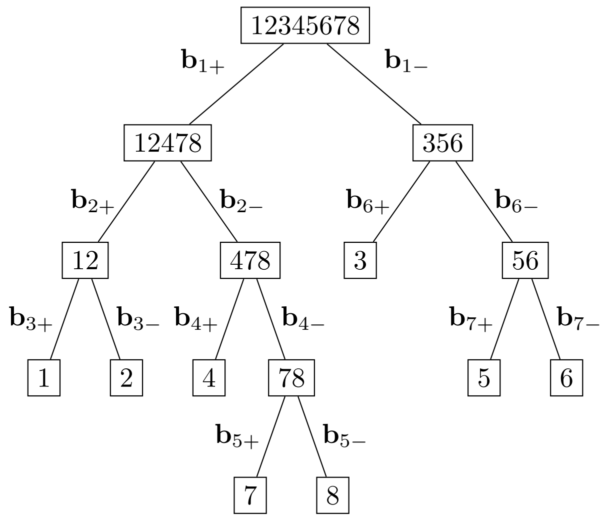

Figure 4.2: Example of a bifurcation matrix $\mathbf{B}$ subdividing eight alternatives into seven bifurcation vectors $\mathbf{b_i}$. Here, $\mathbf{b_{i+}}$ indicates the positive elements of $\mathbf{b_i}$, while $\mathbf{b_{i-}}$ indicates the negative elements; all other elements have a value of zero. In addition to providing an alternate way of visually depicting the bifurcation matrix, this also provides the reader with an algorithm for generating such a matrix.

A more complex example in which eight alternatives are being compared is described by the following matrix $\mathbf{B}$:

$$\begin{bmatrix} \mathbf{b_1} \\ \mathbf{b_2} \\ \mathbf{b_3} \\ \mathbf{b_4} \\ \mathbf{b_5} \\ \mathbf{b_6} \\ \mathbf{b_7} \end{bmatrix} = \begin{bmatrix} 1 & 1 & -1 & 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & 0 & -1 & 0 & 0 & -1 & -1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \end{bmatrix}$$

Here, the vector $\mathbf{b_1}$ begins by subdividing the space into two groups: (12478) and (356). The group (12478) is the further subdivided by rows $\mathbf{b_2}$ through $\mathbf{b_5}$, while the group (356) is further subdivided by rows $\mathbf{b_6}$ and $\mathbf{b_7}$. Figure 4.2 shows this matrix as a binary tree, subdividing the set of alternatives until only single-alternative 'leaves' populate the ends of all the branches. As this example should make clear, there is no requirement that the groups be of equal size. However, when converting $\mathbf{B}$ to $\mathbf{U}$, scaling factors (described below) are necessary to ensure that groupings are scaled correctly, in order to meet the requirements specified in Equation 4.6.

The translation of a bifurcation matrix **B** into an orthonormal basis **U** begins by identifying the scaling factors for each row $\mathbf{b_i}$:

$$\begin{aligned} \varpi_i &= \sum \{j \in \mathbf{b_i} | j = 1\}, \varpi_i > 0 \\ \omega_i &= -\sum \{j \in \mathbf{b_i} | j = -1\}, \omega_i > 0 \end{aligned} \tag{4.9}$$

Here, ϖ_i is the count of the positive elements in row $\mathbf{b_i}$, and ω_i is the count of the negative elements in that row. These scaling factors are essential because they permit each row to be normalized to a common unit length of 1.0. The values of each element in $\mathbf{u_i}$ is then computed based on the value of the associated element in $\mathbf{b_i}$ and the element counts ϖ_i and ω_i:

$$\mathbf{u}_{\mathbf{i},j} = \begin{cases} \sqrt{\frac{\omega_i}{\varpi_i(\varpi_i+\omega_i)}} & \text{if } \mathbf{b}_{\mathbf{i},j} = 1 \\ -\sqrt{\frac{\varpi_i}{\omega_i(\varpi_i+\omega_i)}} & \text{if } \mathbf{b}_{\mathbf{i},j} = -1 \\ 0 & \text{otherwise} \end{cases} \tag{4.10}$$

Note that, if the bifurcation matrix is not constructed properly, the matrix **U** resulting from applying Equation 4.10 will fail to meet the requirements of an orthonormal basis (in terms of its norms and inner products). The most reliable method for creating contrasts is to designate groups using a tree diagram like Figure 4.2, then assign values to **B** accordingly before proceeding to Equations 4.9 and 4.10.

4.4.4 Orthonormal Basis Specification Using Principal Component Analysis

Another approach is to perform a principal component analysis (PCA) on the dataset $\mathbf{X}^\circ$:

$$\mathbf{X}^* = [\mathbf{W}(\mathbf{X}^\circ)^\intercal]^\intercal, \mathbf{W} = \begin{bmatrix} \mathbf{w_1} \\ \vdots \\ \mathbf{w_d} \end{bmatrix} \tag{4.11}$$

Here, $\mathbf{X}^\circ$ denotes the CLR transformation of **X** and **W** denotes a matrix of the first d eigenvectors of the covariance matrix $\mathbf{X}^\circ [\mathbf{X}^\circ]^\intercal$ (Jolliffe, 2002). Although PCA will produce a matrix of D eigenvectors, the last eigenvector may be discarded because $\mathbf{X}^\circ$ necessarily has a singular covariance matrix, so the final eigenvector contains no information. **W** is thus a special case of **U**, and Equation 4.11 is effectively identical to Equation 4.5.

PCA is a good strategy for selecting an orthonormal basis if (1) there are sufficiently rich data to accurately estimate covariance, (2) it is reasonable that the alternatives might not be wholly independent of one another, and (3) there are no over-riding theoretical reasons to examine particular contrasts. Under these circumstances, PCA is likely to reveal the relationships that explain the preponderance of the variance. When there is sufficient power, this approach also permits factor analysis of the resulting orthogonal factors.

On the other hand, many compositional datasets are not sufficiently rich to provide a robust characterization of covariance. Under these circumstances, Equation 4.11 will still yield a usable orthonormal basis, but its contrasts may be idioscyncratic to the available data, or otherwise arbitrary. If the ILR transformation is required (e.g., for hypothesis testing), or if the data are relatively sparse, using the 'default basis' or manually constructing a basis using bifurcation are both more transparent analytic methods.

Chapter 5

Applications for Modeling and Inference

Because transformation makes it possible to obtain unbiased parameter estimates using familiar parametric tools, the way is open for testing various models of choice and behavior. This chapter provides several simple examples of models that might be tested. The emphasis is both on providing a sense of how to interpret the resulting parameters, and also how to structure data matrices in order to obtain parameter estimates using regression methods.

5.1 The Barycentric Matching Model

Armed with these tools, the barycentric matching model (Equation 1.12) may interpreted as a compositional linear model (Equation 3.9), to which the ILR transformation may be applied:

$$\mathbf{y}^* = \boldsymbol{\kappa}^* + \alpha \cdot \mathbf{x}^* \tag{5.1}$$

Note that despite its similarity to Equation 1.11, there is as much information contained in Equation 5.1 despite reducing the complexity of the space by one dimension. This is because although Equation 1.11 represented the data in D dimensions, the constraint of equilibrium demands that those data be confined to a d-dimensional manifold. The ILR transformation serves to rotate that manifold into alignment with the axes, such that the extraneous dimension is omitted from the analysis.

Typically, studies of choice consist of presenting a series of different schedules to an organism and observing the proportions of responding to each. Subsequent analyses will employ an explanatory matrix $\mathbf{X}^*$ of c different "configurations" (typically, different payout schedules) and a matrix $\mathbf{Y}^*$ of responses to those schedules. Each row of these matrices corresponds to a vector of d elements, ILR-transformed from a composition of size D.

$$\mathbf{X}^* = \begin{bmatrix} \mathbf{x_1^*} \\ \vdots \\ \mathbf{x_c^*} \end{bmatrix}, \qquad \mathbf{Y}^* = \begin{bmatrix} \mathbf{y_1^*} \\ \vdots \\ \mathbf{y_c^*} \end{bmatrix}$$
$$\mathbf{x_i^*} = \left(\mathbf{x^*_{i,1}}, \ldots, \mathbf{x^*_{i,\mathit{d}}}\right), \quad \mathbf{y_i^*} = \left(\mathbf{y^*_{i,1}}, \ldots, \mathbf{y^*_{i,\mathit{d}}}\right)$$

Hereafter, single elements in composition matrices will be identified using the notation $\mathbf{x^*_{i,\mathit{j}}}$ to identify item j from composition i in the matrix $\mathbf{X}^*$.

Estimating the parameters for Equation 5.1 given data of this kind can easily be accomplished using multiple linear regression, here represented in matrix form:

$$\begin{bmatrix} \mathbf{y^*_{1,1}} \\ \vdots \\ \mathbf{y^*_{1,\mathit{d}}} \\ \mathbf{y^*_{2,1}} \\ \vdots \\ \mathbf{y^*_{c,\mathit{d}}} \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 0 & \mathbf{x^*_{1,1}} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & \mathbf{x^*_{1,\mathit{d}}} \\ 1 & \cdots & 0 & \mathbf{x^*_{2,1}} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & \mathbf{x^*_{c,\mathit{d}}} \end{bmatrix} \begin{bmatrix} \kappa_1^* \\ \vdots \\ \kappa_d^* \\ \alpha \end{bmatrix}$$

To conserve space, design matrices of this kind will be presented in the following fashion:

$$\left[\mathbf{y^*_{i,\mathit{j}}}\right] = \left[\varrho_j^d, \mathbf{x^*_{i,\mathit{j}}}\right] \left[\boldsymbol{\kappa}^*, \alpha\right]^\intercal \tag{5.2}$$

Here, ϱ_j^d denotes a vector of length d in which element j is a 1 and all others are zero.

Despite the potentially high dimensionality of compositional data, the form of Equation 5.1 dramatically constrains the permitted range of predicted values. Figure 5.1 shows the positions of normative behaviors $\mathbf{x_i^*}$ and of observed behaviors $\mathbf{y_i^*}$. The arrows connecting each $\mathbf{x_i^*}$ to the origin are the *normative vectors*, in that they correspond to the operation $\oplus\mathbf{x_i}$ and serve as the point of reference for the independent variable. Given the model specified by Equation 5.1, the dashed line passing through $\mathbf{y_i^*}$ must be parallel with the normative vector, because the parameters $\boldsymbol{\kappa}^*$ and α may only reposition and rescale the normative vector; they may not rotate or distort it.

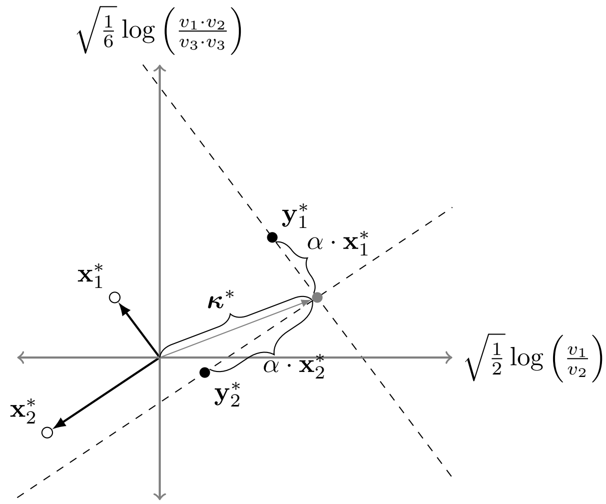

Figure 5.1: Given the normative vector $\mathbf{x}_i^*$, Equation 5.1 requires that the point $\mathbf{y}_i^*$ lie along a line parallel to the vector $\mathbf{x}_i^*$, with all such lines converging at the point $\boldsymbol{\kappa}^*$.

If Equation 5.1 is true, then given any two or more normative vectors, there must be a point on which the lines of possible response must converge, and that point must equal $\boldsymbol{\kappa}^*$. Note that this is true even in three or more dimensions, where arbitrary skew lines need not intersect.

A more subtle assumption is that the scalar value of α must be uniform across all compositions $\mathbf{x_i^*}$. This limits the possible outcomes of the regression, a problem originally identified as "symmetric approximation by leading term" by Natapoff (1970). This assumption is one of the great traps of traditional studies of choice, because it ensures that most data will appear to be well-represented by a single α value even when the underlying phenomenon does not obey the assumption of sphericity.

When choice proportions are influenced by multiple orthogonal factors (such as the reward rate R and the reward magnitude M), the barycentric model may be extended as a simple multiple regression:

$$\mathbf{y}^* = \boldsymbol{\kappa}^* + \alpha_R \cdot \mathbf{x}_R^* + \alpha_M \cdot \mathbf{x}_M^* \tag{5.3}$$

Here, each factor makes a contribution according to its own α parameter, and these factors are not assumed to interact (although adding an interaction term is straightforward). As discussed in Chapter 8, meta-analysis of a collection of animal studies suggest that treating these factors as linear and orthogonal in ILR-transformed space is supported by the bulk of the data.

One way in which α might fail to be uniform is that the value might shift gradually over time

Choice Models

matching model	free parameters $\boldsymbol{\kappa}^*$	α	equation
barycentric	d	1	(5.1)
multi-$\boldsymbol{\kappa}^*$	$\leq c \cdot d$	1	(5.4)
multi-α	d	$\leq c$	(5.5)
discriminability	d	d	(5.6)
multivariate	d	d^2	(5.7)

Table 5.1: Complexity in choice models described in this chapter. Here, c indicates the number of compositions in the dataset, and d corresponds to the dimensions of the Euclidean geometry $\mathbb{R}^d$ derived from the simplicial geometry $\mathcal{S}^D$. In the multinomial case, m corresponds to the number of explanatory variables. The data are presumed to be divided into q lotteries, where $q = 1$ unless otherwise specified.

(as a result of aging, for example). Another is that α might be consistent over time but might differ with respect to each of the contrasts specified by the orthonormal basis.

In the event of any of the above assumptions being violated, a more sophisticated model from the following subsections can be employed. Table 5.1 summarizes the different extensions of the barycentric matching model in terms of their parametric complexity, discussed in the following subsections.

5.2 Violation of Scale Invariance

One of the hallmarks of power laws in general is that they are *scale invariant*. That is, such models predict relative changes regardless of the scale. Because of the constraints of closure, it initially appears as though all compositional models will consequently be power laws, as closure's purpose is to 'cancel out' the scale information.

Although power laws are very popular, however, they are also very rarely appropriate in practice. For most phenomena, scale invariance is a property that is strictly limited to particular scale ranges, and most experimental examinations fail to test phenomena over a sufficiently wide range of scales

(Stumpf & Porter, 2012). The consequences of this can be seen quite clearly in the study of psychophysics. For example, although power laws describe the acuity of human hearing over a range of volumes that might be called 'conversational,' pronounced departures from power laws are nevertheless observed at extreme volumes, as well as at normal volumes in hearing-impaired patients (Marozeau & Florentine, 2007). Similarly, studies examining the subjective experience of the passage of time have identified scale invariance in many species over a range running from seconds to tens of hours (dubbed the 'interval timing' range) but breaking down for very short intervals ('ballistic timing') and intervals near 24h ('circadian timing') (Buhusi & Meck, 2005).

Given the possibility that model parameters might change as some function of scale, it is important to test the blind spot that compositions display with respect to scale. In keeping with the principles behind the ILR transformation, it is crucial that any scale data introduced into the model be orthogonal to the compositional contrasts. Consequently, a covariate that consists only of the information lost during closure is a suitable covariate for measuring the effects of scale.

Figure 5.2 provides a visual sense of how information regarding scale adds a new dimension to the data. Unlike the values in a composition (which are subject to the closure constraint), scale information is a proper ratio variable as per the taxonomy of Stevens (1946). In order to remain entirely consistent with the principle that independent variables should be orthogonal to one another, the only appropriate measure of a composition's scale is the geometric mean of its observed components. Any other value (such as the arithmetic mean) will produce a distorted model.

The emphasis on the geometric mean, as opposed to the arithmetic mean, is surprising to many analysts, particularly those who are accustomed to dealing with monetary rewards. The proposal, for example, that a coin toss with outcomes ($10 vs. $12) is on the same scale as a different cointoss with outcomes ($5 vs. $24) seems unusual given that the arithmetic mean is so much higher in the second pair than in the first. This, however, is a direct consequence of the way in which orthogonality is defined for compositional systems. Because power laws become linear under log-transformation, a model that seeks to examine scale in similar terms is limited to the geometric mean because, when log-transformed, this becomes an arithmetic mean and thus consistent with the Euclidean operations of the resulting sample space.

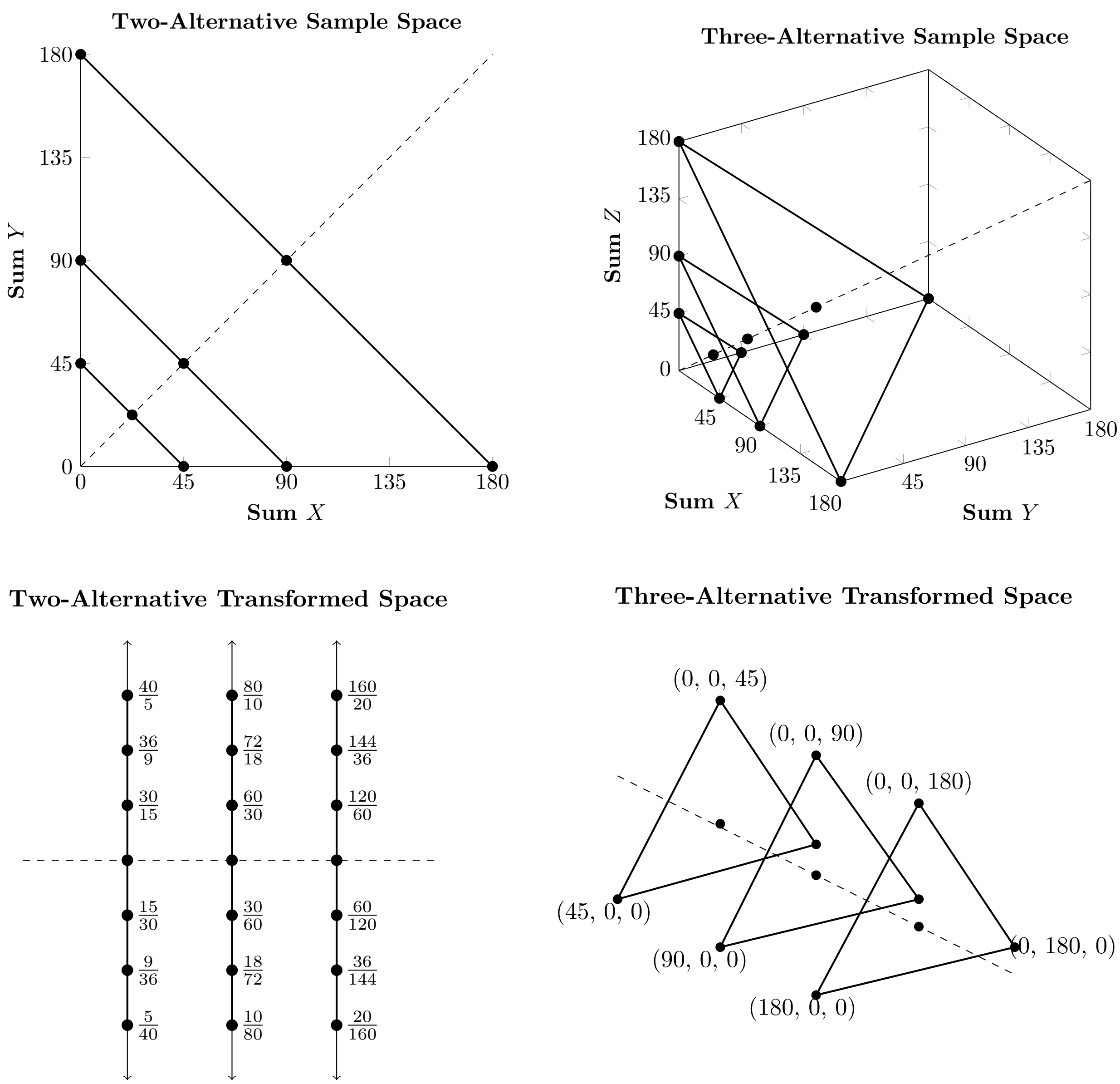

Figure 5.2: Although $\mathcal{C}(40, 5)$ and $\mathcal{C}(80, 10)$ correspond to the same composition, they differ by a factor of 2 with respect to their absolute scale. This scale variable may be included by using the geometric mean of the outcomes as an orthogonal dimension.

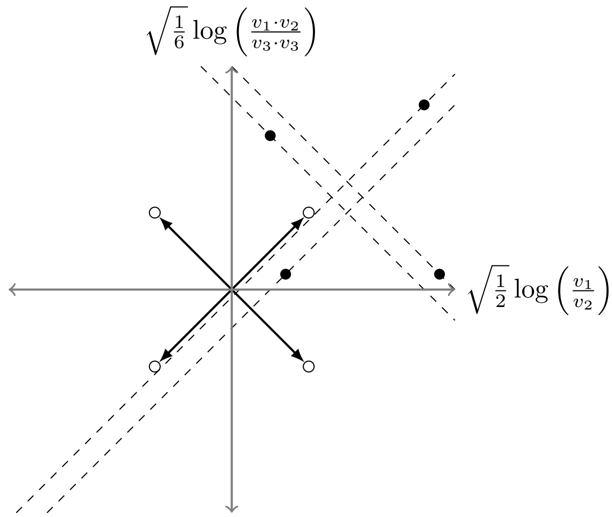

Figure 5.3: An example of non-convergent values of $\boldsymbol{\kappa}^*$. Given that each of the dashed lines (indicating possible positions for the dependent variable) must be parallel to one of the solid lines (indicating the normative vectors associated with the independent variable), there is no single point $\boldsymbol{\kappa}^*$ where the dashed lines meet.

5.3 Non-Convergent Bias

In traditional matching, the bias vector $\boldsymbol{\kappa}$ is interpreted as an organism's "point of indifference." That is: If an organism is entirely insensitive to shifts in the proportions of outcomes (i.e. if $\alpha = 0$), then they are presumed to chose with the proportions $\boldsymbol{\kappa}$ regardless of the values of $\mathbf{x}$. However, although this convergence is a requirement of traditional matching, a scenario that violates it is easily specified:

$$\mathbf{y_1^*} = \boldsymbol{\kappa}_1^* + (\alpha \cdot \mathbf{x_1^*})$$

$$\mathbf{y_2^*} = \boldsymbol{\kappa}_2^* + (\alpha \cdot \mathbf{x_2^*})$$

$$\boldsymbol{\kappa}_1^* \neq \boldsymbol{\kappa}_2^*$$

Figure 5.3 graphs a similar scenario, with the potential consequence that specifying a single point $\boldsymbol{\kappa}^*$ is no longer justified. When a standard barycentric matching model (Equation 5.1) is fit to such data, both the $\boldsymbol{\kappa}$ and α parameters have the potential to be distorted.

A straightforward solution is to modify Equation 5.2 to permit each composition to set its own point of indifference:

$$\left[\mathbf{y}^*_{\mathbf{i},j}\right] = \left[\underbrace{\varrho^d_j, \ldots, \varrho^d_j}_{c \text{ vectors}}, \mathbf{x}^*_{\mathbf{i},j}\right] \left[\boldsymbol{\kappa}^*_1, \ldots, \boldsymbol{\kappa}^*_c, \alpha\right]^\intercal \tag{5.4}$$

This approach is only advisable, however, when c is small and when such a distinction appears theoretically motivated, as the resulting model will have $(c \cdot d) + 1$ free parameters.

Another approach is to think of $\boldsymbol{\kappa}^*$ as a decomposable entity. The traditional treatment of $\boldsymbol{\kappa}^*$ as "bias" recognizes that an organism may be biased along many dimensions. A man may be left handed, and may also have a phobia of the color red. If these two biases are compositionally independent, then each will act as an independent perturbation, along with any additional influences (i.e. $\boldsymbol{\kappa} = \boldsymbol{\kappa}_{\text{hand}} \oplus \boldsymbol{\kappa}_{\text{red}} \oplus \boldsymbol{\kappa}_{\text{other}}$). Given proper experimental design, each of these biases could be included as explanatory variables in the regression model.

5.4 Inconsistent Sensitivity

While $\boldsymbol{\kappa}$ has largely been treated as an error variable, the sensitivity parameter α is frequently of experimental interest (Davison & McCarthy, 1988). This enthusiasm to cause variation in α is at odds with the simultaneous analytic assumption that it otherwise be invariant across time and scenario. Figure 5.4 presents an example in which each composition has its own value for α.

Equation 5.2 may easily be modified by specifying a vector $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_c)$ that gives each scenario its own sensitivity:

$$\left[\mathbf{y}^*_{\mathbf{i},j}\right] = \left[\varrho^d_j, \left(\varrho^c_i \cdot \mathbf{x}^*_{\mathbf{i},j}\right)\right] \left[\boldsymbol{\kappa}^*, \boldsymbol{\alpha}\right]^\intercal \tag{5.5}$$

This analysis will produce a single uniform bias across the c configurations, as well as a discrete α for each configuration. More compact models may be designated as well; $\boldsymbol{\alpha}$ may be a vector of any length from 1 to c.

In many circumstances, this is a economical solution, as it increases the number of free parameters from $d+1$ up to a maximum of $d+c$. Furthermore, variations in α already represent a familiar theoretical construct for many researchers, and the more flexible deployment of multiple α parameters in mixed model analyses may actually reduce the overall parameterization of data by allowing multiple experimental conditions to be intermixed into a single dataset. For example, if experimenters are contrasting choice performance in well-rested vs. sleep-deprived individuals,

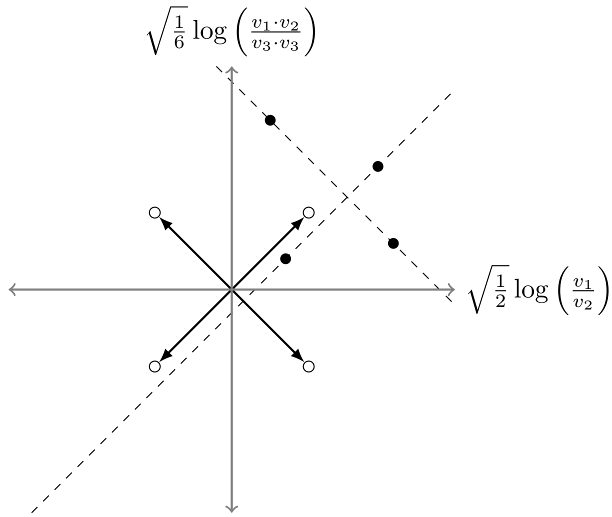

Figure 5.4: An example of inconsistent values of α. Here, despite each of the normative vectors (solid lines) being equal in length, the values of the dependent variable are such that each composition is a different distance from the point where the lines converge. Such inconsistency corresponds to a model in which the α parameter associated with each dependent composition has a different value by which it rescales the normative vector.

and determine that the evidence does not justify giving each condition its own bias parameters, the overall model complexity across the full experiment would drop from $(2d + 2)$ to $(d + 2)$ free parameters.

5.5 Nonuniform Discriminability

A notable failure of Equation 5.1 was reported in experiments manipulating the *discriminability* of response alternatives (Davison & Jenkins, 1985; Davison & Nevin, 1999). These studies argue that both the general matching law (Equation 1.8) and its more recent extension as the barycentric matching model (Equation 1.12) conflate an organism's ability to distinguish between the schedules responsible for determining outcomes and their ability to distinguish between the alternatives themselves. This problem is exemplified by the data introduced in Chapter 8 using data published by Davison & McCarthy (1994).

Figure 5.5 shows an example of how relative discriminability might manifest in ILR-transformed data: The contrast between v_1 and v_2 (represented by the horizontal axis) is much wider in the

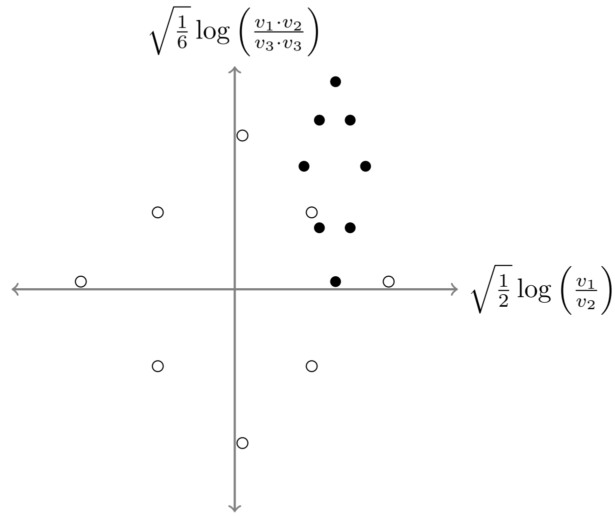

Figure 5.5: Non-uniform discriminability of the objective conditions $\mathbf{x}_i^*$ (white points) and the subjective choices $\mathbf{y}_i^*$ (black points), such that discriminations corresponding to the x-axis contrast are subjectively more difficult than those corresponding to the y-axis contrast.

objective case (white points) than in the subjective one (black points). If the model is constrained such that the subjective composition lines lie parallel to the normative composition lines, as in Equations 5.2, 5.4, and 5.5, the pattern of results in Figure 5.5 could be mistakenly interpreted as evidence for both non-convergent bias and inconsistent sensitivity.

A simpler account arises when the objective symmetry of the orthonormal basis is contrasted with the observed asymmetry of subjective experience. If one axis represents a difficult discrimination for subjects, while another represents an easy one, then the orthonormal basis adopted during the ILR transformation does not correctly capture the subjective differences experienced by subjects. In other words, one unit of movement along the x-axis in Figure 5.5 is not necessarily equal to one unit of movement along the y-axis.

The data in Figure 5.5 can easily be analyzed by asigning each of the orthonormal axes a distinct sensitivity parameter:

$$\left[\mathbf{y}_{\mathbf{i},j}^*\right] = \left[\varrho_j^d, \left(\varrho_j^d \cdot \mathbf{x}_{\mathbf{i},j}^*\right)\right] [\boldsymbol{\kappa}^*, \boldsymbol{\alpha}]^\intercal \tag{5.6}$$

Nonuniform discirminability here plays a role similar to that of eccentricity in the definition of an ellipse. Equation 5.6 effectively splits the analysis into a simple linear regression along each of the dimensions, so in d dimensions, the model has $2d$ free parameters (d from the parameters of $\boldsymbol{\kappa}^*$ and

d from the values for $\boldsymbol{\alpha}$).

The judicious selection of the orthonormal basis is crucial to this discriminability analysis, and should be approached with care. For an example of this, see Chapter 8.

5.6 Isometric Covariance

Although non-uniform discriminability increases model complexity considerably by treating each axis of the isometric transform as an *independent* linear regression, it cannot adequately account for *interdependent* linear effects that arise as a result of covariance. Doing so requires a full multivariate regression:

$$\mathbf{Y}^* = \mathbf{h}^\intercal [\boldsymbol{\kappa}^*] + \mathbf{X}^* [\boldsymbol{\aleph}]^\intercal , \quad \boldsymbol{\aleph} = \begin{bmatrix} \boldsymbol{\alpha_1} \\ \vdots \\ \boldsymbol{\alpha_d} \end{bmatrix}$$
$$\mathbf{h} = \underbrace{(1, \ldots, 1)}_{n \text{ ones}} \quad \boldsymbol{\alpha_i} = (\alpha_1, \ldots, \alpha_d) \tag{5.7}$$

Here, $\boldsymbol{\aleph}$ represents a full $d \times d$ design matrix, consisting of sensitivity vectors $\boldsymbol{\alpha_i}$. With the inclusion of the bias parameter $\boldsymbol{\kappa}^*$, Equation 5.7 has $(d \cdot D)$ free parameters. However, if $\boldsymbol{\aleph}$ is a diagonal matrix, the model effectively collapses to Equation 5.6.

The covariance matrix then arises from Equation 5.7 is presumed to be multivariate-normal. When translated back into the geometry of the simplex, the resulting probability distribution is a *logistic-normal distribution* (Aitchison & Shen, 1980; Egozcue et al., 2012).

Although this multivariate analysis can robustly represent a very wide range of possible effects, it does so at the risk of being exceedingly overparameterized for any but the most exhaustive datasets. In practice, many patterns of choice can be represented by simpler models, and full-covariance models are more likely to serve as the due diligence for model validation than they are to emerge as persuasive theoretical constructs in their own right.

Chapter 6

Considerations & Troubleshooting

Applying compositional methods to a broad range of data will inevitably pit the analyst against the limitations of the method. In some cases, these require that analysts make explicit assumptions, which must then be validated empirically.

Two potential complications are addressed in this chapter. The first, and most likely to arise, is the question of how to move forward if one component in a composition has a value of zero (seemingly putting the composition at odds with the constraints of closure). The other is the topic of the degree to which compositional models can be reconciled with multinomial logistic models, which make use of very similar but not quite identical model assumptions.

6.1 Dealing with Zeros

A common problem in dealing with choice data is that, occasionally, one of the alternatives will not get selected within the period being observed, compromising the requirement that all values in a composition be positive. It is important to distinguish between a missing value (about which nothing is known) and a recorded zero. This dissertation does not address the topic of missing values (for details, see Martín-Fernández et al., 2003), focusing instead on recorded zeros, which may be either *rounded zeros* or *structural zeros*.

A structural zero (also called an "essential zero") is a value that is known with confidence to represent a null value. Compositional analysis can still be applied to such data that contains structural zeros, by working with subcompositions confined to a facet of the full simplex. With a correctly-selected orthonormal basis, problematic dimensions may be singled out. However, structural zeros are theoretically challenging because their "zeroness" is itself informative, especially in studies of choice. Because the absolute refusal to choose a particular alternative is very different from the incidental avoidance of that alternative, structural zeros cannot be lightly ignored (for further discussion, see Martín-Fernández et al., 2011).

Contrastingly, a rounded zero may refer to a datum whose true value lies below the precision of the recording instrument (making it a "censored" datum). While structural zeros can often be planned for prior to the beginning of an experiment, rounded zeros may appear at any time in studies of choice (provided a subject simply does not choose a given alternative). In these cases, it may be advisable to substitute the observations $\boldsymbol{\beta}$ with a replacement vector $\mathbf{r}$, such that if $\beta_i = 0$, then $r_i = \delta_i$.

Martín-Fernández et al. (2011) recommend that the replacement vector $\mathbf{r}$ be computed using a "multiplicative replacement strategy" that otherwise preserves the compositional relationships within the data:

$$\mathbf{r} = \mathcal{C}\left(\mathbf{B}^r\right), \mathbf{B}_i^r = \begin{cases} \delta_i & \text{if } B_i = 0 \\ B_i & \text{if } B_i > 0 \end{cases} \tag{6.1}$$

The vector $\mathbf{B}^r$ is an unscaled replacement vector of the observed data, to which the closure operation is performed to produce $\mathbf{r}$. Because the nonzero observations are unmodified, the compositional metrics associated with the subcomposition consisting only of those values will be unaffected by the replacement.

The amount of distortion resulting from a replacement corresponds directly to the degree that substitution fails to preserve the metric properties of the observed data. Sanford et al. (1993) recommend $\delta_i = 0.55 \cdot \delta_r$, where δ_r represents the maximum rounding-off error. Choice data typically take the form of integer counts, making the minimum observable value for $\beta_i = 1$; hereafter, $\delta_i = 0.55$ is assumed to be suitable.

There is unanimity in the field that all substitutions must be accompanied by an analysis that measures how much distortion is introduced by any substitution and to validate the resulting model parameters (Aitchison & Egozcue, 2005; Martín-Fernández et al., 2003; Martín-Fernández & Thió-Henestrosa, 2006). These "sensitivity analyses" consist of examining distributional metrics for different putative values of δ_i. Aitchison (1986) specifies that suitable test values fall in the range

$\frac{\delta_r}{5} \leq \delta_i \leq 2\delta_r$. A straightforward approach to these validation tests is to use bootstrap methods (Wehrens et al., 2000) to compare the compositional geometric mean and total compositional variance (defined in terms of perturbation and powering rather than addition and multiplication) of the full dataset (for various values of δ_i) to those metrics in subsets of the data that exclude the replacement compositions.

6.2 Multinomial Logistic Regression As An Alternate Approach

In econometrics and public health, the multinomial logistic regression has become the standard approach for dealing with proportions of choice distributed among more than two alternatives (Hosmer & Lemeshow, 2000; Kwak & Clayton-Matthews, 2002). This approach is powerful, but a close examination reveals that it not only suffers the same difficulties noted above, but introduces others as well.

In practice, two forms of multinomial logistic regression are commonly employed: those that use d log-ratios of probabilities, and those that use D log-ratios comparing each probability to a "centered" composite value. These two approaches are therefore isomorphic to the ALR and CLR transformations respectively, differing only with respect to whether the model is interpreted as transformed-log-linear or logistic. Importantly, these two varieties of multinomial logistic regression have the same problems as their log-linear counterparts: The d log-ratios require the specification of an arbitrary "reference" category, while the D centered log-ratios produce an over-parameterized model with a singular covariance matrix. In most cases, the ALR-equivalent approach is used. Despite extensive application, there remains no canonical method for specifying the reference category (Egozcue et al., 2012), and comparisons across studies remains challenging.

Another prominent concern in multinomial logistic regression is that it requires "independence of irrelevant alternatives" (Hausman & McFadden, 1984). That is, for choices to be successfully modeled by this regression, each alternative must be perfectly discriminable from every other alternative, with no interaction between alternatives. This means that, despite its widespread use, this regression requires analysts to make assumptions that are objectionable, especially in human populations.

In addition to these problems, which are already of central concern to compositional analysis, fitting the parameters of a multinomial logit may require numerical procedures, a shortcoming shared by multinomial probit models.

In short, multinomial logistic regression should not be viewed as a panacea to the problems discussed in the preceding sections, as it does not change the problems Aitchison identified regarding geometry of the sample space. Fortunately, it also benefits from the same eventual solution to those problems.

6.2.1 Applying Compositional Ideas To Multinomial Models

Centered multinomial logistic regressions make use of a linear predictor function that relates M explanatory variables to N outcomes, defined as follows:

$$\varkappa_{n,i} = \kappa_n + (\alpha_{1,n} \cdot x_{1,i}) + \ldots + (\alpha_{M,n} \cdot x_{M,i}) \tag{6.2}$$

Here, the function's value for observation i of outcome n is given by a linear sum consisting of an intercept coefficient κ_n and parameter coefficients $\alpha_{m,n}$ that associate the m^{th} explanatory variable to the n^{th} outcome. These seek to explain the observed compositions in the matrix $\mathbf{Y}$, with one row per composition i and N columns (outcomes).

The centered multinomial logistic regression is an application of the *softmax* function, a popular transfer function used in machine learning (R. S. Sutton & Barto, 1998):

$$\mathbf{y}_{\mathbf{i},j} = \frac{\exp\left(\varkappa_{j,i}\right)}{\sum_{n=1}^{N} \exp\left(\varkappa_{n,i}\right)} \Big| \sum_{n=1}^{N} \mathbf{y}_{\mathbf{i},n} = 1 \tag{6.3}$$

The requirement that the probabilities for the various outcomes sum to 1 is a form of closure, but the denominator (which plays the role of ð in the closure operation) is the sum of all of the linear predictor functions in the model. The value of this denominator is typically estimated numerically.

This problematic denominator may be eliminated in two steps, beginning with the CLR transformation (Equation 4.3):

$$\mathfrak{g}\left(\mathbf{y}_{\mathbf{i}}\right) = \frac{\sqrt[N]{\prod_{n=1}^{N} \exp\left(\varkappa_{n,i}\right)}}{\sum_{n=1}^{N} \exp\left(\varkappa_{n,i}\right)}$$

$$\operatorname{clr}\left(\mathbf{y}_{\mathbf{i},j}\right) = \log\left(\frac{\exp\left(\varkappa_{j,i}\right)}{\sqrt[N]{\prod_{n=1}^{N} \exp\left(\varkappa_{n,i}\right)}}\right)$$

Because all items in $\mathbf{y_i}$ have the same denominator, the geometric mean $\mathfrak{g}(\mathbf{y_i})$ necessarily also has it, and the ratio causes it to cancel out. However, as in other cases of the CLR transformation, the resulting data is now subject to the equilibrium constraint (Equation 4.3).

The ilr transformation simplifies the form of the equation and cancels out the denominator:

$$\mathbf{y}^*_{\mathbf{i},j} = \sum_{n=1}^{N} (\mathbf{u}_{\mathbf{i},n} \cdot \varkappa_{n,i}) \tag{6.4}$$

Finally, the original explanatory coefficients $\beta_{m,n}$ from the softmax function can be solved for using partial differentiation.

It is important, however, not to lose sight of the original goal of such an analysis: To transform compositional data in the geometry $\mathcal{S}^N$ into the geometry $\mathbb{R}^{N-1}$, in order to describe it in terms of linear estimators in an independent geometry $\mathbb{R}^M$. Unless the softmax function is central to the analysis, an analyst could just as easily fit the following model:

$$\mathbf{y}^*_{\mathbf{i},j} = \varkappa_{j,i}$$

This approach can be conceived of as examining the impact of the M explanatory variables on a particular isometric logratio j, with a total of d ratios, doing so without asymmetry and without requiring that a logistic function be invoked. This permits analysts to perform all analyses previously undertaken using multinomial regression in a real geometry instead.

Using compositional transformation instead of multinomial logistic regression also bypasses the problem of independence from irrelevant alternatives. As Chapter 5 demonstrates, these analyses can successfully account for differential discriminability and covariance in compositional data, doing so without need for numerical approximation.

Part II

Experimental Methods & Results

Chapter 7

Experimental Context

'Choice' and 'decision-making' are immense topics spanning many academic disciplines. Given such a scope (and corresponding variety in technical vocabularies), miscommunication is all too easy. The focus of the experimental data in this dissertation is choice behaviors that have the following characteristics:

- The available choice alternatives are a discrete set present in a confined space. The studied behavior is constrained to this domain.
- Apart from this constraint in the scope of the task, subjects are free to move and respond at will.
- Subjects will make a large number of responses in any given experiment.
- The consequences of these choices are either a reward delivery or a null outcome. Acutely aversive outcomes (such as electric shock) are not considered.
- Individual choices are uncertain, insofar as any choice may or may not, at any point in time, yield a reward.
- Individual choices are not especially costly, with respect to the various kinds of costs (e.g., the expenditure of effort, the loss of time, and the loss of resources).

Superficially, imposing these restrictions on the present experiments enormously constrain the scope of the research. However, because choice was an unmanageably broad topic to begin with, these constraints are necessary if a precise analysis is to be made. Naturally, each of these constraints may be relaxed, and doing so in each case would yield compelling research. However, there are several reasons why these constraints provide the best starting point for a rigorous analysis.

All experimental paradigms rely on useful simplification. This is done for two reasons: To minimize the random effects of factors outside of experimental control on obtained measurements, and to observe whether changes in experimental conditions yield consistent fixed effects. Both work in favor of capturing phenomena in isolation and measuring them with enough precision to be able to build models and subsequently make predictions. Constraining a choice task to a small space with unambiguous response alternatives works toward these objectives, as does collecting large samples of data.

With respect to the other constraints listed, however, the main benefit is that these describe the vast majority of choices made by most organisms under most circumstances. Large-scale choices that are known in advance to have permanent effects on life direction (buying a house, choosing a college major, adopting a child) are made very infrequently by any given human and are unheard of in other species (who lack the capacity to dwell on questions like 'Where will I be in five years?'). Instead, most choices have low stakes, yield unremarkable outcomes, and are made almost continuously. For the most part, these choices are so 'micro' that they lie below the criterion of interest of microeconomics, and happen so automatically that they reside at the hinterlands of philosophy's debates about free will.

Despite being individually trivial, however, these very small decisions accumulate into substantial effects as a result of their sheer volume. It is on the basis of the sum, and not of the individual gambles, that the house always wins. Furthermore, these small-scale, frequent decisions are not limited to domains that can be measured in simple econometric terms. For example, microaggressions and other manifestations of prejudice may appear 'harmless' to the perpetrator, but can result to 'structural' manifestations of racism or sexism, even among individuals who describe themselves as being committed to an egalitarian view (Brooks & Purdie-Vaughns, 2007).

These many small choices are also crucial to voluntary behavior, and so to free will (Neuringer & Jensen, 2010). Humans frequently make choices that are adaptive without using deliberative strategies, and without being able to articulate how they were influenced by feedback. At the same time, these choices are routinely characterized by a degree of randomness that renders them difficult to predict in exact terms. Consequently, in order to study voluntary behavior as a process, it is essential to begin with the vast majority of those behaviors that are identified as voluntary, even if they are not behaviors that appear individually to be of much consequence.

Finally, the decision to avoid paradigms in which choices are costly is merely a strategy to simplify the experimental topic. The subject of aversive outcomes is clearly of substantial importance, and may differ qualitatively from choices made in strictly rewarding scenarios. Nevertheless, a comprehensive understanding of a narrowly defined phenomenon is more productive than a superficial understanding of a broad one. Extending the method to aversive outcomes will be a necessary direction in future work, but is treated as lying beyond the scope of the present study in order to achieve depth at the expense of breadth.

Chapter 8

Previously Published Data

As previously described, Baum's "generalized matching law" (Equation 1.8) is among the most successful models for describing overall proportions of response in concurrent choice designs, despite its parametric simplicity. Shortly after its publication, a review of available concurrent datasets found that it was able to account for over 90% of the variance in most cases (Baum, 1979; Davison & McCarthy, 1988). This success has not been limited to the study of concurrent choice, however. Equation 1.8 is effectively a log-odds function (Barnard, 1949), itself an extension of the logistic function (Reed & Berkson, 1929; Berkson, 1944). Log-odds functions have been so successful at modeling phenomena in every area of psychology that a recent review describes them as "ubiquitous" (H. Zhang & Maloney, 2012).

Equation 1.8 is also a power law, which may seen as an extension of the "psychophysical law" proposed by Stevens (1957) as a reinvention of the psychophysical functions developed by Weber and Fechner. Furthermore, Equation 1.8 belongs to a set of identities observed by the barycentric matching model, shown in Equation 1.12 (Jensen & Neuringer, 2009).

Contention persists regarding the theoretical viability of power laws in explaining psychophysical phenomena. In some cases, tests that challenge model assumptions reveal scenarios in which they yield incorrect predictions (e.g. Augustin, 2008), while in other, more sophisticated models incorporate known difficulties into a more complete framework (e.g. Luce, 2002). In such endeavors, it is crucially important distinguish between a model that may be incorrect and a model that may be incomplete. Equation 1.8 is a mathematically convenient form of Equation 1.12 when dealing with two alternatives, because it lends itself to linearization through the use of a logarithmic transformation. The full model proposed by Equation 1.12 has been difficult to test in part because it does not follow a well-established parametric form.

Controversy also surrounds the theoretical importance of Equation 1.8 for explaining choice under concurrent schedules. Because Equation 1.8 is a molar description of behavior (that is, a general description of many observations; Baum, 2002), and because it presents prediction in terms of scale-invariant ratios, its explanatory power for understanding behavioral processes has been questioned (e.g. Navakatikyan & Davison, 2010).

Limits of Equation 1.8 have also been demonstrated empirically. Davison & Jenkins (1985) propose an alternative model in which the stimuli that signal the different schedules of reward are confusable to varying degrees. When this approach was extended to three alternatives by Davison & McCarthy (1994), parameter estimates could not be made to reconcile with Equation 1.8 when two of the schedules were highly confusable relative to a third alternative.

Rather than conclude that apparent empirical violations of Equation 1.8 invalidate it as a compelling model of choice behavior, the approach described here favors the view that it is incomplete, and that many of its shortcomings can be remedied without discarding the log-odds characteristics that have made it such a powerful model of molar data. By reconceptualizing Equation 1.8 as a relationship between mathematical objects called compositions, practical analysis of matching data can include any number of alternatives. Furthermore, as described below, this approach clarifies how to extend the model in ways that can accommodate both violations of scale invariance and differential discriminability.

The following re-analyses are adapted from results reported by Jensen (2014).

8.1 Reward Frequency vs. Magnitude: Elliffe et al. (2008)

A multiple regression may take any number of factors into consideration, and this may be exploited to contrast how different aspects of reward contribute to responding. Elliffe et al. (2008) reported an experiment in which two concurrent variable interval schedules varied the frequency of reward delivery between a (300s : 33s) to a (33s : 300s) ratio, and, upon reward delivery, dropped a different number of food pellets for each lever, varying between a (1 : 7) ratio and a (7 : 1) ratio. If frequency and magnitude are orthogonal manipulations, then each may be converted into a log-

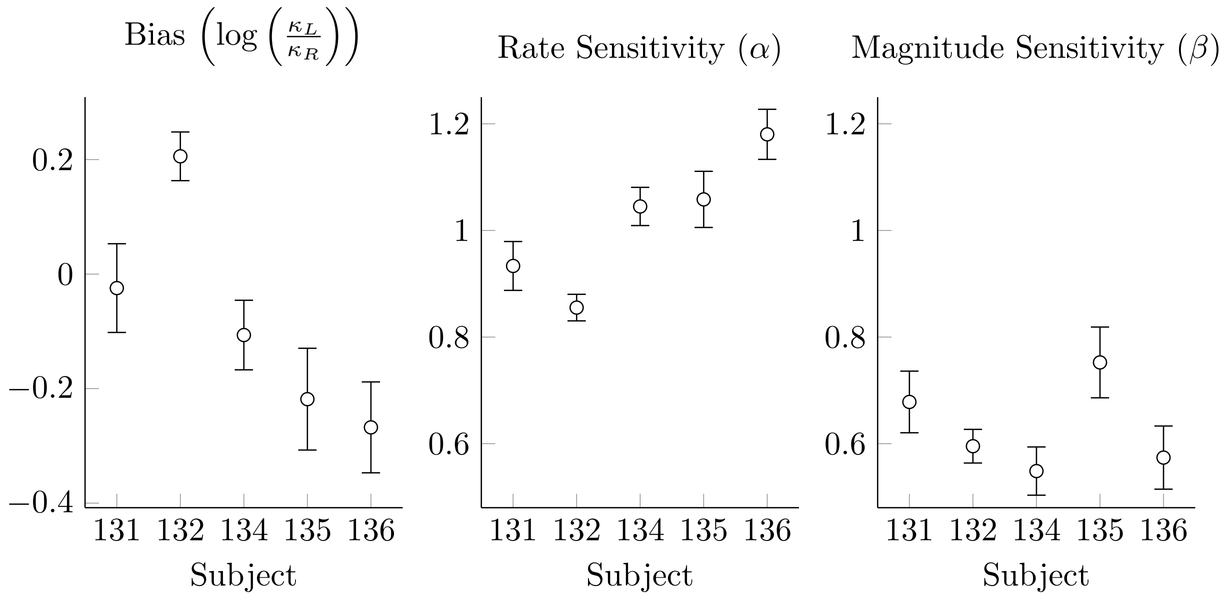

Figure 8.1: Estimated model parameters for Equation 8.1 from the re-analysis of data published by Elliffe et al. (2008). Error bars correspond to 1 standard error.

ratio, as per the ALR transformation (which may be used without concern in the two-item case), yielding the following equation:

$$\mathbf{B}^{@} = \boldsymbol{\kappa}^{@} + \alpha\mathbf{R}^{@} + \beta\mathbf{M}^{@} \tag{8.1}$$

Here, the ratio $\mathbf{B}^{@}$ corresponds to the ratio of responses $\log\frac{B_1}{B_2}$, as per Equation 7. $\mathbf{R}^{@}$, in turn denotes the log-ratio of reward probabilities, while $\mathbf{M}^{@}$ denotes the log-ratio of reward magnitudes.

Figure 8.1 plots the estimated parameters in Equation 8.1 for each of the five subjects. Although some individual differences are evident, the overall pattern is clear: a higher slope parameter associated with the rate of reward than with the magnitude of reward, and little or no operandum bias.

The goodness of fit obtained by Equation 8.1 is very high. Figure 8.2 plots the best fitting plane described by Equation 8.1 for subjects 132 (the best-fitting subject) and 135 (the worst fit), as well as the raw data. The lines drawn for each point show the residual distance from the best-fitting plane.

Although Elliffe et al. argue that relative reward rates and magnitudes are interdependent, their regressions do not include the interaction term that would test this proposal directly. Instead,

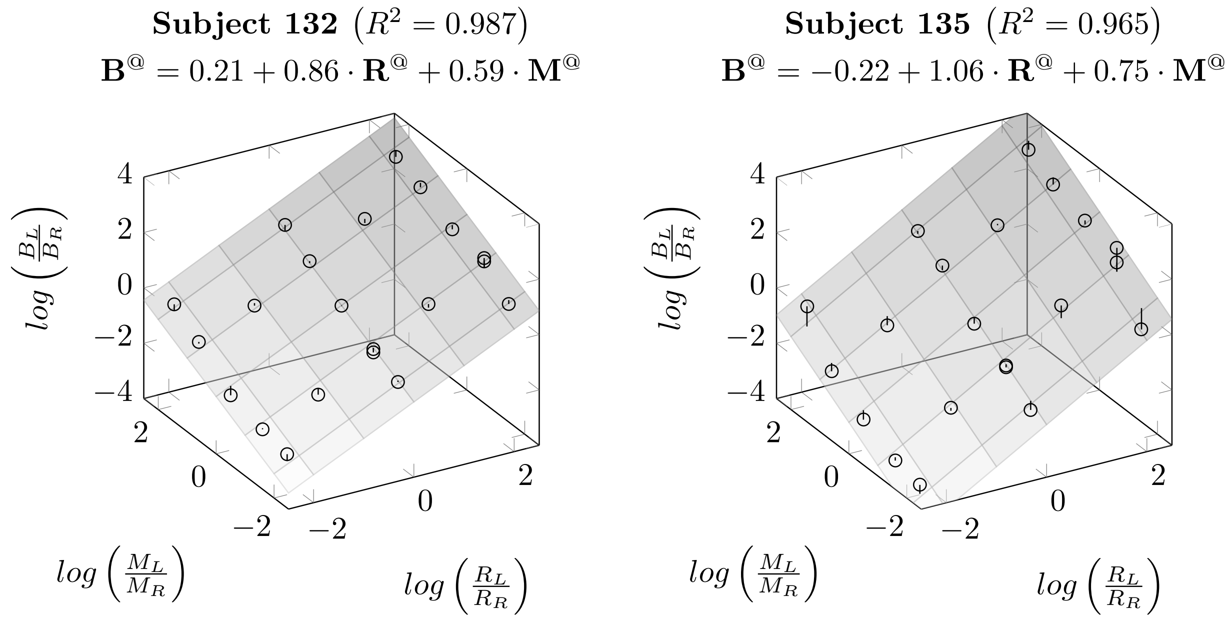

Figure 8.2: The best-fitting plane for two subjects from Elliffe et al. (2008), as well as the raw data used to fit those functions. Lines indicate the residual distance from the plane of the function to each datum.

they rely on simple regressions applied to subsets of the data. Cording et al. (2011) test for this interaction term as part of a residual meta-analysis of five studies that manipulated reward magnitude. According to this meta-analysis, there is not persuasive evidence for non-linearity in the studies examined.

In conclusion, a compositional reanalysis of the data published by Elliffe et al. (2008) yields a model that accounts for nearly all of the variance and shows no indication of a departure from linearity. Although this analysis is largely familiar, it nevertheless provides a bridge that analysts can use to acquaint themselves with the basic concepts outlined in this paper.

8.2 Discriminability & Covariation: Davison & McCarthy (1994) and Davison (1996)

As was emphasized in the discussions of compositional transformation in Chapter 4, the translation from a two-alternative problem to an n-alternative problem is non-trivial, and observing linear functions in two-alternative log-odds plots is no guarantee that the model will generalize. As the following example will demonstrate, the compositional perspective not only provides a way of understanding when and why models of choice fail, but also provides a framework within which more successful models can be built with a minimal increase in model complexity.

One of the most striking demonstrations of the insufficiency Baum's Generalized Matching Law (Equation 1.8) was a 3-alternative scenario reported by Davison & McCarthy (1994). In their study, pigeons responded to three operanda using a switch key procedure, where each operandum was governed by a different concurrent variable interval schedule. The procedure was then extended using the same subjects, with further data reported in Davison (1996). Throughout these experiments, two of the operanda were identified by fixed colors: green-yellow (with a wavelength of 560 nm) and orange-red (630 nm). The third alternative was assigned a different wavelength W, in order to manipulate how easy it was to discriminate from the other two. In Davison & McCarthy (1994), data was collected at $W = 600$ nm (an easy discrimination) and $W = 623$ nm (which was difficult to discriminate from 630 nm). Davison (1996) then reported further data collected at $W = 563$ nm, 570 nm, 615 nm, 619 nm, and 627 nm.

The resulting data cannot be accounted for by Equation 1.8, which yields inconsistent predictions depending which two alternatives are included in the ratio. For example, when comparing 560 nm to 630 nm using Equation 1.8, subjects appeared sensitive to the schedule; however, data from the same session comparing 623 to 630 appeared insensitive. Because compositions are structured in a recursive fashion (e.g. the ternary plots in Figure 2 are bounded by 3 edges), it follows that Equation 5.1 is also unable to accommodate the reported patterns in the data. To understand why, consider the effects of bias $\boldsymbol{\kappa}$ and sensitivity α in Equation 5.1. The bias term repositions the barycenter of a set of points, and sensitivity rescales that set; however, neither parameter has the means to distort a set of points in such a way that can explain the inconsistency in the pairwise results. Another way to describe the problem is in terms of a violation of the assumption of sphericity. In order for a model similar to Equation 6 to be successful in describing the data reported in these studies, it must accommodate the ability to rescale the data differently along multiple axes. However, given the constraints of SALT, described above (see also Natapoff, 1970), a matching model cannot assign an independent sensitivity parameter to each response alternative.

On the basis of the inconsistency in pairwise sensitivity estimates, and the assumption that

SALT forces Equation 5.1 to only have one global sensitivity, the results reported in Davison & McCarthy (1994) and Davison (1996) were presented as support for the contingency discriminability model (Davison & Jenkins, 1985), which takes the following form:

$$\frac{B_1}{B_2} = \lambda \frac{R_1 - \rho R_1 + \rho R_2 - \omega}{R_2 - \rho R_2 + \rho R_1 - \omega} \tag{8.2}$$

The λ parameter serves a role similar to the bias, while the ρ parameter is analogous to sensitivity. The function of the ω parameter is less clear-cut, providing a background of "subtractive punishment."

Equation 8.2 is problematic for several reasons. To accommodate two-operandum data, it introduces a third parameter, doing so because the model is otherwise unable to accommodate overmatching (i.e. cases in which $\alpha > 1$). Furthermore, although Equation 8.2 models varying discriminability well for two operanda, there is no straightforward way to generalize to more than two alternatives. Davison & McCarthy (1994) provide a three-operandum generalization, which requires a different ρ_{ij} parameter for every pair of operanda and quietly omits the λ parameter (whose general form is far from clear). The resulting function can only be fit numerically.

In order to assess the validity of the Equation 8.2, N. P. Sutton et al. (2008) performed a residual meta-analysis of 20 studies. Because Equation 8.2 made use of addition and subtraction, it is expected to be non-linear when the data are log-transformed, so testing for non-linearity in the residuals of the log-transformed data provides a robust test of Equation 8.2's predictions. This meta-analysis found no evidence of non-linearity on log-scaled data, suggesting that Equation 8.2 is unlikely to generalize.

Although the simplest incarnation of Equation 5.1 cannot be reconciled with these data, the ILR transformation nevertheless provides a means by which the compositional qualities of Equation 1.8 can be reconciled with the differential discriminability in the data published by Davison & McCarthy. In addition to eliminating the extraneous dimension responsible for SALT, the orthonormal basis $\mathbf{U}$ in Equation 5.5 permits the analyst to specify the contrasts of interest as a multivariate normal distribution whose length and breadth are described different sensitivity parameters. This allows discriminability to be understood as a covariance matrix in the ilr-transformed sample space.

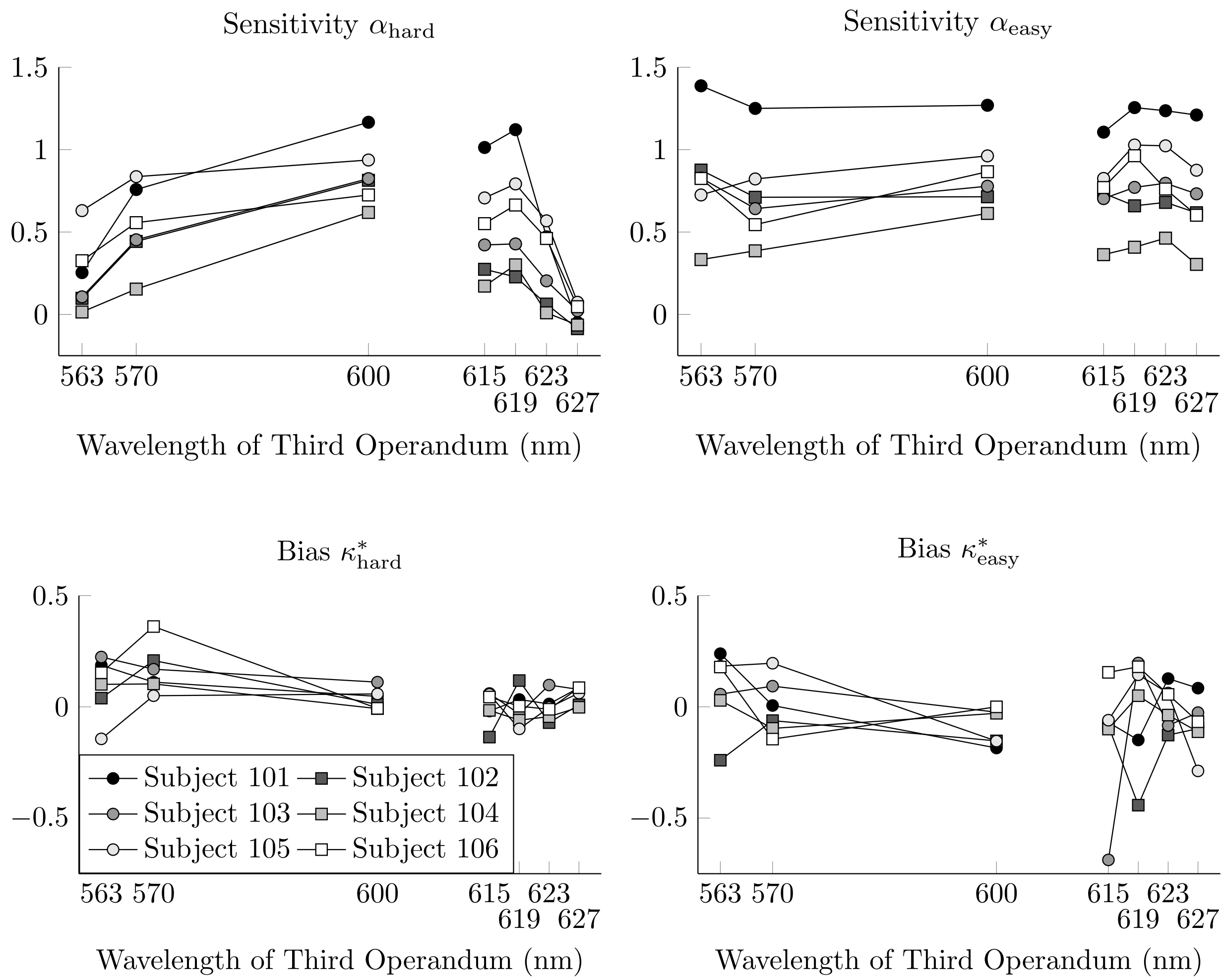

Figure 8.3: Estimated model parameters from the re-analysis of data published by Davison & McCarthy (1994) and Davison (1996). The points on the left-hand side of each plot are obtained using a different orthonormal base than those on the right-hand side, as indicated in Equation 8.3.

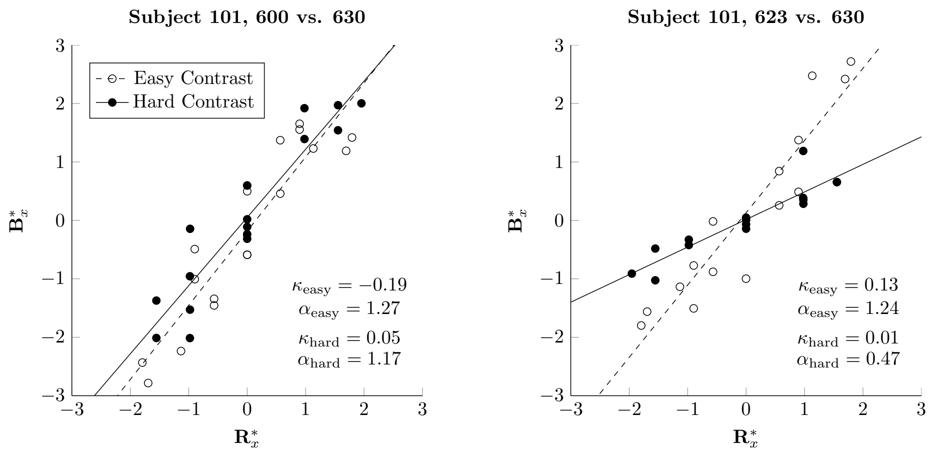

Figure 8.4: Best-fitting lines for Subject 101 in the (600 vs. 630) and (623 vs. 630) conditions, as reported by Davison & McCarthy (1994). The "hard contrast" refers to the $\mathbf{u}_{xz}$ (left plot) and $\mathbf{u}_{yz}$ (right plot) contrasts in Equation 8.3, whereas the "easy" contrast compares the third, distinct alternative (560 nm) to the composite of the other two. Solid lines were fit to the black points and dashed lines were fit to the white points using entirely independent regressions.

To re-analyze Davison and McCarthy's data, two different $\mathbf{U}$ matrices were used:

$$
\begin{aligned}
\mathbf{U}_{xz|y} &= \begin{bmatrix} \sqrt{\frac{1}{2}} & 0 & -\sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{6}} & -\sqrt{\frac{2}{3}} & \sqrt{\frac{1}{6}} \end{bmatrix} \\
\mathbf{U}_{yz|x} &= \begin{bmatrix} 0 & \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} \\ \sqrt{\frac{2}{3}} & -\sqrt{\frac{1}{6}} & -\sqrt{\frac{1}{6}} \end{bmatrix}
\end{aligned}
\tag{8.3}
$$

Here, the first column in each matrix corresponds to the 560 nm stimulus and the second column corresponds to the 630 nm stimulus. The third column corresponds to the varying stimulus, whose wavelength W was manipulate experimentally. The $\mathbf{U}_{xz|y}$ case was used for $W = 563$ nm, 570 nm, and 600 nm. The first contrast is between the 560 nm operandum and the varying alternative, while the second contrast compares 630 nm to a composition of the items in the first contrast. The $\mathbf{U}_{yz|x}$ case was used for $W = 615$ nm, 619 nm, 623 nm, and 627 nm, because it contrasts the longest wavelength with the varying stimulus. For convenience, the first pairwise contrast in each case is hereafter labeled as "hard" (comparing the manipulated wavelength W to its closest fixed counterpart), whereas the second composite contrast is labeled as "easy." Since each contrast is orthogonal to the other in ILR-transformed space, each was fitted with its own bias and sensitivity parameters, resulting in four free parameters from two regressions. Parameters were assessed separately for each contrast.

Figure 8.3 plots the four free parameters for all six subjects, across all seven wavelengths. The break in the lines between 600 nm and 615 nm signals that two different U matrices were used, so parameters on the left side of each figure describe a different kind of distortion that those on the right side. Overall, this analysis shows a decrease in the sensitivity of similar stimuli (α_{hard}) as a function of that similarity, without any lack of sensitivity in the composite contrast (α_{easy}). Bias was generally minimal.

Parameters were estimated using standard regression techniques. Figure 8.4 shows the regression lines used to fit Subject 101's parameters for the 600 nm and 623 nm cases. These lines, which were representative of fits for all subjects, explain a considerable degree of the variance and do not require numerical estimation techniques or specialized software. Furthermore, the data do not show signs of non-linearity.

In conclusion, compositional analysis of choice can accommodate differential discriminability with a minimal increase in model complexity. Once an analyst has grown comfortable with ILR transformation, either traditional "single global sensitivity" models may be fit, or theoretically compelling contrasts may instead be singled out using appropriate orthonormal bases.

Chapter 9

Concurrent Choice: Four vs. Six vs. Eight Alternatives

Although the consistent efficacy of compositional methods in existing datasets is encouraging, many basic questions about choice remain unanswered in light of the insufficiencies of traditional analyses. One of the most basic questions is, “How does changing the number of alternatives impact behavior?” This chapter will show how the compositional approach facilitates the analysis of data whose dimensionality is far greater than can be visualized easily. From these results, a series of tools are described for understanding and characterizing that behavior.

The validity of matching models in realistically complex situations is at stake in this work. If the orderly relationships observed in two-alternative experiments falls apart under these more complex conditions, then the degree to which those models can be applied to understanding behavior outside the lab will be minimal. There are, however, several ways in which orderly behavior might be observed.

The initial focus is on the sensitivity parameter α in the barycentric matching model (Equations 1.12, 5.1), for which unbiased estimates are made possible by the compositional paradigm. Changing the number of alternatives in the task might influence α in several ways. The naïve prediction from a cognitive perspective is that lower estimates of α should be expected when more alternatives are available, because adding alternatives increases the computational load that the task imposes on the subject (and in particular, on working memory), and this in turn should make subjects less effective at processing task-related information (Ackerman, 1988). This intuition has been popularized in the form of the "paradox of choice" (Schwartz, 2004). However, systematic reviews of the literature suggest that this phenomenon is not consistently observed, and may be the result of idiosyncratic confounds (including experimenter expectancy) (Scheibehenne et al., 2010). Consequently a second plausible possibility is that the α parameter will remain relatively stable across conditions.

In order to examine this question directly, ten rats were exposed to a series of concurrent schedules in which either four, six, or eight alternatives were simultaneously available, each with a differing degree of relative richness. The studies presented in this dissertation instead make use of the 'Turn-Based Foraging' paradigm, in which each selection of an alternative 'forages' at that location. Rewards are set up randomly (and in secret) at all locations, and rewards awaiting collection are delivered whenever the animal forages at its location.

Although a handful of studies have used the Turn-Based Foraging paradigm (Lau & Glimcher, 2005; Jensen & Neuringer, 2008, 2009), it is not the standard paradigm in the study of choice. In comparative animal studies, choices have more commonly been studied using 'concurrent variable interval' schedules, in which reward availability depends on the passage of time rather than on subject-initiated turns. Contrastingly, the dominant paradigm in behavioral economics is the use of 'repeated lotteries,' in which participants are asked to make a series of independent gambles. Each of these methods has shortcomings that the Turn-Based Foraging paradigm helps to address.

One difficulty with variable interval (VI) schedules is that there is considerable ambiguity in the literature about what the word 'variable' entails; some studies sample from an exponential distribution directly (the agreed-upon best practice), while others use discretized exponential approximations (e.g. Fleshler & Hoffman, 1962), and still others make use of non-exponential distribution (e.g. Alsop & Elliffe, 1988). Problematically, these procedural differences impact behavior, complicating comparisons across the VI literature (Taylor & Davison, 1983; Elliffe & Alsop, 1996).

VI schedules also increase the subjective complexity of the task, because a subject must track not only the relative richness of the alternatives, but also their own overall response rate. If rewards become available only once per minute (on average), then subject making one response per second is 'wasting' considerable effort. Evidence suggests that subjects are sensitive to these temporal dynamics (Nevin, 2003), which helps explain why VI schedules routinely elicit lower response rates than variable ratio (VR) schedules (Baum, 1993).

Turn-Based Foraging also permits very easy calculation of the objective probability of reward on any given response. Given the logic detailed in Algorithm 1, the probability is a simple function of the number of turns t_i since the last response to the alternative C_i, which follows the cumulative geometric distribution (Jensen & Neuringer, 2008):

$$\Pr\left(\text{Reward}|C_i\right) = 1 - \left(1 - P_i\right)^{t_i} \tag{9.1}$$

The more turns an organism spends away from an alternative, the more opportunities that alternative will have had to set up a reward. Thus, as t_i becomes large, the value of $(1 - P_i)^{t_i}$ becomes small and the cumulative probability of a reward approaches 1.0.

This simple relationship between the distribution of effort and the cumulative probability of being rewarded provides Turn-Based Foraging with another of its strengths: Optimal behavior consists of intermixing the response alternatives at an equilibrium point. This contrasts with the repeated lotteries popular in behavioral economics, where 'optimal' consists either of exclusively choosing one option over another, or displaying indifference (Vulkan, 2000). Because not all behavior can be reduced to either exclusive preference or strict indifference, a methodology is required that permits experimental investigation of the range between these alternatives.

9.1 Methods

9.1.1 Subjects

Subjects were 10 male albino Sprague-Dawley rats (Charles River, NY), weighing 450-650 g. Although familiar with the operant chambers and trained to press a single lever on a continuous reinforcement (CRF) schedule prior to running, they otherwise began naïve to the task. Subjects were pair-housed and were given one hour of access to food following each day's experimental session, with additional time provided in the event that any animal began to lose weight. Additionally, subjects were given free access to food from Friday afternoon to Sunday afternoon during each week.

Algorithm 1: The dependent concurrent variable ratio schedule (depVR), as implemented in Rat Experiment 1.

```
Data: number of alternatives D, setup probabilities P
begin
    F ← Zeroes(D)                      /* Boolean array for storing rewards */
    repeat
        x ← Response()                          /* Subject selects an alternative */
        if F(x) == 1 then
            F(x) ← 0;                  /* If food is available, remove counter */
            Reward();                                      /* Reward delivery */
        for i = 1 to D do
            if Rand() < P(i) then
                F(i) ← 1;      /* If setup condition is met, set counter to 1 */
    until 100 rewards earned or 60 minutes elapsed
```

9.1.2 Apparatus

Data were collected using four operant chambers manufactured by Med Associated (Model ENV-008). Each chamber measured 30.5 cm x 24.1 cm x 21 cm, with a grid floor of evenly-spaced stainless steel rods (4.8 mm diameter). The front and back walls were clear plastic, while the left and right walls consisted of modular steel slots that could be reconfigured. Rewards were Bio-serv-brand 45-mg food pellets delivered into a recessed trough positioned 2 cm above the chamber floor in the center slot of the right wall. Between two and eight non-retractable levers were made available during any particular session. The layout of the chamber, along with the positional indices of the eight possible lever locations, are depicted in Figure 9.1. Additionally, each chamber had an illuminated house light and a fan, and was enclosed in a sound-attenuating wooden box (75 cm x 61 cm x 38 cm). The operant chambers were collectively controlled by a PC running MedPC-IV software.

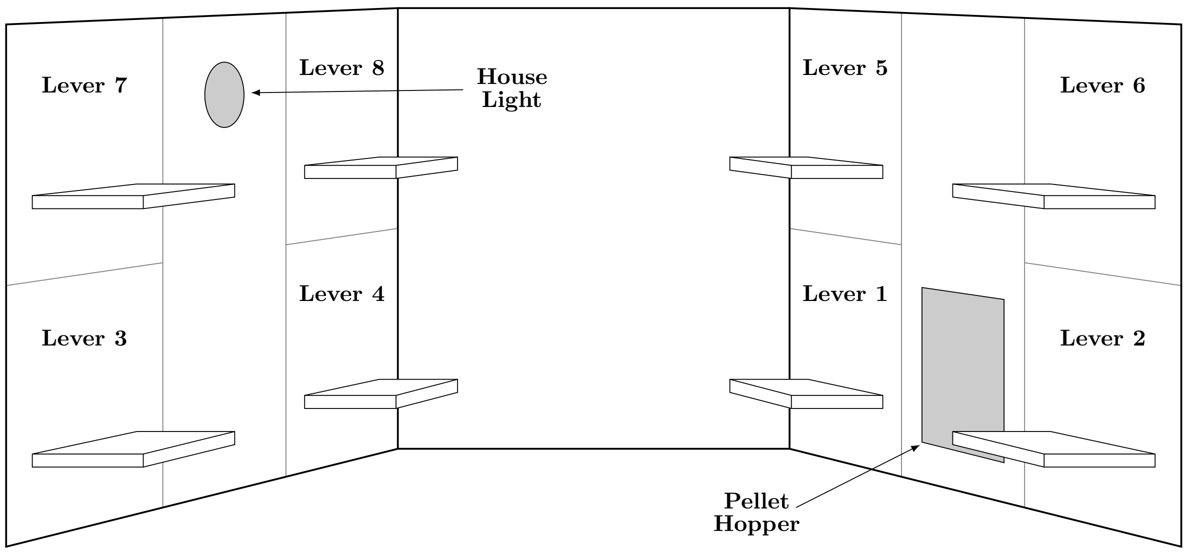

Figure 9.1: Operant chamber layout for the eight-lever condition. Lever were fixed in place and could not retract. In the six-lever phases, Lever 1 and Lever 2 were not available and were replaced with a smooth steel wall segment . In the four-lever phases, Levers 1, 2, 7, and 8 were not available.

9.1.3 Procedure

In order to implement the Turn-Based Foraging paradigm, rewards were scheduled according to a 'dependent concurrent variable ratio' (depVR) schedule. In such a schedule, rewards are made available probabilistically every time the subject responds. However, depVR schedules differ from traditional variable ratio schedules because a response to *any* alternative creates an opportunity for rewards to set up on *every* alternative, and rewards that set up on other alternatives remain available until collected. The schedule is described precisely by the pseudocode presented in Algorithm 1. The probabilities of reward setup are described in Table 9.1.

Each phase consisted of ten sessions, which were run five days a week (Monday to Friday). Sessions run until either 100 rewards had been earned, or until 60 minutes had elapsed, whichever came first.

Phase	Lever 1	Lever 2	Lever 3	Lever 4	Lever 5	Lever 6	Lever 7	Lever 8
1 (Eight Levers)	.0422	.0357	.0617	.0552	.0097	.0065	.0227	.0162
2 (Eight Levers)	.0552	.0162	.0357	.0065	.0617	.0227	.0097	.0422
3 (Eight Levers)	.0162	.0617	.0097	.0357	.0227	.0552	.0422	.0065
4 (Eight Levers)	.0617	.0422	.0065	.0162	.0552	.0097	.0357	.0227
5 (Eight Levers)	.0065	.0097	.0162	.0227	.0357	.0422	.0552	.0617
1 (Four Levers)	–	–	.0178	.0417	.0774	.1131	–	–
2 (Four Levers)	–	–	.1131	.0774	.0417	.0178	–	–
3 (Four Levers)	–	–	.0417	.1131	.0178	.0774	–	–
4 (Four Levers)	–	–	.0774	.0178	.1131	.0417	–	–
1 (Six Levers)	–	–	.0759	.0134	.0580	.0491	.0223	.0313
2 (Six Levers)	–	–	.0134	.0759	.0223	.0580	.0313	.0491
3 (Six Levers)	–	–	.0491	.0223	.0313	.0134	.0759	.0580
4 (Six Levers)	–	–	.0223	.0491	.0759	.0313	.0580	.0134

Table 9.1: Schedule probabilities in Rat Experiment 1

9.2 Results

Over the course of the experiment, each subject made between 60,000 and 64,000 responses across the thirteen different phases. In order to characterize this substantial body of responses, both the molar level of analysis (i.e. phase-level descriptive statistics) and the molecular level (i.e. changes over time) must be considered. Molar effects are considered first, followed by molecular methods.

9.2.1 Molar Analysis

As described in Chapter 5, the barycentric matching model described proportions of behavior in terms of an exponent α, called *sensitivity*, and a compositional vector $\boldsymbol{\kappa}$, called *bias* (see Equation 5.1). To obtain parameter estimates, each subject's responses to each alternative were summed over the last seven sessions in each phase. A regression analysis (Equation 5.2) was then performed using proportions of response ($\mathbf{B}$, for "behavior") as the dependent variable, while the reward setup probabilities, listed in Table 9.1 were used as the independent variable, $\mathbf{R}$.

The data can be represented in a parsimonious fashion using the ILR transformation, but making use of it requires specifying an orthonormal basis $\mathbf{U}$ (Equation 4.6). The molar analysis of Experiment 1 made use of the following basis in the 8-lever case:

$$\mathbf{U}_8 = \begin{bmatrix} \mathbf{u_1} \\ \mathbf{u_2} \\ \mathbf{u_3} \\ \mathbf{u_4} \\ \mathbf{u_5} \\ \mathbf{u_6} \\ \mathbf{u_7} \end{bmatrix} = \begin{bmatrix} 0 & 0 & \sqrt{\frac{2}{8}} & \sqrt{\frac{2}{8}} & -\sqrt{\frac{2}{8}} & -\sqrt{\frac{2}{8}} & 0 & 0 \\ 0 & 0 & \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} & 0 & 0 \\ 0 & 0 & \sqrt{\frac{2}{24}} & \sqrt{\frac{2}{24}} & \sqrt{\frac{2}{24}} & \sqrt{\frac{2}{24}} & -\sqrt{\frac{4}{12}} & -\sqrt{\frac{4}{12}} \\ 0 & 0 & 0 & 0 & 0 & 0 & \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} \\ \sqrt{\frac{6}{16}} & \sqrt{\frac{6}{16}} & -\sqrt{\frac{2}{48}} & -\sqrt{\frac{2}{48}} & -\sqrt{\frac{2}{48}} & -\sqrt{\frac{2}{48}} & -\sqrt{\frac{2}{48}} & -\sqrt{\frac{2}{48}} \\ \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad (9.2)$$

Here, each column refers to a lever, while each row represents an empirical contrast. Row $\mathbf{u_1}$ estimates *only* how [lever 3 & lever 4] contrasts with [lever 5 & lever 6], while row $\mathbf{u_2}$ estimates *only* the contrast between lever 3 and lever 4. Since the 4- and 6-lever cases used subsets of the levers, the orthonormal bases for their transformations are subsets of this matrix: $\mathbf{U}_4$ consists only of $\mathbf{u}_{\mathbf{1:3},3:6}$, i.e. rows 1 through 3 and columns 3 through 6, as these are the three contrasts devoted to levers 3 through 6.

As noted in Chapter 4, there is no uniquely ideal orthonormal basis for analyzing compositions of a given size. It is instead up the analyst to specify the contrasts that are interesting, doing so in a principled manner. Importantly, because the present analysis makes use of the barycentric matching model (Equation 1.12), *any* valid orthonormal basis will yield an identical estimate of α and a precisely similar estimate of $\boldsymbol{\kappa}$ (that is, identical under rotation). Consequently, although the contrasts specified by Equation 9.2 are interesting with respect to some of the obvious questions regarding the factors contributing to relative bias, they are also in a sense arbitrary, and do not bias the resulting analysis. Note that equivalent parameter estimates would be obtained using CLR-transformed data as well.

Forming a complete picture of behavior at the molar level requires understanding both the bias and sensitivity parameters. The sensitivity parameter is considered first, followed by bias.

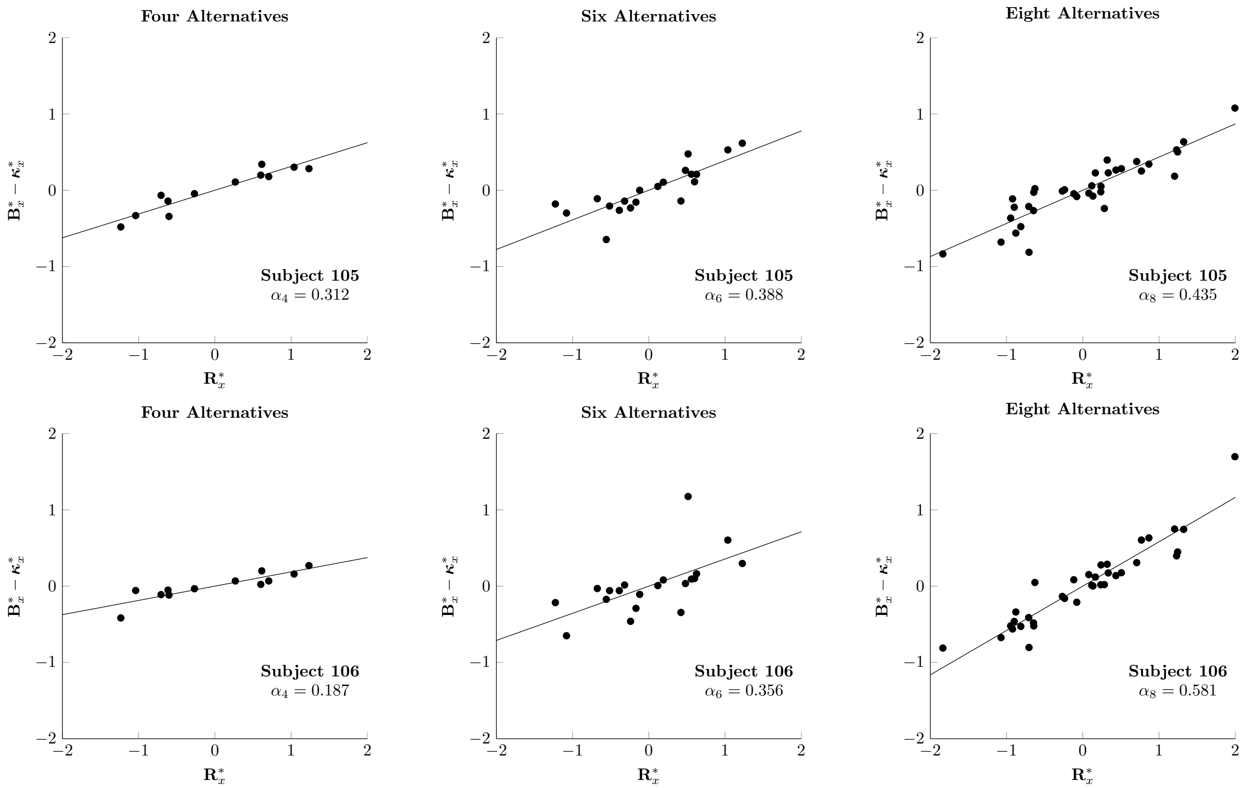

Figure 9.2: Plots of the sensitivity parameters for Subjects 105 and 106 in Experiment 1. Bias parameters κ and sensitivity parameter α relating the CLR-transformed programmed probabilities in Table 9.1 to the CLR-transformed proportions of responding summed over the last seven sessions were estimated using Equation 5.2. The bias parameters were then factored out to provide a visual sense of the sensitivity parameters. Sensitivity appears to be lower (i.e. a shallower slope is observed) when there are fewer alternatives than when there are more alternatives.

9.2.1.1 Sensitivity

Because κ acts as a normalizing factor, a plot with $(\mathbf{B}^* - \boldsymbol{\kappa}^*)$ on the y-axis and $\mathbf{R}^*$ on the x-axis is expected to be linear and to pass through the origin. This provides a straightforward way to visually compare the α parameter in each condition (where it corresponds to the slope of the line), as well as to visually assess the variability accounted for. Figure 9.2 does this for Subjects 105 and 106, whose α parameters were typical.

For both subjects 105 and 106, estimates of sensitivity appear to increase as a function of the number of alternatives (that is, $\alpha_4 < \alpha_6 < \alpha_8$). This pattern generally persists for the other eight

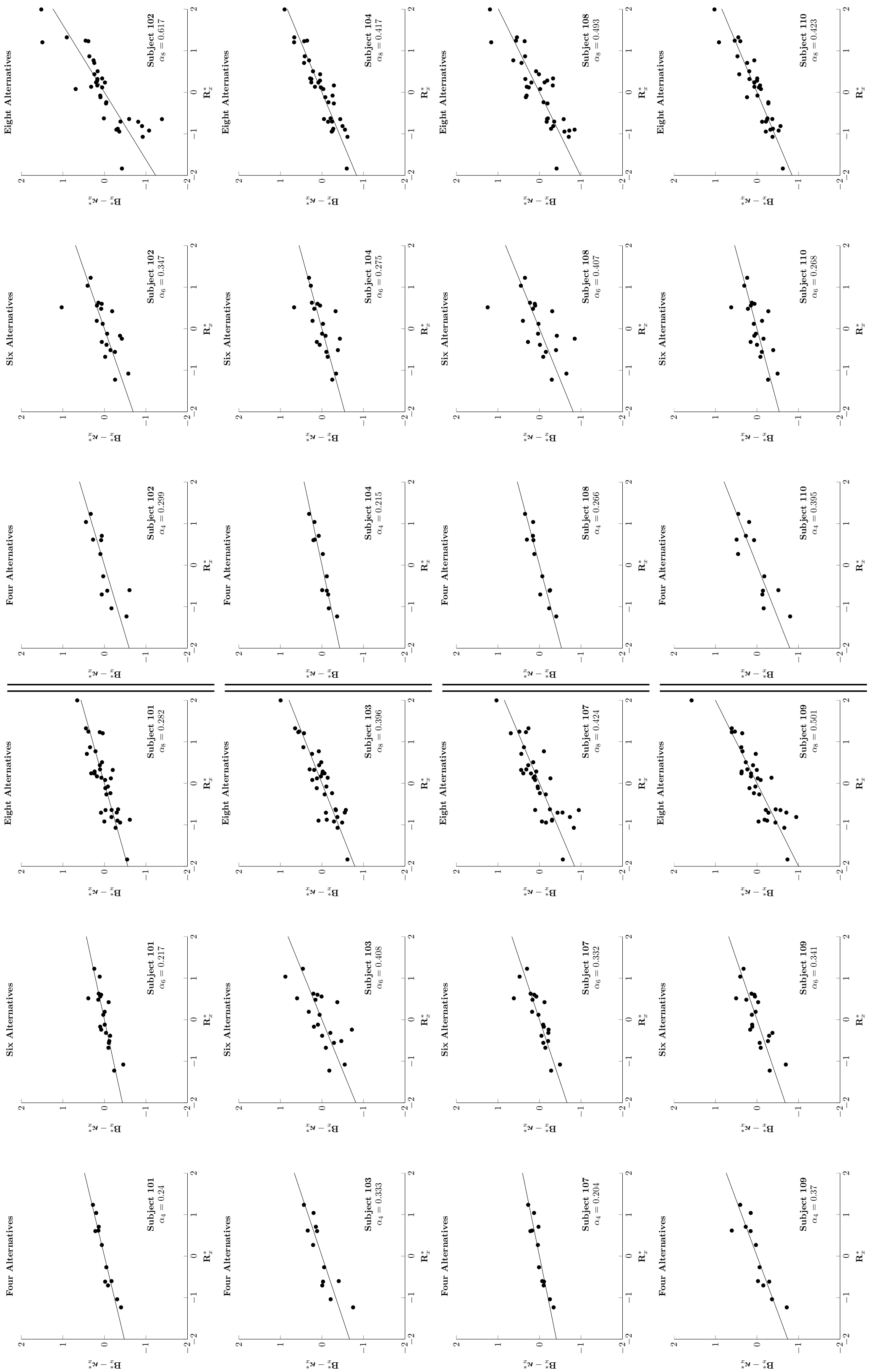

Figure 9.3: Plots of the sensitivity parameters for all other subjects, plotted in the same manner as in Figure 9.2.

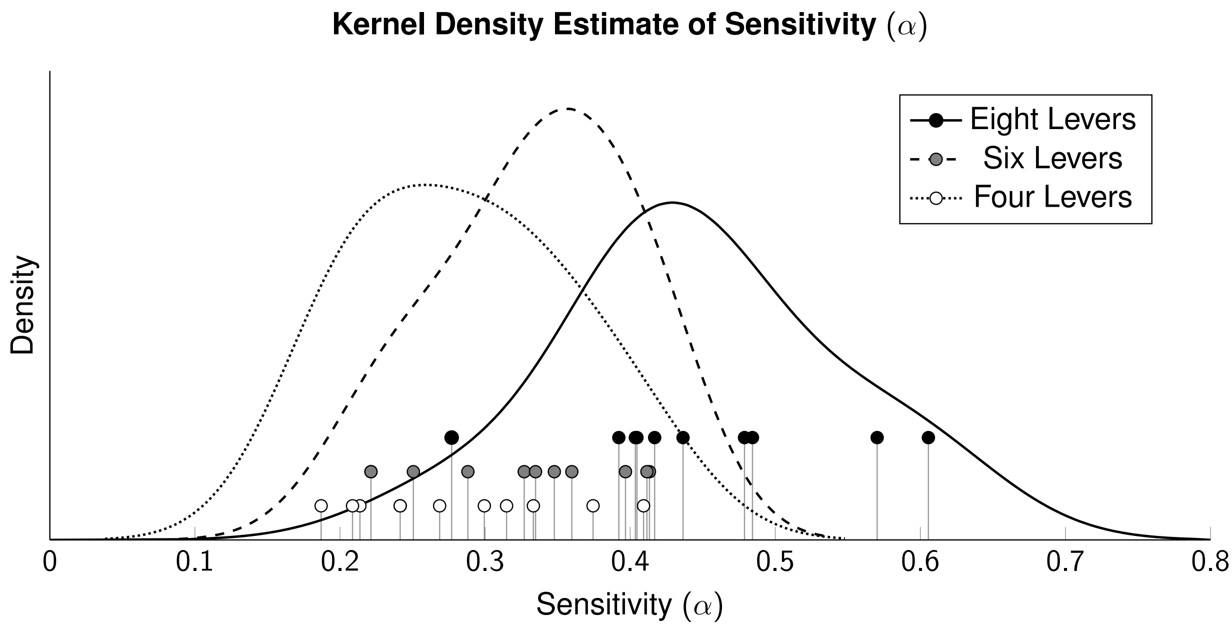

Figure 9.4: Kernel density estimates (KDE) of distributions of sensitivity parameters α for all subjects in all conditions. Markedly higher sensitivity was seen in the eight-lever condition than in the four- and six-lever conditions. The KDE used a Gaussian kernel whose bandwidth was set using the rule of thumb specified by Silverman (1986).

subjects, whose sensitivity is plotted in Figure 9.3.

The distributions of the sensitivity parameter across subjects are estimated in Figure 9.4 using a kernel density estimate (or 'KDE', Silverman, 1986). The population density is obtained by assigning a Gaussian distribution to each parameter estimate for each subject, whose standard deviation (the 'bandwidth,' in the parlance of KDE distribution) was determined using Silverman's rule of thumb:

$$\sigma_{\text{KDE}} = \left(\frac{4}{3n}\right)^{\frac{1}{5}} \hat{\sigma} \tag{9.3}$$

Here, the kernel σ_{KDE} is based on the estimated standard deviation for the sample, $\hat{\sigma}$. This rule of thumb is optimal if the underlying data are Gaussian, and is fairly robust in other cases. (see Silverman, 1986, for details).

Based on Figure 9.4, there appears to be more overlap between the 4-lever and 6-lever conditions than between the 8-lever condition. A non-parametric two-way analysis of variance (commonly identified as the Friedman test; see Quade, 1984) found a significant difference among the sample means ($\chi^2(2) = 13.4, p < .002$). However, post-hoc multiple comparisons test (Rhyne & Steel, 1965) only confirm a significant difference between the 4-lever and 8-lever conditions.

Because the KDE in Figure 9.4 suggests the data were *approximately* normal, a repeated-measures analysis of variance was also performed. This parametric analysis yielded a highly significant effect ($F(2, 106) > 16.2, p < .001$), and Holm-Šidák-corrected[1] t-tests yielded significant differences between the 8-lever condition and both the 4-lever condition ($p < .02$) and the 6-lever condition ($p < .03$). The 4- and 6-lever conditions did not differ significantly ($p = .12$).

9.2.1.2 Bias

Figure 9.5 plots the bias parameters estimated for each lever using color-coded box-and-whisker diagrams. Note that although Equation 5.1 was used to obtain parameter estimates (which relied on the ILR transformation), Figure 9.5 instead uses the CLR-transformed parameters $\boldsymbol{\kappa}^{\circ}$, to make the per-lever effects clear (as opposed to using the ILR contrasts, whose interpretation requires that the reader recall exactly which levers are being contrasted with which for every parameter). Translation between CLR and ILR representations can be achieved without any loss of information using Equation 4.5, so the two representations are equivalent.

Cursory examination of Figure 9.5 suggests that subjects displayed less bias in their responding in the 8-lever case (the bias parameters associated with each lever appear generally closer to the origin in that condition). However, whereas comparing sensitivity across conditions is relatively straightforward, comparing bias is somewhat less so because of the differing dimensionality of the conditions. A comparison of κ_{L3} in the 4-, 6-, and 8-lever conditions is inappropriate because of the closure constraint: $\boldsymbol{\kappa}$ is a composition, so κ_{L3} is a *relative* measure, and its interpretation changes as a function of what choice alternatives it is relative *to*.

This difficulty can be overcome by instead performing an analysis of $\boldsymbol{\kappa}^*$. The ILR transformation converts the composition $\boldsymbol{\kappa}$ into a set of orthogonal contrasts that isolate which operanda are being compared. For example, in the orthonormal basis $\mathbf{U}$ defined by Equation 9.2, the contrast $\mathbf{u_2}$ contrasts lever 3 and lever 4, and only those levers, regardless of how many other alternatives were included in that phase of the experiment.

[1] This is a correction for multiple comparisons that is uniformly more powerful than the commonly-used Bonferonni correction. See Ludbrook (1998) for details.

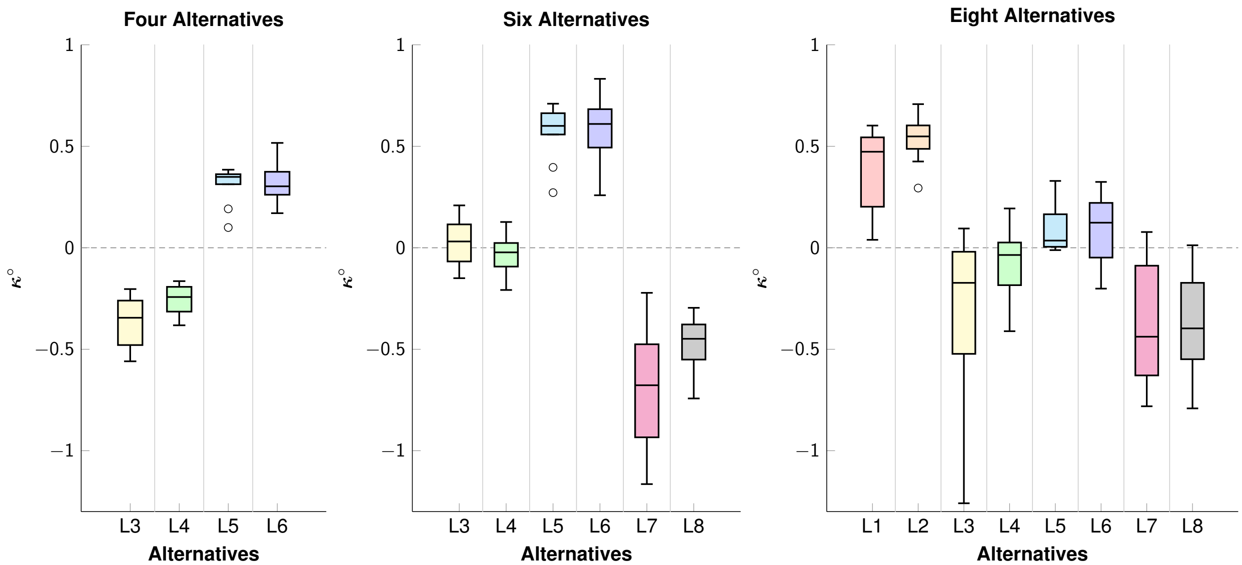

Figure 9.5: Box-and-whisker plots of the distributions of CLR-transformed bias parameters κ° for all subjects in each condition.

Because $\mathbf{u_6}$ and $\mathbf{u_7}$ are only present in the 8-lever condition, they have no basis for comparison. For the rest, comparisons were made using a separate repeated-measures analysis of variance of the absolute deviation for each factor. Significant effects for condition were only observed for $\mathbf{u_1}$, comparing levers 3 & 4 to levers 5 & 6 ($F(2, 18) > 10.8, p < .001$) and $\mathbf{u_4}$, comparing levers 3, 4, 5, & 6 to levers 7 & 8 ($F(1, 9) > 10.8, p < .001$). In post-hoc Holm-Šidák comparisons, the 8-lever condition is significantly lower than the other conditions in both cases, while the remaining conditions do not differ from one another. Table 9.2 shows the mean absolute contrast for each condition, as well as the standard error estimated from the repeated-measures ANOVAs.

9.2.1.3 Summary

These analyses paint an unexpected picture: Subjects appear to be *more* sensitive to the schedule of rewards in the 8-lever condition than in either the 4- or 6-lever conditions, and also appear to be *less* biased. Thus, by both metrics provided by generalized matching, subjects were closer to optimal performance when the task was at its most complex. Some trends pointed toward the 6-lever conditions engendering more optimal performance than the 4-lever condition, but none of these were significant.

Contrast	Levers Compared	4 Levers	6 Levers	8 Levers	Standard Error
$\mathbf{u_1}$	(3,4) vs. (5,6)	0.6318	0.5788	0.3281	0.0492
$\mathbf{u_2}$	(3) vs. (4)	0.1016	0.0523	0.1862	0.0442
$\mathbf{u_3}$	(5) vs. (6)	0.0770	0.1055	0.0739	0.0170
$\mathbf{u_4}$	(3,4,5,6) vs. (7,8)	N/A	0.9939	0.4598	0.0860
$\mathbf{u_5}$	(7) vs. (8)	N/A	0.2593	0.1479	0.0496

Table 9.2: Mean ILR-Transformed Bias Contrasts (Absolute) in Rat Experiment 1. The 'Levers Compared' column shows which alternatives are included in each contrast.

In order to examine the behavioral processes that gave rise to these parameters, an extensive molecular analysis follows.

9.2.2 Molecular Analysis

While an analysis of steady-state patterns of responding show unambiguously that subjects adapted their behavior to changes in the schedule (as evidenced by sensitivity parameters greater than zero), these molar descriptions do little to illuminate what gives rise to these changes in behavior. This approach also, by design, ignores data immediately following a transition to a new schedule. To obtain a better understanding of the process by which decisions evolve over time, a time-series (or 'molecular') analysis was undertaken.

The primary tool used for the analysis was a form of change-point analysis called the CPR algorithm (Jensen, 2014 (projected)). This method, which is summarized in Appendix B, assumes that data arise stochastically from an underlying distribution, and that this underlying distribution may change abruptly an unknown number of times. These 'change-points' are identified using a Bayesian algorithm that weighs the evidence in favor of a more complex model that includes a change vs. the evidence favoring a less complex model that lacks a change. This process unfolds recursively, determining whether a change-point is justified in a segment of data, dividing the data at the most appropriate point if a change is detected, and repeating the process on each resulting subsegment. This process continues until the evidence is deemed to support no further changes. An important feature of this approach is that it automatically corrects for model complexity, subdividing the data only a much as the evidence supports. This, in turn, means that each segment

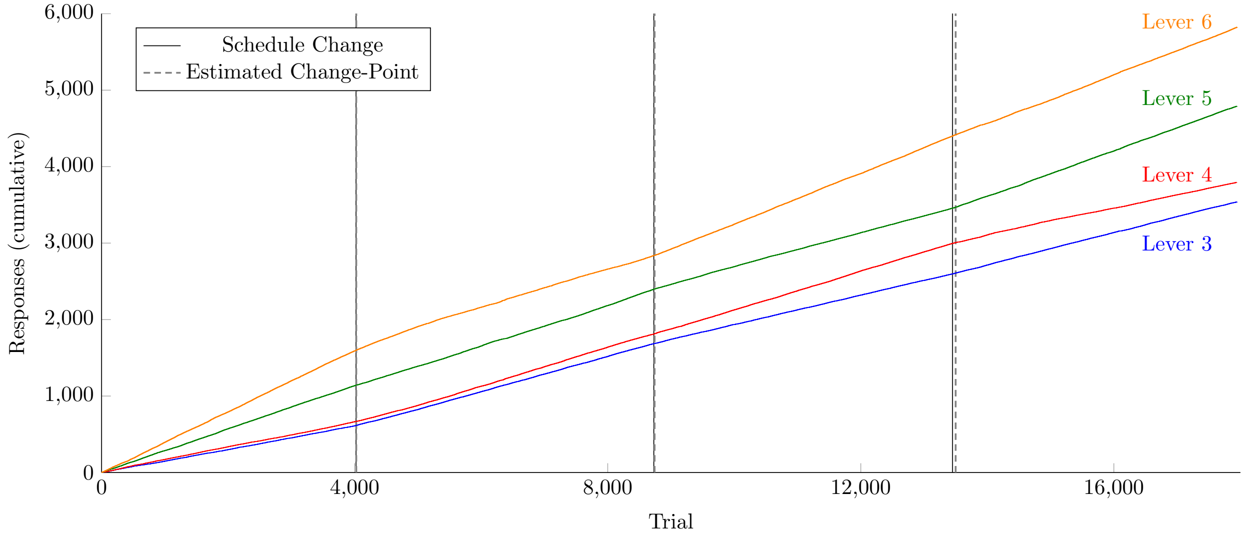

Figure 9.6: Cumulative record of responses for Subject 104 in the four-alternative condition. A change-point analysis using a multinomial distribution was performed to determine when proportions of response changed. Vertical black lines indicate when the schedule of rewards changed, and the dashed lines indicate when a change-point was detected.

can be treated as having been sampled from a steady-state process. The CPR algorithm is a general approach that has been developed for many different distributions. In the present case, we assume that the data in each steady-state segment is sampled randomly from a multinomial distribution whose parameters we would like to estimate.

If the CPR algorithm is applied to a subject's entire response history, relatively few change-points arise. For example, Figure 9.6 depicts the cumulative record of responses made by Subject 104 in the 4-lever condition. In this case, exactly one change-point (dashed lines) is identified for each change in the schedule (solid lines), often following so closely on the heels of the schedule change that the two nearly overlap. The substantial bias favoring levers 5 and 6 over levers 3 and 4 are evident in their higher overall rates of selection, and the relatively shallow inflections at each change-point are reflective of Subject 104's low sensitivity in this condition. This result is typical, but it is also misleading, because of the assumption that responses are sampled stochastically. An examination of the conditional probabilities underlying the overall behavior reveals that behavior is instead highly structured.

There are two substantial sources of structure in responding that render a simple multino-

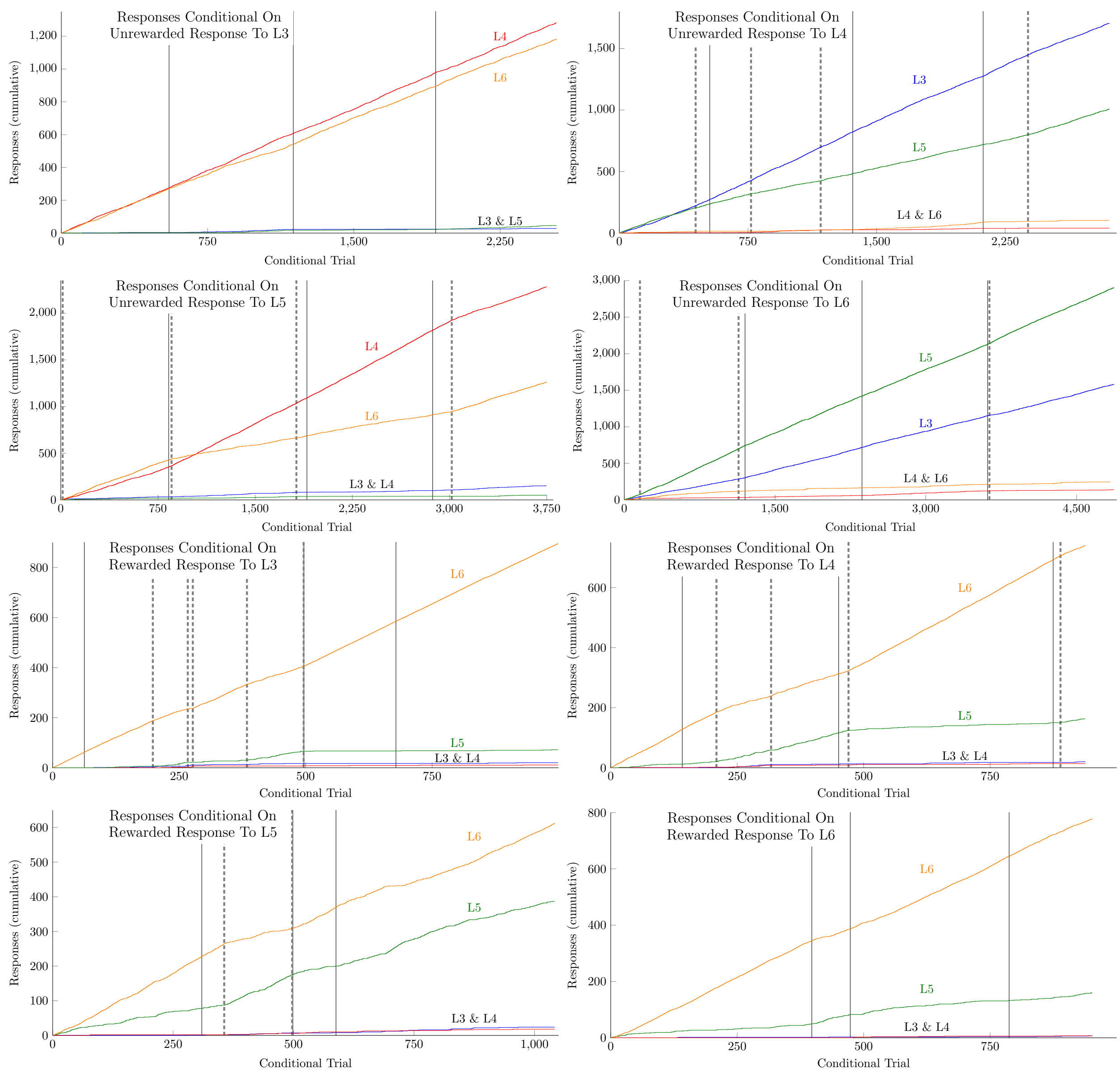

Figure 9.7: Cumulative record of conditional responses for Subject 104 in the four-alternative condition. The top row consists of responses following an unrewarded trial, while the bottom row consists of responses following a rewarded trial. Vertical black lines indicate when the schedule of rewards changed, and the dashed lines indicate when a change-point was detected.

mial model invalid. The first is that the conditional probabilities of selecting each operandum given the prior choice deviate substantially from the stochastic expectation. The second is that responses following the delivery of a reward differ substantially from those following unrewarded trials. These criteria split Subject 104's approximately 18,000 responses in the 4-lever condition into eight categories. The cumulative plots and corresponding change-point analyses that result from this subdivision are depicted in Figure 9.7.

As these plots reveal, Subject 104's choices were substantially contingent on the preceding circumstance. Following an unrewarded trial, the next response was often to the lever counter-clockwise from the previous lever. Following a rewarded trial, responses were most commonly made to Lever 6. Figure 9.7 also reveals that some conditional relationships changed more frequently than others. For example, the choice made following an unrewarded response to lever 3 was invariant across schedules, whereas responses conditional on lever 4 changed frequently.

The resulting behavior can be described by two first-order Markov chains (i.e. two conditional probability tables), $\mathcal{B}_0$ (for responses following non-rewarded trials) and $\mathcal{B}_1$ (for responses following rewarded trials):

$$\mathcal{B}_0 = \begin{bmatrix} \mathcal{C}(\mathbf{B}|\neg R_3) \\ \mathcal{C}(\mathbf{B}|\neg R_4) \\ \mathcal{C}(\mathbf{B}|\neg R_5) \\ \mathcal{C}(\mathbf{B}|\neg R_6) \end{bmatrix} = \begin{bmatrix} \Pr(B_3|\neg R_3) & \Pr(B_4|\neg R_3) & \Pr(B_5|\neg R_3) & \Pr(B_6|\neg R_3) \\ \Pr(B_3|\neg R_4) & \Pr(B_4|\neg R_4) & \Pr(B_5|\neg R_4) & \Pr(B_6|\neg R_4) \\ \Pr(B_3|\neg R_5) & \Pr(B_4|\neg R_5) & \Pr(B_5|\neg R_5) & \Pr(B_6|\neg R_5) \\ \Pr(B_3|\neg R_6) & \Pr(B_4|\neg R_6) & \Pr(B_5|\neg R_6) & \Pr(B_6|\neg R_6) \end{bmatrix}$$

$$\mathcal{B}_1 = \begin{bmatrix} \mathcal{C}(\mathbf{B}|R_3) \\ \mathcal{C}(\mathbf{B}|R_4) \\ \mathcal{C}(\mathbf{B}|R_5) \\ \mathcal{C}(\mathbf{B}|R_6) \end{bmatrix} = \begin{bmatrix} \Pr(B_3|R_3) & \Pr(B_4|R_3) & \Pr(B_5|R_3) & \Pr(B_6|R_3) \\ \Pr(B_3|R_4) & \Pr(B_4|R_4) & \Pr(B_5|R_4) & \Pr(B_6|R_4) \\ \Pr(B_3|R_5) & \Pr(B_4|R_5) & \Pr(B_5|R_5) & \Pr(B_6|R_5) \\ \Pr(B_3|R_6) & \Pr(B_4|R_6) & \Pr(B_5|R_6) & \Pr(B_6|R_6) \end{bmatrix}$$

Here, R_i is used as a shorthand for "reward earned as a result of a response to alternative i", while $\neg R_i$ denotes "no reward earned as a result of a response to alternative i." This formulation also draws attention to the fact that each row in the table is a composition whose contents are independent of the other rows.

At each point in Figure 9.7 when a change-point is identified, *only* the components in the corresponding row change. The conditional probabilities themselves are estimated by the CPR algorithm merely by computing the sum of the observations in each category, adding 0.5 to each

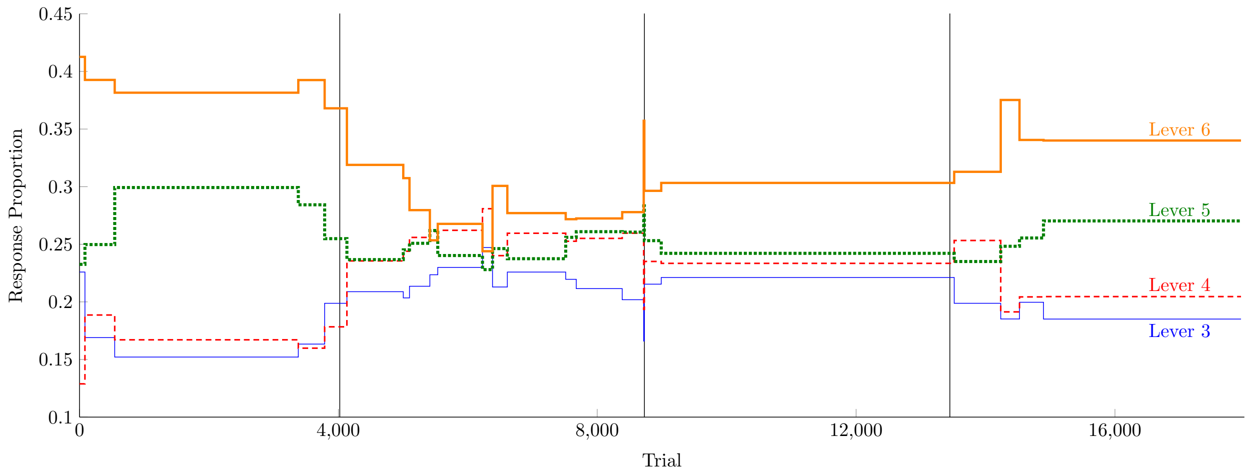

Figure 9.8: Estimated stationary response proportions for Subject 104 in the four-alternative condition, given the conditional matrices for unrewarded responding ($\mathcal{B}_0$) and rewarded responding ($\mathcal{B}_1$).

score[2], and then applying the closure operation.

A model of behavior consisting of Markov chains is straightforward to simulate. Figure 9.8 shows the estimated overall response rate, taking both rewarded and unrewarded behavior into account. These rates effectively provide a principled estimate of the slopes plotted in Figure 9.6, while also displaying sensitivity to transitions in behavior too subtle for the CPR algorithm to detect at the molar level. For example, although the transition from the first to the second schedule at around trial 4,000 engendered an immediate shift in behavior, a period of further calibration is evident over the next few thousand responses.

Furthermore, given any contingency table for which all states are positive recurrent, the *stationary distribution* (i.e the overall molar rate of occurrence for each outcome) is defined at the limit as follows:

$$\pi_j = \lim_{n\to\infty} p_{i,j}^{(n)} \qquad \text{where } p_{i,j}^{(n)} = \sum_{r\in D} p_{i,r}^{(k)} \cdot p_{r,j}^{(n-k)} \tag{9.4}$$

The stationary distribution may be calculated arithmetically using the following linear relationship,

[2]As in any Bayesian method, the CPR algorithm requires that the analyst specify the prior distributions for each parameter. Adding 0.5 to each score corresponds to the *Jeffreys prior* (Jeffreys, 1946), which has the desirable property that it is 'minimally informative,' meaning that its biasing influence on the posterior distribution for the parameters is the smallest possible. See Jensen (2014 (projected)) for further details.

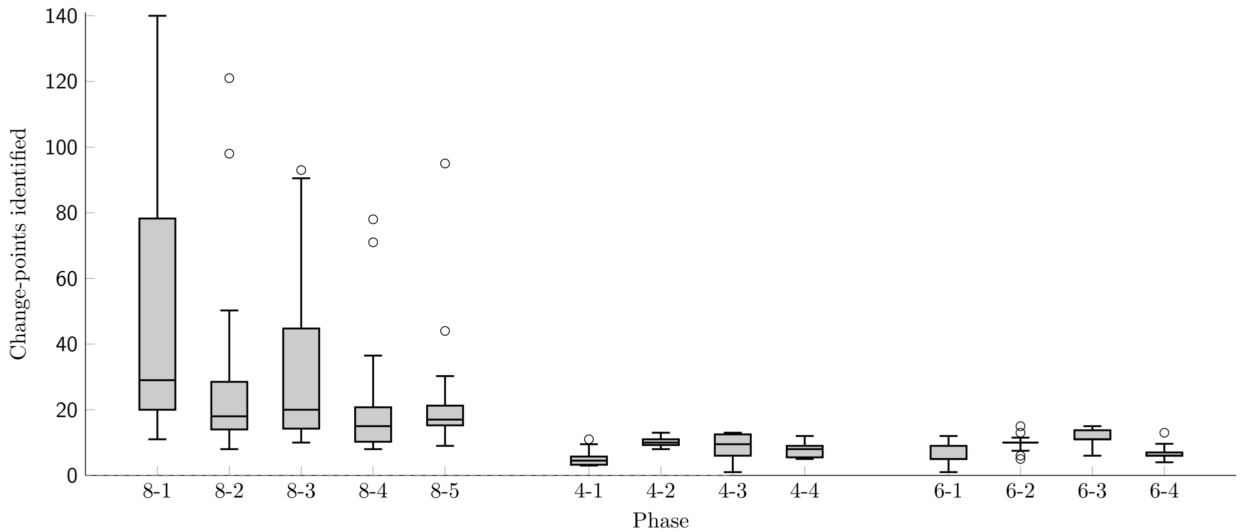

Figure 9.9: Number of change-points detected per phase in Experiment 1.

thanks to the reduced degrees of freedom implied by the closure constraint:

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathcal{B}$$

Since the vector $\boldsymbol{\pi}$ can be solved arithmetically, it can be used as the basis for applying the tools of molar analysis previously described to much narrower windows of time.

These analyses do not speak to the degree to which this approach to describing the behavior is consistent across animals. A simple summary statistic to examine is the number of change-points detected in each phase of the experiment, plotted in Figure 9.9. This plot reveals that, very consistently, subjects made detectable changes to their behavior frequently during the early phases of the experiment (i.e. during the eight-lever phases), but made changes much less frequently in the later (four- and six-lever) phases.

Performing the appropriate statistical test of these differences is complicated by (1) the clear changes in variance from one phase to the next and (2) the nested characteristics of the data (in which phase was subordinate to number of levers). To accommodate these complications, the 130 values were rank-transformed (Conover & Iman, 1981) and subjected to a mixed-model nested ANOVA, in which lever count and phase were treated as fixed effects, phase was nested within lever count, subject ID was treated as a random effect, and subject ID and lever count were permitted to interact. The fixed effects were significant, with the effect of lever count being most pronounced

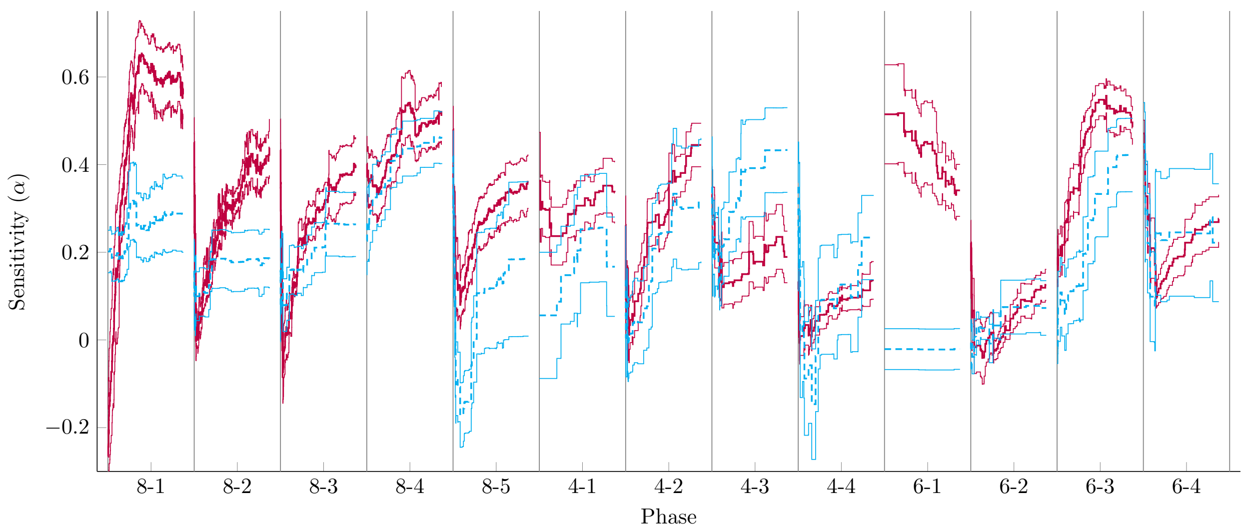

Figure 9.10: Mean sensitivity parameter across subjects, estimated on a trial-by-trial basis assuming consistent bias for each condition.

($F(2, 18) > 40.8, p < .0001$), followed by the effect of phase ($F(10, 90) > 5.4, p < .0001$). Although a significant interaction was seen between subject ID and lever count ($F(18, 90) > 2.8, p < .001$), there was not a significant effect for subject ID ($F(9, 18.06) = 0.95, p = .51$).

9.2.2.1 Group Averaging

If $\mathcal{B}_0$ and $\mathcal{B}_1$ provide an adequate description of behavior at a given point in time, being able to compute the stationary distribution of a behavior allows bias-corrected sensitivity estimates (of the sort plotted in Figure 9.2) to be obtained at any time point. Because this 'trial-wise' estimate of sensitivity does not require summing responses across multiple sessions, it allows an assessment of the refinement of behavior over time.

Figure 9.10 plots the mean sensitivity across subjects of responses following unrewarded trials in red and rewarded trials in blue (thin lines represent one standard error), tracking the first 3,500 trials in each condition. These suggest that, in general, subjects adapted their behavior over the course of multiple sessions. Being group means, these 'learning curves' are not representative of individual performance (which tend consist of much more abrupt transitions). Nevertheless, they suggest that, in general, adjustments are made throughout each phase.

The benefits of these adjustments are generally quite small, as measured by rewards obtained.

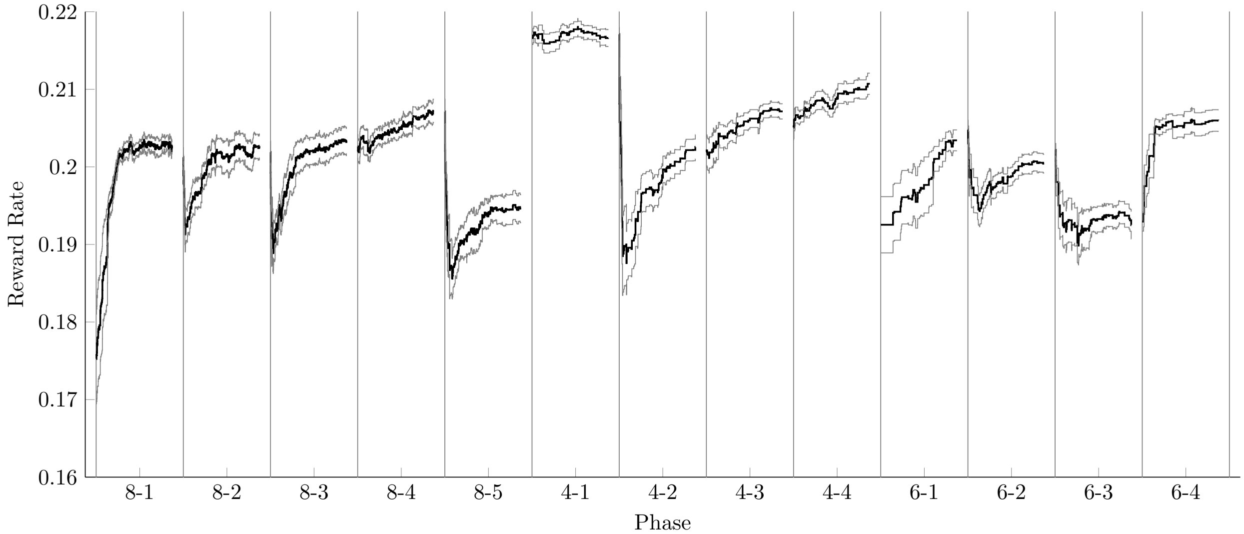

Figure 9.11: Mean reward rate across subjects, based on simulations using the programmed reward probabilities and the response models described by each subjects conditional response matrices.

Figure 9.11 plots the mean reward rate across subjects, as determined from simulations of behavior using $\mathcal{B}_0$ and $\mathcal{B}_1$ (with gray lines indicating one standard error). Across subjects and conditions, the range of reward rates is narrow, nearly always falling between 0.19 and 0.21 rewards per response. Following a period of adjustment to a schedule change, the reward rate tended to stabilize, even as further refinements were made to the patterns of response (as indicated by the changing sensitivity).

9.2.2.2 Summary

Despite providing an excellent fit for proportions of behavior summed over time, the molar analyses mask response structure that can be observed by considering first-order conditional probabilities. When examined in these terms, patterns of behavior changed as a function of the preceding response, as well as the receipt of a reward.

A change-point analysis was used to determine when the conditional probabilities of responding changed, yielding the stationary Markov processes $\mathcal{B}_0$ and $\mathcal{B}_1$ as models of behavior following unrewarded trials and rewarded trials, respectively. Since the stationary distribution of a positive recurrent Markov process can be solved for arithmetically, these matrices enable the tools of molar analysis to be deployed over short time scales.

Analyses of the estimated sensitivity and reward rate over the course of time series suggest that behavior underwent ongoing refinement. However, reward rates differed only slightly over this period, suggesting perhaps that some other factor was driving the refinement of behavior.

9.2.3 Information-Theoretic Analysis

Individual differences in choice behavior can make it difficult to describe in general terms, much less to theorize about. This is particularly true when evaluating conditional probabilities, which are already awkward to visualize with four alternatives and become exponentially more so as alternatives are added.

While sensitivity and bias provide descriptions of overall patterns of behavior, they do at the expense of an understanding of the *process* by which behavior emerges. This motivates an additional analysis described below, in which measures from information theory were used to provide a summary description of behavioral complexity.

9.2.3.1 Entropy Rate

Knowing both the conditional probabilities and the stationary distribution enables the calculation of the *entropy rate*, which provides a measure of how much additional information, on average, is transmitted by the behavior with each response (McMillan, 1953). This can be computed from the model with the following equation:

$$H\left(\mathcal{B}\right) = -\sum_{i,j \in D} \pi_i \cdot \Pr\left(i|j\right) \cdot \log_2\left(\Pr\left(i|j\right)\right) \tag{9.5}$$

Here, the information (or entropy) of each response as measured in bits (due to the use of a base-2 logarithm) is computed by calculating the entropy of each row, multiplying it by the stationary distribution.

The entropy rate of a complex phenomenon provides a good summary of the cognitive load associated with its processing. For example, the entropy rate of written English has been studied both at the level of individual symbols (between 1.0 and 1.5 bits per character B. Schneider, 1996) and word order (about 5.7 bits per word Montemurro & Zanette, 2011). Surprisingly, the entropy rates of different natural languages appear similar, even when employing dramatically different alphabets and grammar (M. A. Changizi, 2011). This surprising consistency has been present as evidence of a basic cognitive limit in human language processing (Pellegrino et al., 2011). Similar

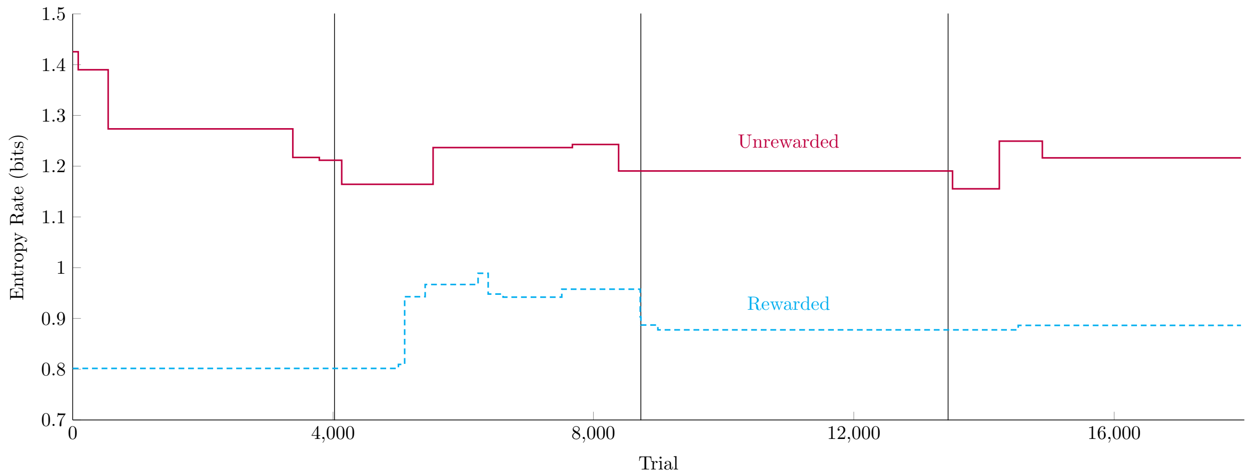

Figure 9.12: Entropy rate of conditional responding for Subject 104 in the four-alternative condition.

arguments have been put forth for vision (M. A. Changizi, 2008) and music processing (Pearce & Wiggins, 2004).

The entropy rate of Subject 104's 4-lever responding (as estimated from the model consisting of $\mathcal{B}_0$ and $\mathcal{B}_1$) is presented in Figure 9.12. This shows that although the overall frequencies of responding changed as a function of the schedule (sometimes being close to equiprobable, and other times clearly favoring one alternative over the others), the entropy rate of responding was generally uniform. Furthermore, responses following reward delivery ($\mathcal{B}_1$) had a lower entropy rate than the rest ($\mathcal{B}_0$), a pattern consistent with greater bias observed immediately after reward delivery.

Analyzed across subjects and across conditions, the entropy rates clearly differ as a function of the number of response alternatives. This was confirmed by rank-transforming each subject's mean entropy rate per phase (Conover & Iman, 1981) and performing a mixed-model ANOVA with phase nested within lever count, subject ID treated as a random variable, and subject ID allowed to interact with lever count. Entropy rates differed significantly as a function of lever counts in both unrewarded trials ($F(2, 18) > 223.4, p < .0001$) and rewarded trials ($F(2, 18) > 12.6, p < .0005$). An effect of phase was also significant in the unrewarded trials ($F(10, 90) > 4.4, p < .0001$), but not in the rewarded trials ($F(10, 90) = 1.11, p = .36$). This effect appears to be driven by a gradual decrease over the course of the 8-lever phase; the entropy rate appears otherwise stable over the course of the 4- and 6-lever conditions that follow. Throughout, the entropy rate associated with

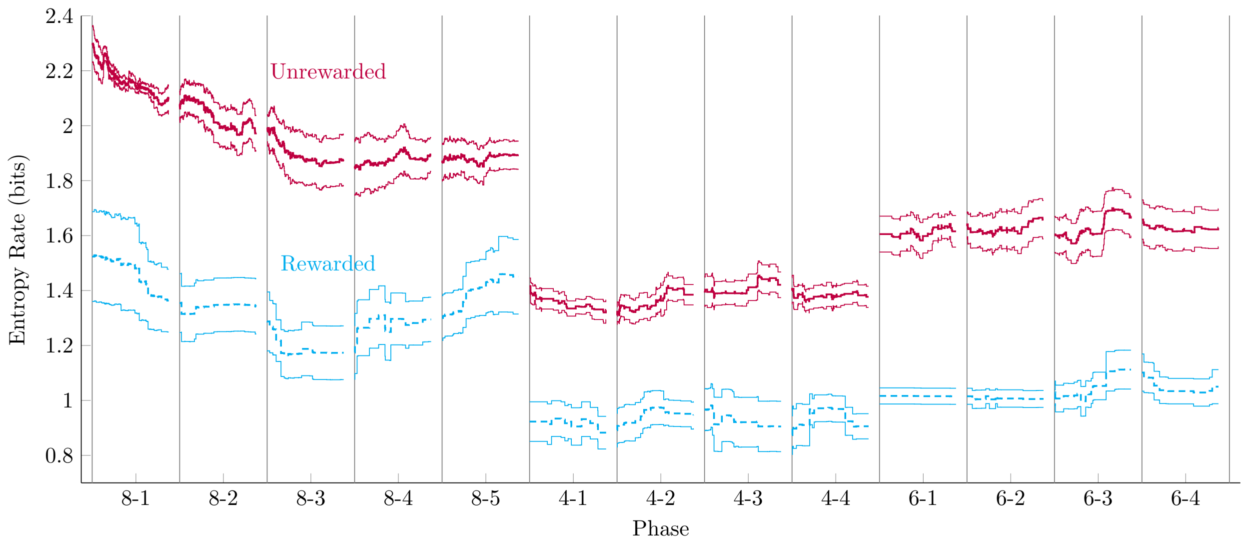

Figure 9.13: Mean entropy rate across subjects, estimated based on the response models described by each subjects conditional response matrices.

responses following rewards is lower than otherwise, consistent with the bias toward responses near the food-delivery trough.

9.2.3.2 Divergence Rate

A distinction must be made between the *complexity* of a behavior and the *cognitive load* that it demands of the organism performing the behavior. Because the experimental paradigm requires that the subject displace itself in the operant chamber, each conditional probability reflects a a choice made under different circumstances. This can yield *conditional biases* in behavior. For example, because Lever 5 is positioned directly above Lever 1, it is plausible that a subject might be much more likely to display a Lever 5 $\rightarrow$ Lever 1 transition than would be predicted from the overall biases for Lever 5 and Lever 1 alone. This yields an increase in the informational complexity of the behavior without any increased information processing on the part of the subject. Consequently, it is essential to tease apart observed complexity that arises from a subject-environment interaction from information that must be encoded and processed by the subject. Entropy rate provides a measure of the former, but not of the latter.

A simple solution to this problem is to measure the entropy of each subject's 'divergence' from a baseline behavior. This can be accomplished using a metric called the Kullback-Leibler divergence

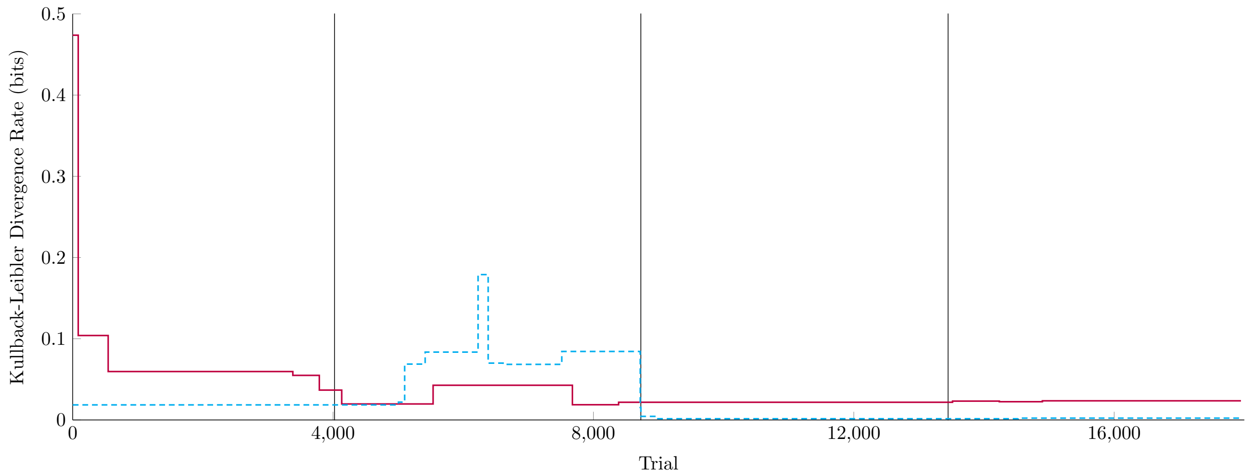

Figure 9.14: Kullback-Leibler divergence rate of conditional responding for Subject 104 in the four-alternative condition.

(or KLD) (Kullback & Leibler, 1951):

$$KLD\left(P\|Q\right) = \sum_{i} P\left(i\right) \cdot \log_2 \left(\frac{P\left(i\right)}{Q\left(i\right)}\right) \tag{9.6}$$

Here, P is an observed frequency distribution, and Q is a theoretical distribution. Both are presumed to have the same number of symbols, but those symbols do not necessarily appear with equal frequency. $KLD\left(P\|Q\right)$ measures the number of additional bits of information needed to transmit a signal using the symbol set P that was originally composed using Q.

As an example, consider the 26 letters in the English language. These letters appear with a particular distribution across the broad corpus of written English, in which E is the most common letter (occurring 12.49% of the time) and Z is the least common (occurring 0.09% of the time), according to a recent exhaustive analysis of the 743.8 billion words in the Google Books database (Norvig, 2013). Let L_{Eng} be this theoretical distribution. Next, suppose that an artist wishes to reproduce large passages of English using tiles from the popular board game Scrabble (Butts, 1938). Although Scrabble tiles have a similar distribution of letters to English, it differs appreciably in some respects. For example, the letter Z appears in the Scrabble letter set 1.02% of the time, over 100 times the rate it appears in long-form written English. Let L_{Scr} be the distribution of tiles observed in Scrabble.

Because L_{Scr} and L_{Eng} are mismatched, the artist in this example will discover that buying

Scrabble board games in bulk will yield a surplus of some letters and a shortage of others. The resulting 'wastage' of tiles is the cost of translating a signal composed using L_{Eng} into a message composed using L_{Scr}. By measuring $KLD\left(L_{Scr} \| L_{Eng}\right)$, we can obtain a measure (in bits) of how large this wastage is. Based on Equation 9.6, this amounts to 0.129 bits per character.

Equation 9.6 does not, however, take into account the response structure as a continuous measure over time. In order to identify the *divergence rate* over time, Equation 9.5 was modified in the following fashion:

$$KLD_T\left(\mathcal{B}\right) = \sum_{i,j \in D} \pi_i \cdot \mathcal{B}_{(i|j)} \cdot \log_2\left(\frac{\mathcal{B}_{(i|j)}}{\mathcal{Q}_{(i|j)}}\right) \quad \text{where } \mathcal{Q}_{\mathbf{i}} = \mathcal{C}\left(\prod_{t=1}^{T}\left(\mathcal{B}_{\mathbf{i}|t}\right)^{-T}\right) \tag{9.7}$$

Here, $\mathcal{Q}$ is the geometric mean of the conditional probability matrix $\mathcal{B}$ at each time t in the time series. For example, the geometric means of all transition probabilities derived from Figure 9.7 yield[3] the following matrices $\mathcal{Q}_0$ and $\mathcal{Q}_1$ for Subject 104's unrewarded and rewarded trials:

$$\mathcal{Q}_0 = \begin{bmatrix} 0.0112 & 0.5047 & 0.0183 & 0.4658 \\ 0.6307 & 0.0085 & 0.3617 & 0.0260 \\ 0.0374 & 0.6177 & 0.0098 & 0.3351 \\ 0.3290 & 0.0248 & 0.6022 & 0.0440 \end{bmatrix} \quad \mathcal{Q}_1 = \begin{bmatrix} 0.0135 & 0.0061 & 0.0288 & 0.9517 \\ 0.0202 & 0.0144 & 0.1511 & 0.8144 \\ 0.0227 & 0.0183 & 0.3943 & 0.5646 \\ 0.0079 & 0.0079 & 0.1682 & 0.8160 \end{bmatrix}$$

These provide an estimate of Subject 104's 'default' behavior across conditions. Note that the appropriateness of this estimation approach depends on thorough counterbalancing in the experimental design. The default transition matrices for all subjects in Experiment 1 are presented in Appendix C.

Figure 9.14 plots the divergence rate for Subject 104 in the 4-lever condition. In clear contrast to the entropy rate (which differed as a function of reward delivery), the divergence rate was consistently low. This indicates that Subject 104 not only employed strategies that maintained relatively uniform levels of entropy, but further that these strategies were reliably similar to the default patterns of transition.

This analysis was extended to all subjects in all conditions. The resulting mean divergence rates are plotted in Figure 9.15. In some respects, these results are similar to the change over time

[3] Because these matrices represent conditional probabilities, each row of each matrix is also subject to the closure constraint, and thus must sum to 1.0

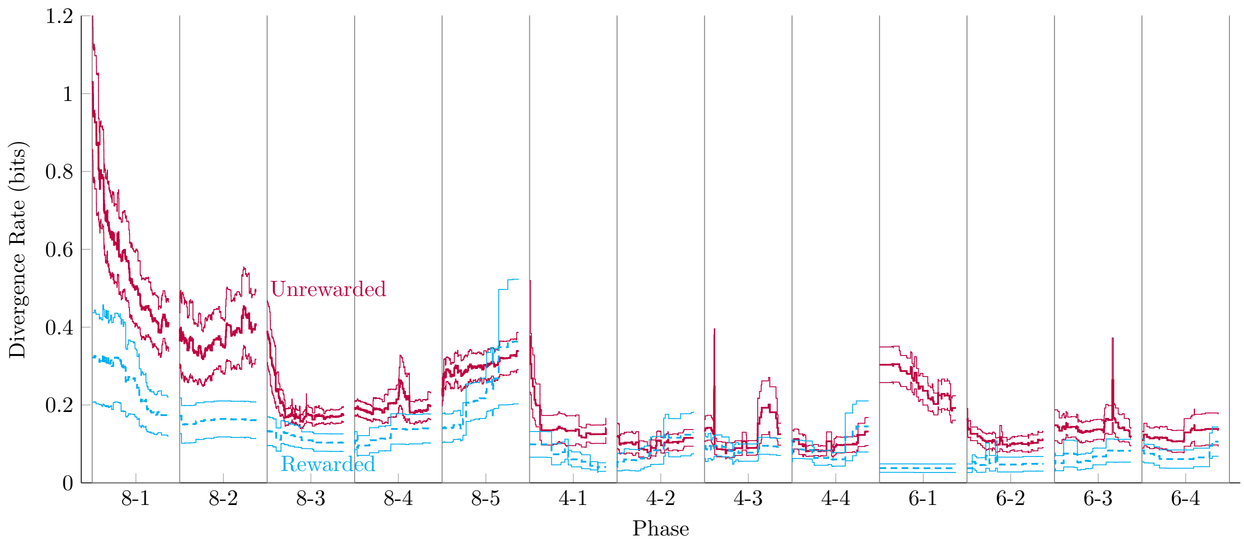

Figure 9.15: Mean Kullback-Leibler divergence rate across subjects, estimated based on the response models described by each subjects conditional response matrices.

in the mean entropy rate: Some differences remain between conditions, and responses following a reward continue to display less structure than others. However, a general effect of time is also evident, with substantial divergence from the default pattern during the early configurations of the schedule, and generally low divergence in the 4- and 6-lever conditions.

As in the case of the entropy rates, each subject's mean entropy rate per phase was rank-transformed (Conover & Iman, 1981) and subjected to a mixed-model ANOVA with phase nested within lever count, subject ID treated as a random variable, and subject ID allowed to interact with lever count. Divergence rates differed significantly as a function of lever counts in both unrewarded trials ($F(2, 18) > 15.3, p < .0002$) and rewarded trials ($F(2, 18) > 4.2, p < .04$). A significant difference was also as a function of phase during unrewarded trials ($F(10, 90) > 8.2, p < .0001$), but not during rewarded trials ($F(10, 90) = 0.88, p = .55$).

9.2.4 Transitions and Schedule Configuration

In order to reconcile the relationship between trial-by-trial estimates of sensitivity (Figure 9.10) and the proposal that behavior is best understood as deviations from the default transition matrices $\mathcal{Q}_0$ and $\mathcal{Q}_1$, it is necessary to characterize what would constitute a mismatch between a particular strategy and a particular schedule. Doing so provides some basis for determining whether or not a subject's performance can be anticipated for some unknown schedule, given some known transition matrix.

Conceptually, a mismatch is easy to imagine. In the four-lever phases, for example, subjects tended to travel along the walls unless interrupted by a reward delivery, and they did so without making many repeated responses. An example of this shown in Figure 9.7. Because these transitions were much more common than other transitions, subjects could be expected to perform poorly when the two best response alternatives were positioned at opposing diagonals from one another, because the optimal approach to such a scenario would be to travel back and forth along the diagonal. This was precisely the kind of scenario established by phase 4-4.

From this, a more general principle can be articulated: A response alternative may be considered an "island" if it is associated with a rich schedule, so long as all transitions that are favored by a subject from that alternative lead to poor schedules. Under these circumstances, the subject must 'travel through' a bad patch to get to another rich schedule, and this will lead to excessive sampling of the poor alternatives relative to the rich alternatives.

Although this principle is intuitive, it is difficult to visualize for complex scenarios. Consequently, it would be helpful to be able to quantify the degree of 'islandness' that arises from the interaction between a transition matrix $\mathcal{Q}$ and a set of reward schedules $\mathbf{R}$. Let this index be denoted by $\mathcal{I}_{sland}$, according to the following formula:

$$\mathcal{I}_{sland} = D^2 \cdot \sum_{i,j}^{D} \mathcal{Q}_{i,j} \cdot \pi_i \cdot R_i \cdot R_j \quad \text{given } \mathcal{C}\left(\mathbf{R}\right) \tag{9.8}$$

In this case, π_i corresponds to the frequency of state i in the stationary distribution of $\mathcal{Q}$. If the transition $\mathcal{Q}_{i,j}$ is associated with both a relatively rich schedule R_i and another relatively rich schedule R_j, then the quantity $\mathcal{Q}_{i,j} \cdot R_i \cdot R_j$ is expected to be relatively large. If, on the other hand, either R_i or R_j are poor, then the quantity will be smaller. Note that this only provides a rough approximation, as the reward dynamics of the schedule have a temporal component.

Figure 9.16 shows the relationship between the average $\mathcal{I}_{sland}$ index across subjects to the average sensitivity (α), in each of the 13 phases. As in previous figures, the eight-lever case is plotted in black, the six-lever case in gray, and the four-lever case in white. These sensitivities are averages of the per-trial estimates from Figure 9.10, and as such exhibit considerable variability. Nevertheless, there appears to be a trend that higher sensitivities are achieved when a better

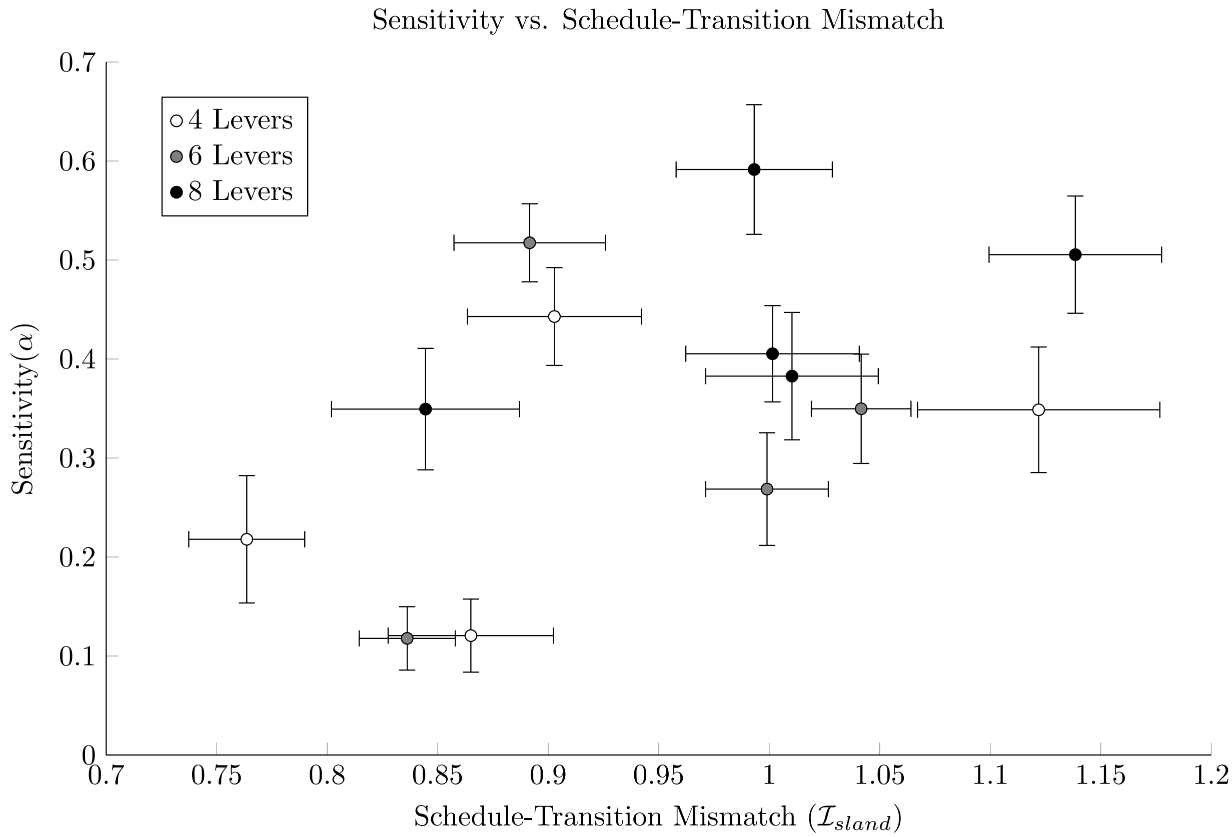

Figure 9.16: Degree of mismatch between transition matrices and schedules. Error bars correspond to standard errors.

correspondence between transition matrix and schedule are observed. A mixed-effects regression model predicting sensitivity confirmed a continuous relationship with $\mathcal{I}_{sland}$ ($F(1,117) > 4.4, p < .04$), as well as a fixed effect for number of levers ($F(2,117) > 6.44, p < .003$). The effect size of these effects were $\eta^2 = .03$ and $\eta^2 = .08$, respectively, corresponding to relatively small effects according to Cohen's rule of thumb (Cohen, 1988). The random effect for individual differences was not significant ($F(9,117) > 1.4, p = .19$).

9.3 Discussion

Ten subjects earned rewards by pressing either 4, 6, or 8 fixed levers. Rewards earned from each lever were governed by concurrent schedules. Under each configuration of levers (first 8, then 4, then 6), subjects experienced different variations of the reward schedules. A series of analyses were performed to assess the manner in which subjects adapted their behaviors to these changes.

A molar analysis of behavior was performed using the barycentric matching model (Equation 1.12), with parameter estimates obtained by conducting a linear regression on the ilr-transformed data (Equation 5.1). The parameters resulting from this analysis were an overall 'sensitivity' to the programmed probabilities of reward (denoted by α) and a composition of 'bias' parameters (denoted by $\boldsymbol{\kappa}$).

Surprisingly, subjects in the 8-lever condition displayed both a higher sensitivity and less bias overall than they did in the 4- and 6-lever conditions. Additionally, although a significant difference was not found between the 4- and 6-lever conditions, there were hints that a small effect might exist, albeit one that Experiment 1 was underpowered to detect.

The direction of the observed differences in both parameters was unexpected. Under an information-processing model, rising task complexity might be expected to increase the associated cognitive load, and thus to degrade the quality of task performance (Lavie, 2010). The perils of 'overload' are taken seriously in domains as disparate as ecology (Sol, 2009) and business (Eppler & Mengis, 2004), suggesting that the metaphor of the brain as an overtaxed computer is deeply compelling.

Despite this intuition, however, there are other lines of empirical evidence suggesting that operational models of cognitive overload (or at least, the interpretation of their experimental measurements) may be overvalued. For example, comparative work suggests that if the relationship between reaction times and number of alternatives provides a means of measuring intelligence (or at least, cognitive capacity), then it follows that pigeons must be more intelligent than humans (Vickery & Neuringer, 2000). In practice, the information processing load of a task is determined not by the task's superficial characteristics, but instead by the way in which information about the task is processed. With this in mind, the analysis turned to a more process-based approach, in order to determine how the behavior emerged.

This process-oriented analysis began by using a change-point analysis (Jensen, 2014 (projected)) to model the behavior. An initial effort applied to the full time-series of behavior found few inflection points. However, when change-point analyses were instead applied to the time series *conditional* on the previous response and the previous reward outcome, substantial response structure emerged, as did a more nuanced understanding of which aspects of behavior changed, and at which times. This yielded a two-fold model, in which the contingency table $\mathcal{B}_1$ governed responses following a reward, while $\mathcal{B}_0$ governed behavior otherwise. Rather than using change-points to segregate the data into entirely isolated epochs, the change-points identified in the conditional time series were used to update individual rows of $\mathcal{B}_0$ and $\mathcal{B}_1$ independently of one another.

Because both $\mathcal{B}_0$ and $\mathcal{B}_1$ represent positive-recurrent stationary Markov processes, their long-term stationary distribution can be solved for arithmetically. These distributions, when normalized using the overall bias vector $\boldsymbol{\kappa}$, permit the sensitivity parameter α to be recalculated for any time point. The means of the resulting 'sensitivity over time' estimates are plotted in Figure 9.10. Although these averages are not representative of individual animals, they nevertheless reveal that changes in behavior continue to occur throughout each phase.

Although the sensitivity parameters changed a great deal over the course of the 3,500 responses plotted in each phase, the rate of rewards per response was confined to a fairly narrow band near 20%. This suggests that although the relative probability of reward continued to motivate behavior (without which estimates of α would remain near zero), the overall reward *rate* was not the primary signal being used to shape and refine that behavior.

In order to explore the computation underlying behavior, two metrics derived from information theory were employed: the 'entropy rate' (Equation 9.5), which measures the additional Shannon entropy provided by each trial (McMillan, 1953), and the 'divergence rate' (Equation 9.7), which measured the average Kullback-Leibler divergence (Kullback & Leibler, 1951) between each trial's estimated conditional probability matrices ($\mathcal{B}_0$ and $\mathcal{B}_1$) and the 'default transition matrices' ($\mathcal{Q}_0$ and $\mathcal{Q}_1$) derived across all conditions. Entropy rate measures suggested, unsurprisingly, that the complexity of behavior was highest in the 8-lever condition and lowest in the 4-lever condition (Figure 9.13). However, the divergence rate across conditions did not follow a similar pattern. Although it was high in the early phases of the experiment, it tended toward toward a relatively low rate that remained stable across conditions (Figure 9.15).

This contrast is important, because the complexity of behavior depends both on the processing that takes place in the brain *and* on the complexity of the environment in which the behavior is being measured. Many relatively simple strategies will result in more unpredictable behavior if employed in a more complex environment. The default strategy described by $\mathcal{Q}_0$ and $\mathcal{Q}_1$ can be considered a subject's 'one-size-fits-all' approach to a problem (whose cognitive load is likely small), and the divergence rate measures how much additional information must be encoded to account for changes in observed behavior over time.

It is not surprising, therefore, that the 8-lever condition yielded both higher divergence rates and higher sensitivities overall. A subject with a low divergence rate has limited itself to only those strategies that differ slightly from some default pattern, and this limitation is unlikely to permit response proportions to exactly match relative reward probabilities. The two are not inextricably linked, however. For example, the mean sensitivity in Phase 8-4 plateaued around 0.5, whereas phase 8-2 reached a mean closer to 0.4. The mean divergence rate for Phase 8-4, however, was merely 0.2 bits per trial, compared to the 0.4 bits per trial observed in Phase 8-2.

Thus, rather than interpreting sensitivity and divergence rate as reflections of the same phenomenon, it is more useful to think of them as tapping into different aspects of the response process. Sensitivity is a useful proxy for a subject's *behavioral efficacy*, measuring ability to adapt to feedback provided by the environment. Divergence rate, however, is a useful proxy for *computational efficiency*, measuring the extent to which subjects are able to minimize the cognitive load associated with adapting to the current schedule by reconciling their own predispositions with the current task demands.

The measure $\mathcal{I}_{sland}$ provides only a rough approximation of the degree of correspondence between a subject's transition bias and the particulars of the current schedule, but using it as a basis for comparison (as in Figure 9.16) yields results consistent with the hypothesis that subjects perform less well when schedules require them to deviate further from their behavioral baseline. However, it remains unclear whether the degree of mismatch is adequate to explain the differences observed, as a main effect of lever count was detected in the corresponding regression analysis.

9.3.1 Confounds

In order to validate the above interpretation of sensitivity and divergence rate, however, a number of confounds must be addressed by subsequent experiments.

First, it is important to demonstrate that the barycentric matching model (which assumes a single sensitivity parameter) is valid, relative to other models that make use of additional parameters. As observed in Chapter 5, each isometric contrast of ILR-transformed data can be assigned its own sensitivity parameter without violating the implicit geometry of compositional data. A more completely counterbalanced 8-lever design is necessary to test this hypothesis effectively. This is undertaken in Experiment 2, described in Chapter 10.

Next, it is essential to determine whether the effects observed in Experiment 1 would arise given *any* configuration of levers, or whether the particular configuration of levers utilized had a determining effect. To test this, Experiment 3 made use of different configurations of four levers, as described in Chapter 11.

Another vitally important factor to counterbalance is the influence of order of experience, as it relates to subjects' age. The gradual decrease in the divergence rate observed in Figure 9.15 does not seem to be mainly driven by number of choice alternatives, but rather by cumulative learning history. The generally-lower sensitivity parameters observed in the 4- and 6-lever conditions may have experienced a similar decrement. Chapter 12 will make use of counterbalanced order of experience to determine the extent to which learning history, rather than number of alternatives, might be responsible for the relative insensitivity of behavior in latter phases of Experiment 1.

Another confound addressed in Chapter 12 (in counterbalanced fashion) is that of the relative richness of the schedule. A reward rate of one pellet per five responses is high relative to many studies reported in the literature, and it is possible that the observed effects would fail to manifest under leaner schedules. For example, if reward deliveries constitute information to be used by subjects, then schedules that are overly lean may not provide sufficient information to discover effective strategies, while schedules that are overly rich may not necessitate any meaningful discrimination among the alternatives.

Chapter 10

Eight Alternative Replication

As surprising as the results of Experiment 1 are with respect to the sensitivity parameter, various potential confounds must be addressed. The first of these is the concern that the failure to fully counterbalance aspects of the design led to a distorted sense of the effects.

The most straightforward of these concerns is the failure to fully counterbalance the eight-lever case (such that each alternative was sampled at each level of schedule richness). Although the compositional matching equation should in principle be robust against incomplete counterbalancing, demonstrating consistent estimates in practice is straightforward to accomplish. Furthermore, replicating the original design effectively double the size of the subject pool, which helps to effectively characterize population parameters.

Additionally, a more fully counterbalanced design is better suited to testing whether a single sensitivity parameter is a suitable assumption, as compared to a model with one parameter per contrast (e.g. Equation 5.5).

10.1 Methods

Experiment 2 was very similar to Experiment 1, by design. Limited to the eight-lever case, the first five phases of Experiment 2 presented subjects with conditions exactly identical to those in Experiment 1. Then, three additional phases rounded out the counterbalancing of the lever schedules.

Phase	Lever 1	Lever 2	Lever 3	Lever 4	Lever 5	Lever 6	Lever 7	Lever 8
1 (Eight Levers)	.0422	.0357	.0617	.0552	.0097	.0065	.0227	.0162
2 (Eight Levers)	.0552	.0162	.0357	.0065	.0617	.0227	.0097	.0422
3 (Eight Levers)	.0162	.0617	.0097	.0357	.0227	.0552	.0422	.0065
4 (Eight Levers)	.0617	.0422	.0065	.0162	.0552	.0097	.0357	.0227
5 (Eight Levers)	.0065	.0097	.0162	.0227	.0357	.0422	.0552	.0617
6 (Eight Levers)	.0357	.0227	.0552	.0617	.0422	.0162	.0065	.0097
7 (Eight Levers)	.0227	.0065	.0422	.0097	.0162	.0357	.0617	.0552
8 (Eight Levers)	.0097	.0552	.0227	.0422	.0065	.0617	.0162	.0357

Table 10.1: Schedule probabilities in Rat Experiment 2

10.1.1 Subjects & Apparatus

Subjects were a new set of 12 male albino Sprague-Dawley rats (Charles River, NY), weighing 450-650 g. As in Experiment 1, subjects were given preliminary training to press a single lever on a continuous reinforcement (CRF) schedule prior to running, but were otherwise naïve.

Unless otherwise noted, subject care protocols and the experimental apparatus were identical to those described in Experiment 1.

10.1.2 Procedure

As in Experiment 1, subjects were presented with an eight-lever Turn-Based Foraging paradigm (Algorithm 1). Probabilities of reward setup in each of the eight phase are described in Table 10.1. As in Experiment 1, each phase consisted of ten sessions, each lasting 60 minutes (or ending earlier, after 100 rewards had been earned).

10.2 Results

As in experiment 1, use of the ILR transformation necessitated the specification of an orthonormal basis **U** (Equation 4.6). The molar analysis of Experiment 2 made use of the following basis in the

8-lever case:

$$\mathbf{U}_8 = \begin{bmatrix} \mathbf{u_1} \\ \mathbf{u_2} \\ \mathbf{u_3} \\ \mathbf{u_4} \\ \mathbf{u_5} \\ \mathbf{u_6} \\ \mathbf{u_7} \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} \\ \sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} \\ \sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} \\ -\sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} \\ \sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} \\ \sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} \\ -\sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} & -\sqrt{\frac{1}{8}} & \sqrt{\frac{1}{8}} \end{bmatrix}$$

As in the basis used in Experiment 1, each column refers to a lever, while each row represents an empirical contrast. Because there is no need to provide identical contrasts across different configurations of the operant chamber, a basis was selected that more intuitively represented the qualitative distinctions among operanda, according to the following scheme:

$\mathbf{u_1}$ = (Levers close to the floor) vs. (Levers far from the floor)

$\mathbf{u_2}$ = (Levers close to the food) vs. (Levers far from the food)

$\mathbf{u_3}$ = (Levers on the subject's left) vs. (Levers on the subject's right)

$\mathbf{u_4}$ = (Levers close to the door) vs. (Levers far from the door)

$\mathbf{u_5}$ = Opposing diagonal corner interaction

$\mathbf{u_6}$ = Left-to-right diagonal interaction

$\mathbf{u_7}$ = Front-to-back diagonal interaction

Figure 10.1 shows the sensitivity and bias parameters across all subjects in Experiment 2. These distributions are largely consistent with the parameters observed for the eight-lever condition in Experiment 1: a mean α of 0.40 (compared to a mean of 0.46 in Experiment 1), and a mean bias vector of $\boldsymbol{\kappa}^\circ = [0.61, 0.55, -0.21, -0.13, 0.15, 0.05, -0.47, -0.55]$ (compared to $\boldsymbol{\kappa}^\circ = [0.39, 0.56, -0.28, -0.09, 0.08, 0.10, -0.39, -0.38]$ in Experiment 1). In these regards, behavior in Experiments 1 and 2 were similar, despite differing in their degree of counterbalancing and their use of different subjects.

An important additional test is to consider whether a single α parameter is appropriate. As noted in Chapter 5, each predictive factor $\mathbf{u_i}$ in an ILR-transformed matrix can support its own

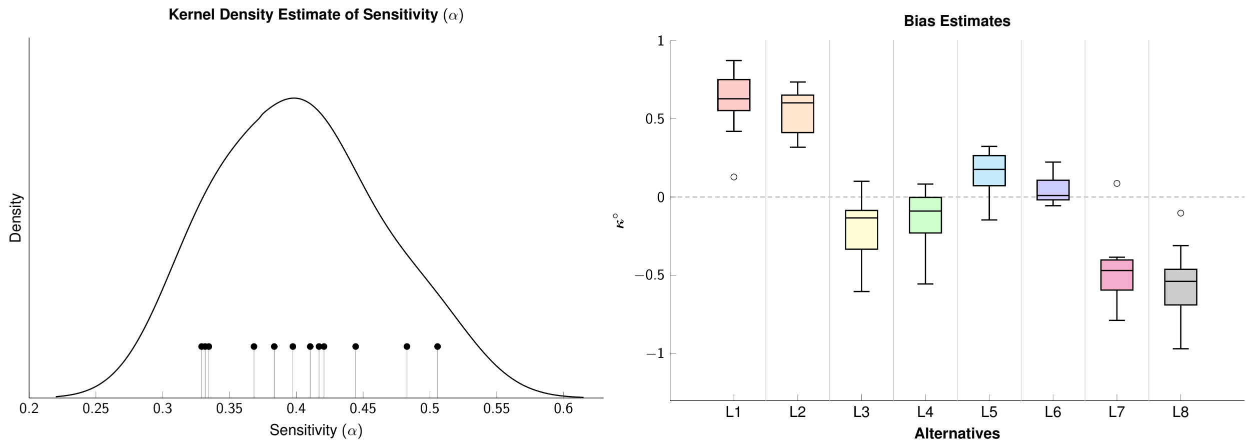

Figure 10.1: Molar parameter estimates for Experiment 2. (Left) Kernel density estimate of sensitivity (α) across the 12 subjects. (Right) Box-and-whisker plots of the distributions of CLR-transformed bias parameters ($\boldsymbol{\kappa}^\circ$).

exponent α_i. Consequently, a regression may be performed on every contrast individually, obtaining bias and sensitivity for each. In an eight-alternative context, this increases the number of free parameters from 8 to 14, but permits the question of whether the contrasts are equivalent to be addressed.

Figure 10.2 shows boxplots of the distributions of sensitivities for each of the contrasts. It appears to imply that the contrast $\mathbf{u_1}$ (proximity to the floor) tended to engender somewhat higher sensitivity than the rest, while $\mathbf{u_3}$ (handedness) tended to yield less sensitive discrimination. Figure 10.2 also suggests that subjects tended to show more varied parameters for the diagonal interactions (especially $\mathbf{u_5}$). Given the differences in variance, a nonparametric Friedman test (Quade, 1984) was performed, and a significant difference in the sensitivities was identified ($\chi^2\,(2) = 13.4, p < .002$). Post-hoc multiple comparisons test (Rhyne & Steel, 1965) found significant differences only between $\mathbf{u_1}$ and $\mathbf{u_3}$ on the one hand, and $\mathbf{u_1}$ and $\mathbf{u_6}$ on the other; all other differences were non-significant.

Although a difference in the contrasts was found, the single-α model nevertheless has the advantage of parsimony. To determine whether a model with multiple α parameters would be appropriate, Akaike's corrected information criterion (AICc) (Hurvich & Tsai, 1989) was computed for three models: An "intercept only" model acting as a null hypothesis, a "single α" model, and a "multiple α" model based on the contrasts in the matrix $\mathbf{U}_8$ above. Table 10.2 shows these AICc scores for

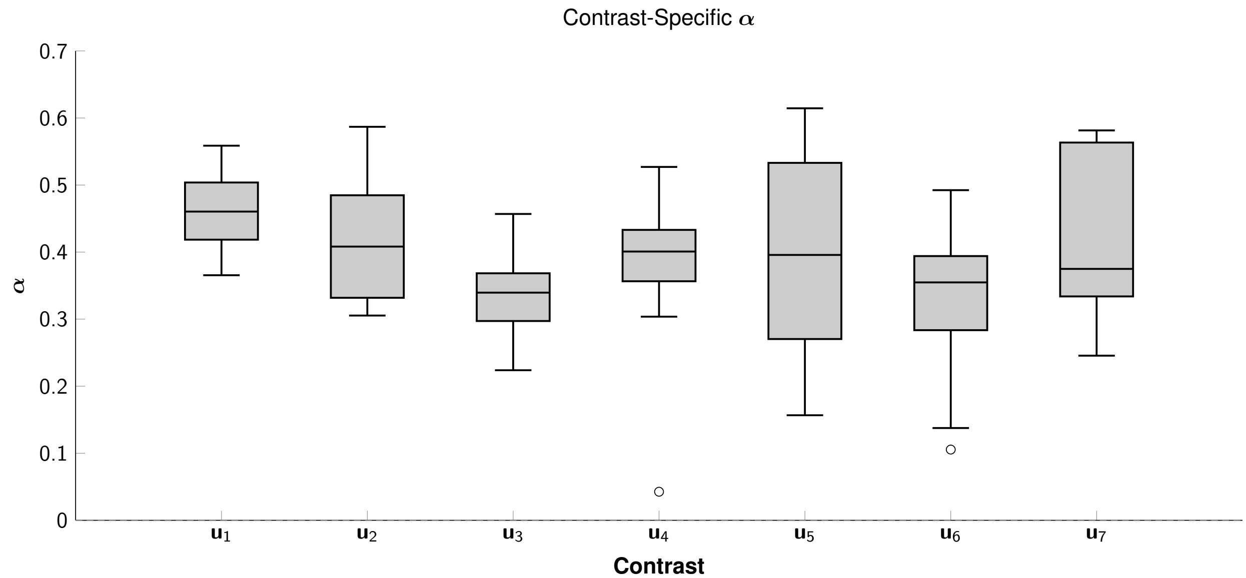

Figure 10.2: Molar sensitivity estimates calculated for each orthonormal contrast in Experiment 2.

each subject and each model. In every case, the single α model yields the lowest score (and thus the best fit of the three). Consequently, a single-α model appears preferable as a default approach.

The number of change-points detected per phase is plotted in Figure 10.3. As in Experiment 1, early phases appear to engender more frequent changes; however, unlike Experiment 1, the present data consist only of eight-lever configurations. A rank-transformed (Conover & Iman, 1981) mixed-model ANOVA was performed, with phase as a fixed effect and subject ID as a random effect. Significant differences were found both with respect to phase ($F(7,77) > 12.6, p < .0001$) and subject ID ($F(11,77) > 3.3, p < .001$). Consequently, even when limiting the focus to eight levers, subjects appear to make more frequent changes to their strategies during the earliest phases of their experience, and make fewer changes later on, a pattern consistent with the hypothesis that early responding is characterized by a more dynamical strategy, whereas later responding consists mainly of small corrections to a general pattern.

As in Experiment 1, change-point analysis was used to detect changes in the conditional response probabilities of subjects, and these were used to obtain trial-specific estimates of their stationary distributions. Given the time-specific stationary distribution and the correcting for bias, trial-specific estimates of sensitivity were obtained. The mean of these estimates across subjects is plotted in Figure 10.4. Although subjects consistently achieved a high sensitivity in the first phase,

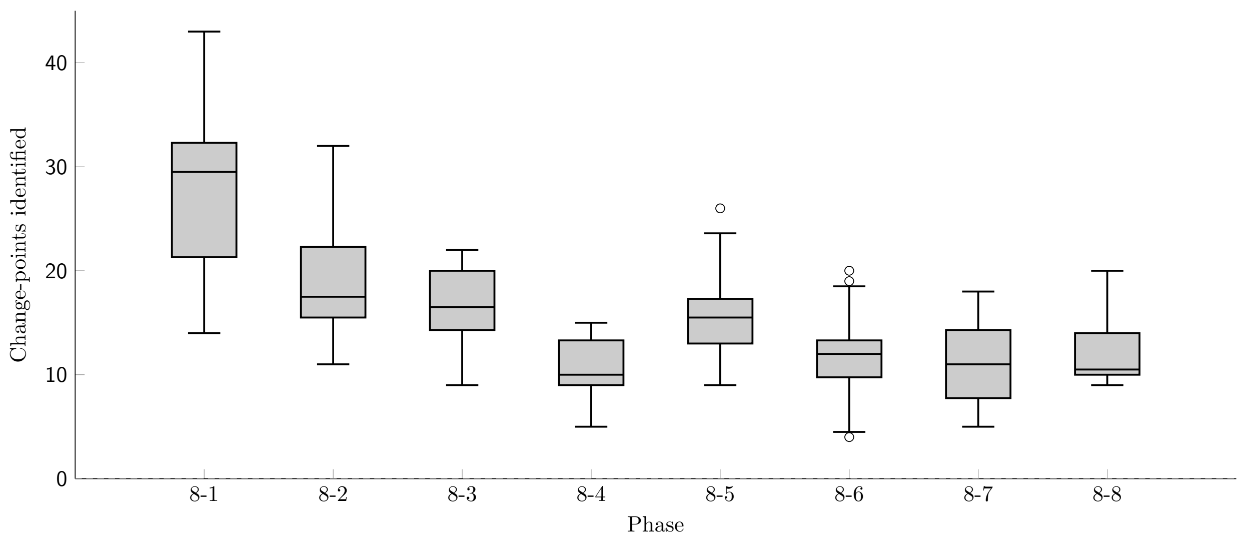

Figure 10.3: Number of change-points detected per phase in Experiment 1.

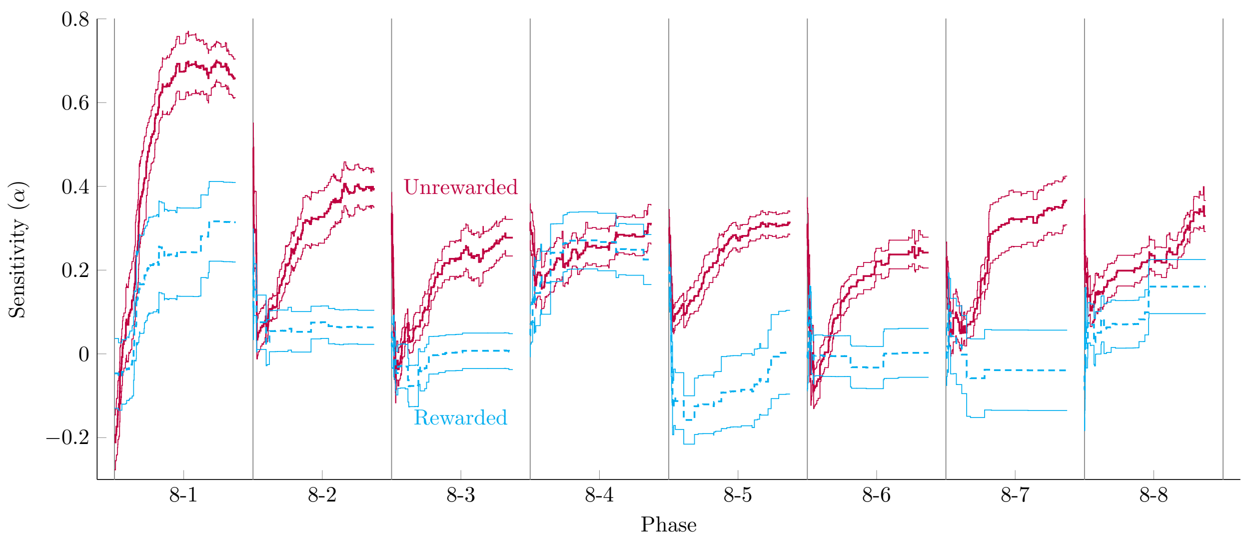

Figure 10.4: Mean sensitivity (α) estimate across subjects for each of the first 3,500 trials in each phase. The blue dashed line indicates sensitivity conditional on having just received a reward, while the red line indicates sensitivity otherwise. Thin lines indicate one standard error of the mean.

Subject	Intercept Only	Single α	Multiple α
1	115.82	34.45	51.99
2	99.52	-0.98	16.98
3	91.24	-13.52	-3.61
4	90.98	-9.72	7.21
5	98.46	-1.80	15.55
6	105.69	30.67	41.39
7	82.43	40.04	46.32
8	94.85	-28.36	-22.65
9	114.57	-18.17	-6.98
10	95.28	-10.94	0.41
11	82.22	0.15	5.12
12	86.78	16.39	29.74

Table 10.2: Akaike's corrected information criterion (AICc) for each molar model in Experiment 2. The lowest score for each subject is highlighted in green.

their subsequent parameters consistently failed to rise to a similar level, a similar result to the eight-level condition in Experiment 1.

Figure 10.5 shows the mean reward rate, also computed on a per-trial basis using each subject's first-order contingency table. As in Experiment 1, rates were relatively invariant, showing a small improvement over the first 3,500 trials of each phase, shifting by at most a few percentage points. Reward rate did not appear particularly well correlated with sensitivity: Lower rates were observed in phases 8-5 and 8-7, despite those phases displaying comparable levels of sensitivity.

Figure 10.6 shows the change in the the mean entropy rate (Equation 9.5) across subjects, as derived from their first-order contingency tables. Unlike sensitivity (which showed a marked change at the onset of each new phase, due to the change in the reward schedule), entropy rates seemed relatively unaffected by the transition between schedules. Instead, a gradual decrease was observed over the duration of the experiment.

This trend is even more evident in the mean divergence rate (Equation 9.7), shown in Figure 10.7. As in Experiment 1, the default transition matrices for each were subject were estimated

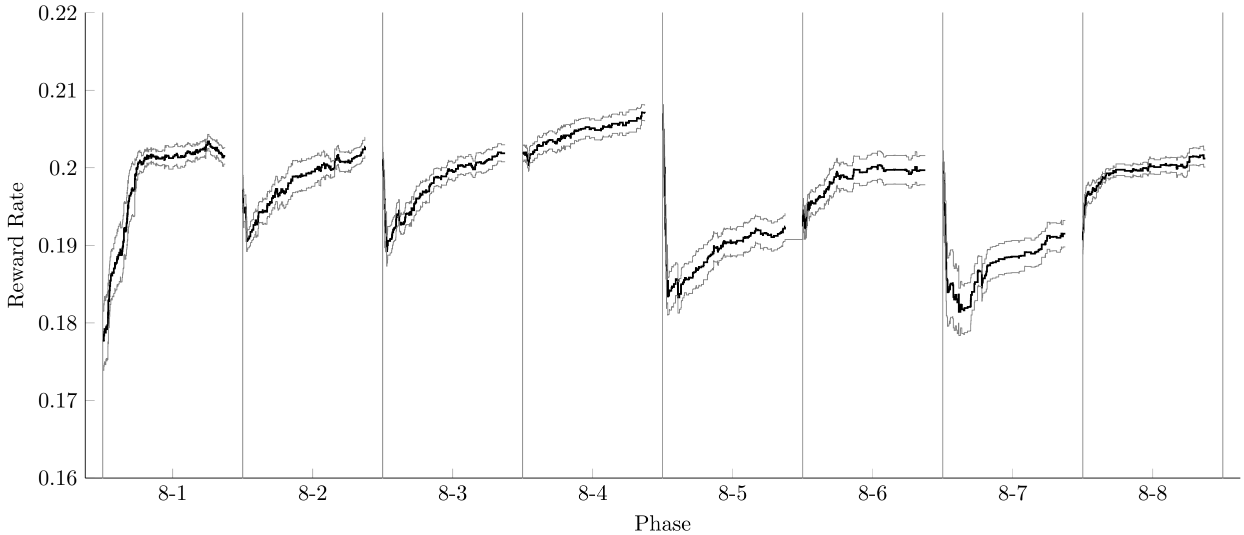

Figure 10.5: Mean rewards earned per response for each of the first 3,500 trials in each phase (as estimated based on 100,000 simulated responses using a subject's first-order contingency table at that time). Thin lines indicate one standard error of the mean.

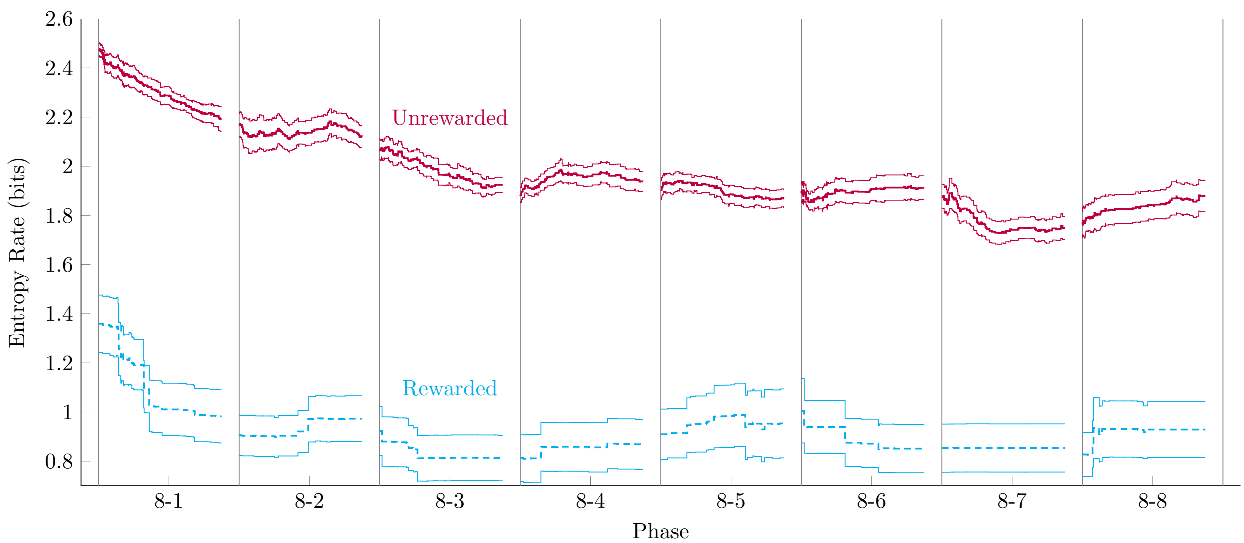

Figure 10.6: Mean entropy rate (Equation 9.5) estimated across subjects for each of the first 3,500 trials in each phase. The blue dashed line indicates sensitivity conditional on having just received a reward, while the red line indicates sensitivity otherwise. Thin lines indicate one standard error of the mean.

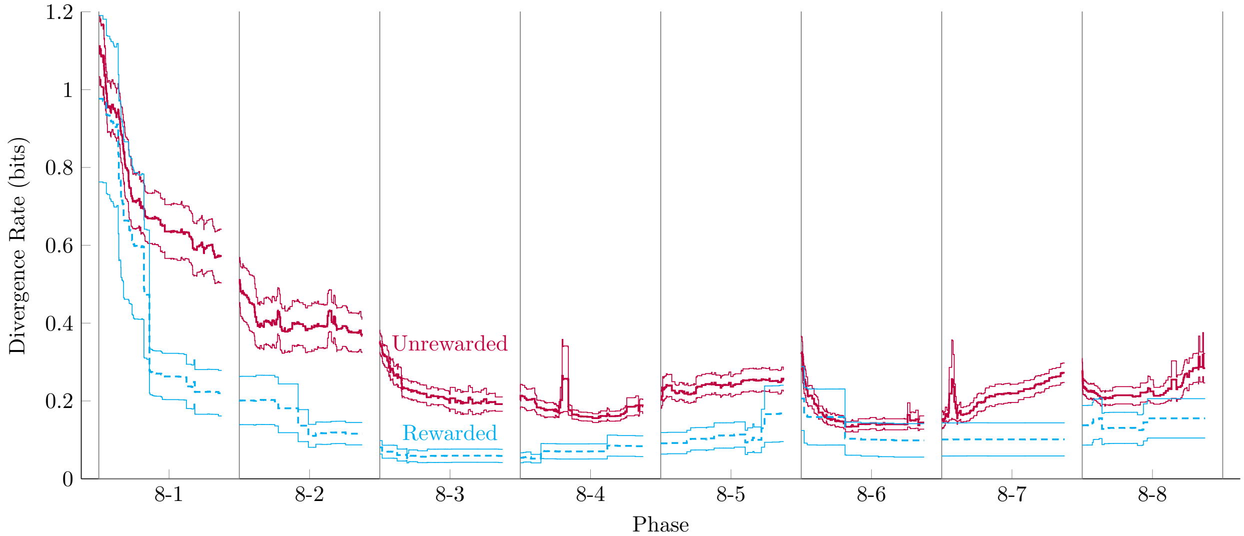

Figure 10.7: Mean divergence rate (Equation 9.7) estimate across subjects for each of the first 3,500 trials in each phase. The blue dashed line indicates sensitivity conditional on having just received a reward, while the red line indicates sensitivity otherwise. Thin lines indicate one standard error of the mean.

(reported in Appendix C), and each subject's trial-by-trial divergence from that default strategy was estimated. By the third phase, subjects had generally settled on a stable divergence rate of 0.2 bits per response, on average. This low rate suggests that the adaptation shown by subjects fell within a narrow informational radius around each subject's default transition biases.

10.3 Discussion

A group of 12 rats replicated the first 5 phases of Experiment 1, following which they completed three additional phases to receive fully counterbalanced experience of the schedule. Both with respect to molar and molecular analyses, this replication yielded equivalent results to those observed in Experiment 1.

The more fully counterbalanced account of behavior permits test of whether making use of different sensitivity parameter for different orthonormal contrasts was valid. According to this analysis, some differences were detected. For example subjects tended to be more sensitive to the contrast of low levers vs. high levers than they were to the contrast of left levers vs. right levers. On this basis, it is reasonable to suppose that an apparatus that exaggerated these sources of bias even further might also yield corresponding differences in sensitivity. However, the differences observed were small and (given the sample size), were not sufficient to justify favoring the more complex model when considered in terms of an information criterion. Thus, although it is reasonable to assume that differences in sensitivity to various contrasts may be achieved by manipulating their physical characteristics, it is also reasonable to assume that these effects are not substantial in this apparatus.

As in Experiment 1, subjects began the experiment using a relatively complex strategy. Their early performance is also marked by a higher sensitivity parameter. This was gradually replaced by the more simplistic strategy over the course of the first two phases. By the third phase, most subjects consistently displayed a divergence rate of less than 0.4 bits per response, a low rate that generally persisted for the rest of the experiment.

Because Experiment 2 counterbalanced across a wider range of schedule configurations, it also presumably provides a better estimate of the 'transition bias' matrices $\mathcal{Q}_0$ and $\mathcal{Q}_1$ (see Appendix C), upon which the calculations of the divergence rate depend. Identifying these matrices in an accurate way is a general problem for this approach, as samples of behavior that are too small (or unbalanced) will yield biased results. The similarity of these matrices in the present experiment to those observed in the eight-lever condition of Experiment 1 provides some support for the notion that the two experiments made use of adequately counterbalanced data.

Because the same subjects are used in Experiment 3, described in the following chapter, further discussion of the results are presented there.

Chapter 11

Counterbalancing of Four Alternative Choice

A difficulty with Experiment 1 is that it is not obvious *a priori* that the four levers selected for use in that experiment provided an appropriate basis upon which to generalize about four-operandum performance in general. It is conceivable, for example, that dramatic differences in performance might arise as a result of the levers selected.

To address this concern, the subjects from Experiment 2 immediately continued in another experiment during which several configurations of four levers were provided.

11.1 Methods

11.1.1 Subjects & Apparatus

The same subjects used in Experiments 2 and 3. They transitioned immediately from Experiment 2 to 3 without interruption of the running schedule. Unless otherwise noted, subject care protocols and the experimental apparatus were identical to those described in Experiment 1.

11.1.2 Procedure

As in previous experiments, the Turn-Based Foraging paradigm (Algorithm 1) was employed to provide subjects with concurrent schedules or reward. The probabilities used in each of 18 phases

Phase	Lever 1	Lever 2	Lever 3	Lever 4	Lever 5	Lever 6	Lever 7	Lever 8
1 (Four Levers A)	–	–	.0178	.0417	.0774	.1131	–	–
2 (Four Levers A)	–	–	.1131	.0774	.0417	.0178	–	–
3 (Four Levers A)	–	–	.0417	.1131	.0178	.0774	–	–
4 (Four Levers B)	–	–	–	–	.1131	.0417	.0178	.0774
5 (Four Levers B)	–	–	–	–	.0178	.1131	.0774	.0417
6 (Four Levers B)	–	–	–	–	.0417	.0774	.1131	.0178
7 (Four Levers C)	–	–	.0774	.0178	–	–	.0417	.1131
8 (Four Levers C)	–	–	.1131	.0417	–	–	.0178	.0774
9 (Four Levers C)	–	–	.0178	.1131	–	–	.0774	.0417
10 (Four Levers D)	.1131	.0178	.0417	.0774	–	–	–	–
11 (Four Levers D)	.0417	.1131	.0774	.0178	–	–	–	–
12 (Four Levers D)	.0178	.0774	.1131	.0417	–	–	–	–
13 (Four Levers E)	.1131	.0178	–	–	.0774	.0417	–	–
14 (Four Levers E)	.0774	.0417	–	–	.1131	.0178	–	–
15 (Four Levers E)	.0178	.1131	–	–	.0417	.0774	–	–
16 (Four Levers F)	.0774	.0178	–	–	–	–	.0417	.1131
17 (Four Levers F)	.0417	.0774	–	–	–	–	.1131	.0178
18 (Four Levers F)	.1131	.0417	–	–	–	–	.0178	.0774

Table 11.1: Schedule probabilities in Rat Experiment 3

in Experiment 3 are described in Table 11.1. These 18 phases made use of six different configurations of the levers, each persisting for three phases before being changed.

Because of the variety of different possible configurations in the experimental chamber, phases only last five sessions apiece, instead of the ten previously used. This was done to ensure that subjects did not become too advanced in age before completing the experiment.

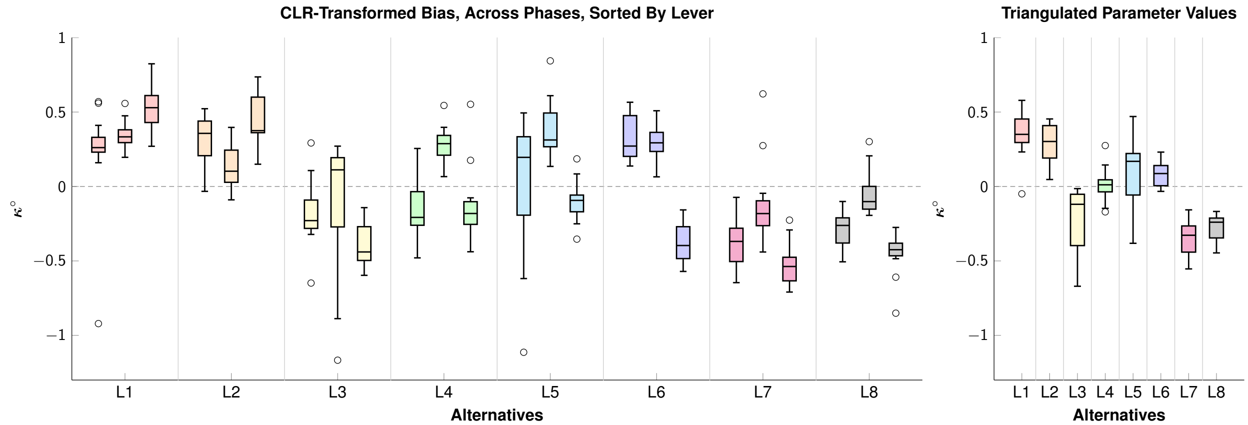

Figure 11.1: Molar CLR-transformed bias estimates (κ_i°) for Experiment 3. (Left) Estimates drawn each configuration separately, sorted by lever. (Right) Parameter estimates obtained by triangulation across conditions.

11.2 Results

Figure 11.1 (Left) shows the CLR-transformed bias parameters across the six conditions, sorted by lever. These do not provide the means for a formal comparison, as each parameter is only strictly interpretable in terms of other parameters. For example, the estimated CLR-transformed bias parameter for Lever 1 can be expected to be closer to zero in Phase 4 than in Phase 6, because Phase 4 compares four levers that are expected to be preferred more overall than the subset in Phase 6.

These bias parameters may, however, be used to obtain a 'triangulated estimate.' Since Experiment 3 was counterbalanced, the geometric mean of a subject's (untransformed) compositional bias parameter for a given alternative combines the information from each subset. For example, κ_{L1} was estimated at 0.311, 0.367, and 0.349 in phases 4, 5, and 6 respectively (such that $\mathfrak{g}\left(\kappa_{\text{L1}}\right) = 0.341$), whereas κ_{L8} was estimated at 0.181, 0.205, and 0.177 in phases 2, 3, and 6 (such that $\mathfrak{g}\left(\kappa_{\text{L8}}\right) = 0.187$). When one such geometric mean is computed per alternative, and the results are CLR-transformed, the "triangulated" bias (Figure 11.1, Right) appears very similar to the patterns obtained in previous experiments.

Figure 11.2 shows the sensitivity estimated in each of the phases. A significant difference was detected between these using the Friedman test ($\chi^2(5) = 20.33, p < .002$), although post-hoc tests only revealed significant differences between Phase 1 and Phase 2 in one case, and Phase 2

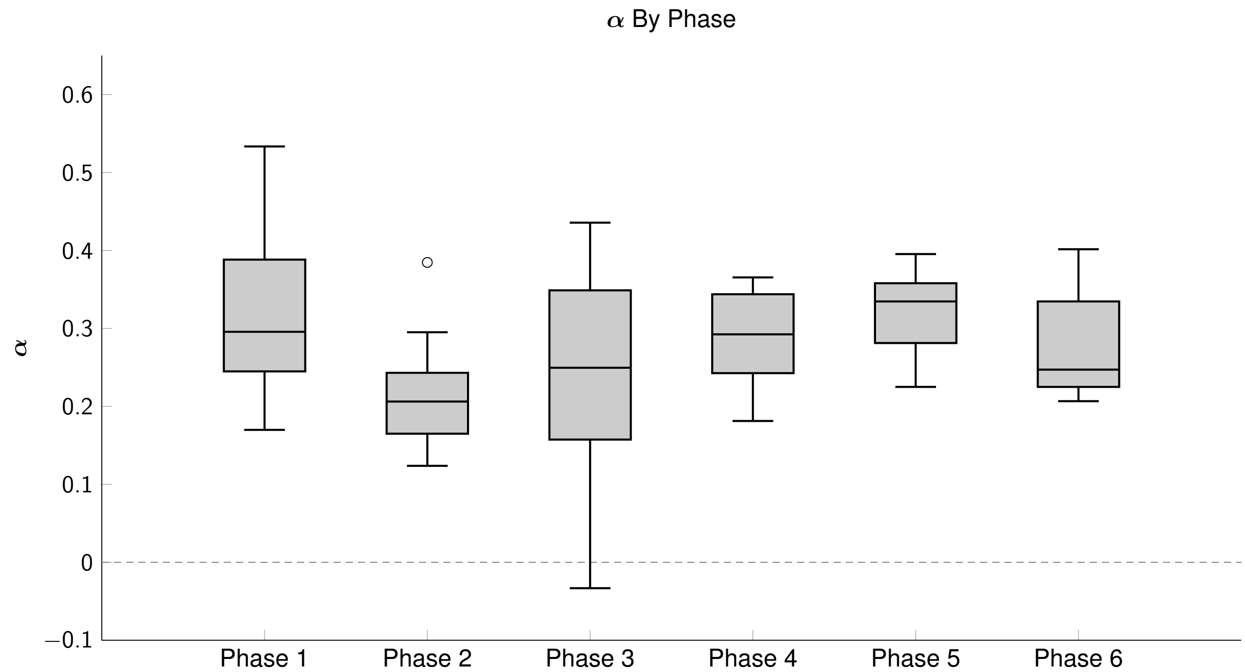

Figure 11.2: Molar sensitivity (α) estimates for each configuration in Experiment 3.

and Phase 5 in another. Another noteworthy characteristic of these data is that the sensitivity parameter was generally lower in this experiment than in the eight-lever condition described in Experiment 2, despite both experiments involving the same subjects. This replicates the finding from Experiment 1 that subjects tended to display a higher sensitivity parameter when using a larger number of alternatives.

Figure 11.3 plots the number of change-points identified in each phase. These counts are generally low, consistent with the hypothesis that subjects later in training only make small modifications to a well-established transition matrix. However, configuration C seemed to engender a higher rate of changes, and higher variability in scores. To test this non-parametrically, the data were rank-transformed (Conover & Iman, 1981) and subjected to a mixed-model nested ANOVA, in which configuration and phase were treated as fixed effects, phase was nested within configuration, subject ID was treated as a random effect, and subject ID and configuration were permitted to interact. The fixed effects were significant, with the effect of configuration being most pronounced ($F(5, 55) > 9.9, p < .0001$), followed by the effect of phase ($F(12, 132) > 3.1, p < .001$). Although a significant interaction was seen between subject ID and configuration ($F(55, 132) > 1.7, p < .01$),

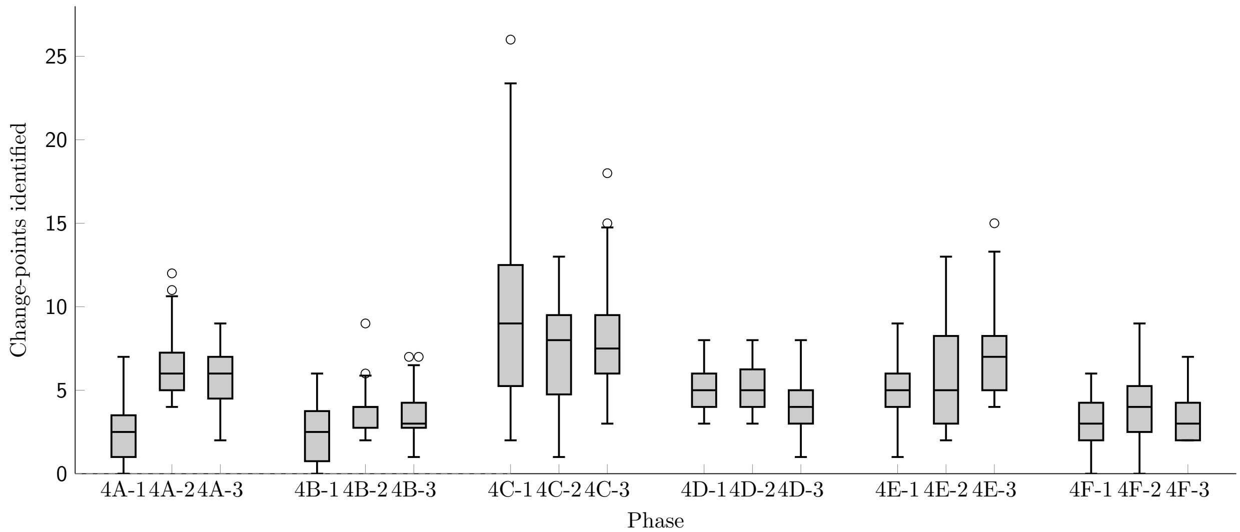

Figure 11.3: Number of change-points detected per phase in Experiment 3.

there was not a significant effect for subject ID ($F(11, 55) = 1.73, p = .09$).

Figure 11.4 shows the estimated sensitivity per trial, based on the time-series analysis of conditional change-points described in Chapter 9. Unlike previous plots of this kind, the analysis is limited to the first 2,000 trials of each phase, as the shorter phase durations resulted in some subjects emitting fewer responses than in previous experiments. As in past experiments, different conditions engendered different levels of sensitivity among subjects. Although a few configurations (4B-1, 4C-3, 4E-3) yielded almost entirely insensitive behavior, subjects nevertheless tended to fall in a range of low α values comparable to their performance of subjects in Experiment 1.

Figure 11.5 shows the reward per trial, also for the first 2,000 trials of each phase. Consistent with earlier results, reward rate was relatively invariant (shifting over a narrow range), and also poorly correlated with sensitivity, consistent with the theme that the calibration of behavior based on differential reward rates is unlikely to provide an adequate explanation for behavior.

Figure 11.6 shows the mean entropy rate for the first 2,000 trials of each phase. Both during rewarded and unrewarded trials, subjects displayed relatively uniform entropy rates despite the changes in the configurations of the levers.

As in previous experiments, the default transition matrices were estimated for each lever configuration, and are provided in Appendix C. Although the mean divergence rates (Figure 11.7) were also generally low, the onset of a change in the configuration of the levers sometimes engendered

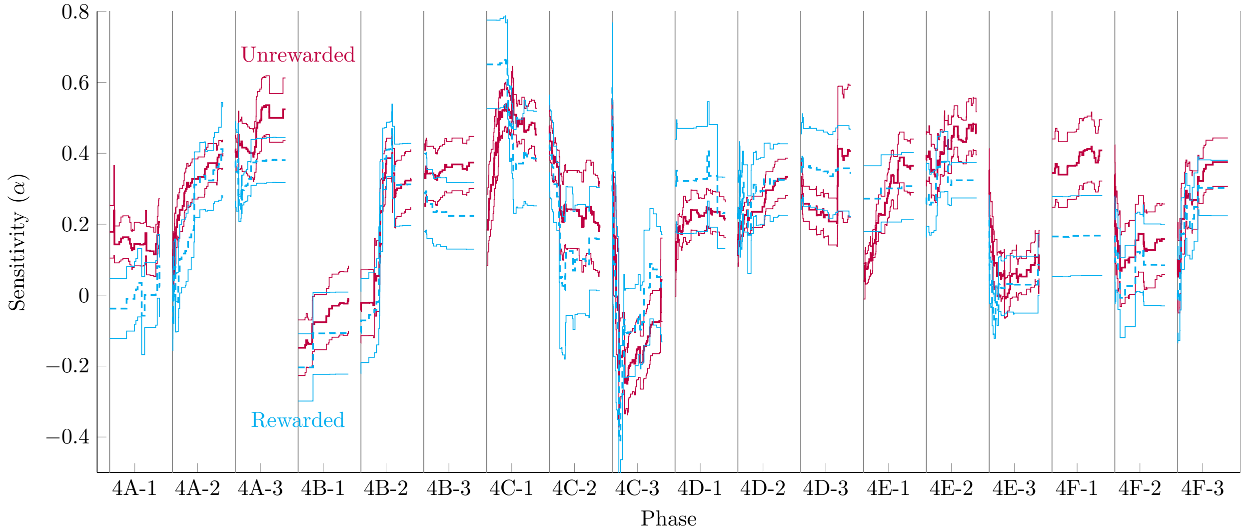

Figure 11.4: Mean sensitivity (α) estimate across subjects for each of the first 2,000 trials in each phase. The blue dashed line indicates sensitivity conditional on having just received a reward, while the red line indicates sensitivity otherwise. Thin lines indicate one standard error of the mean.

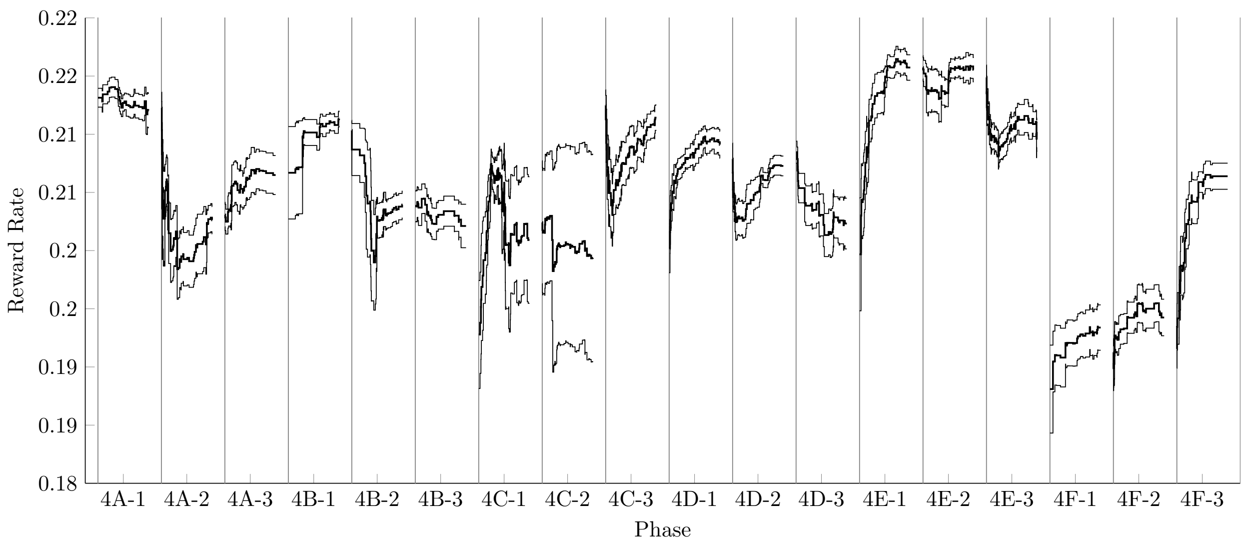

Figure 11.5: Mean rewards earned per response for each of the first 2,000 trials in each phase (as estimated based on 100,000 simulated responses using a subject's first-order contingency table at that time). Thin lines indicate one standard error of the mean.

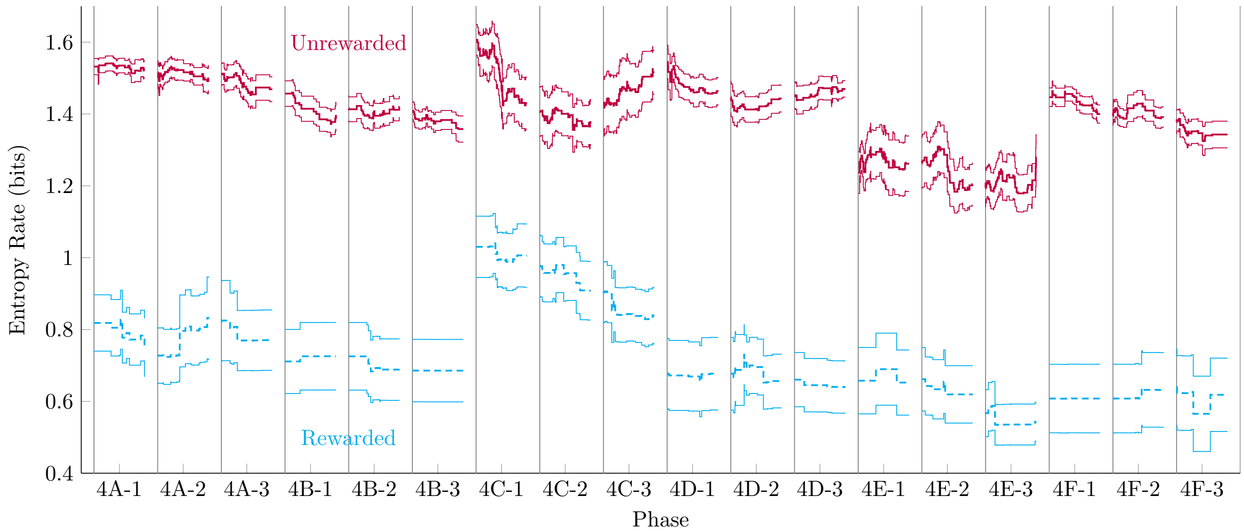

Figure 11.6: Mean entropy rate (Equation 9.5) estimated across subjects for each of the first 2,000 trials in each phase. The blue dashed line indicates sensitivity conditional on having just received a reward, while the red line indicates sensitivity otherwise. Thin lines indicate one standard error of the mean.

a temporary increase. Notable brief periods of elevated divergence took place at the beginning of phases 4C-1 and 4E-1, both scenarios where all four levers were placed on the same wall.

Figure 11.8 shows the average sensitivity in each phase on the y-axis, and the average mismatch index (denoted as $\mathcal{I}_{sland}$, Equation 9.8) on the x-axis. Because Experiments 2 and 3 made use of the same group of subjects, both experiments are plotted together to facilitate comparison. In general, the same weak correlation between the two was observed. However, three distinctive outliers are apparent. The first is the highest sensitivity. This corresponds to the first phase of Experiment 2. In both Experiments 1 and 2, performance was very high in the first phase relative to other phases, so this may be the result of an as-yet uncontrolled-for effect of either age or order of experience. The other two points correspond to sensitivities below zero, seen in phases 4B-1 and 4C-3 of Experiment 3. The reasons for these anomalously low sensitivities, seen consistently across subjects, are not clear.

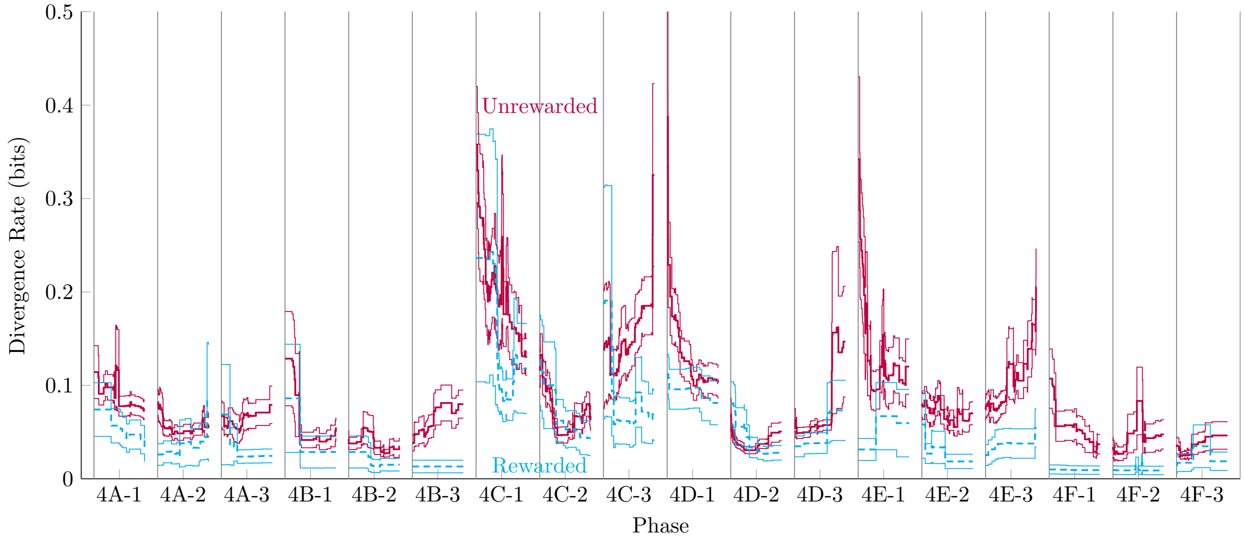

Figure 11.7: Mean divergence rate (Equation 9.7) estimate across subjects for each of the first 2,000 trials in each phase. The blue dashed line indicates sensitivity conditional on having just received a reward, while the red line indicates sensitivity otherwise. Thin lines indicate one standard error of the mean.

11.3 Discussion

Experiment 3 provided a continuation of the work done with the subjects initially used in Experiment 2. Rather than merely replicating the results from Experiment 1, however, this experiment was more ambitious in two respects. The first was its much more involved counter-balancing of different four-lever configurations, and the second was the use of shorter phases (5 sessions per phase, as compared to the previous ten).

Truly systematic counterbalancing of the operant chamber was not feasible. Given eight possible positions, there are 1,680 different possible configurations of the levers. Even assuming that various forms of symmetry may be considered redundant, the minimum number of potentially interesting configurations remains in the hundreds. The six configurations that were used therefore only scratch the surface of the possible arrangements, and the possibility exists that some noteworthy effects might have been missed. Nevertheless, those variations that were included suggest that changing the spatial configuration gives rise to corresponding changes in the behavior.

However, the second issue (namely, the use of shorter phases) presents both promise and difficulty in interpreting these results. On the one hand, it appears as though some degree of consistency

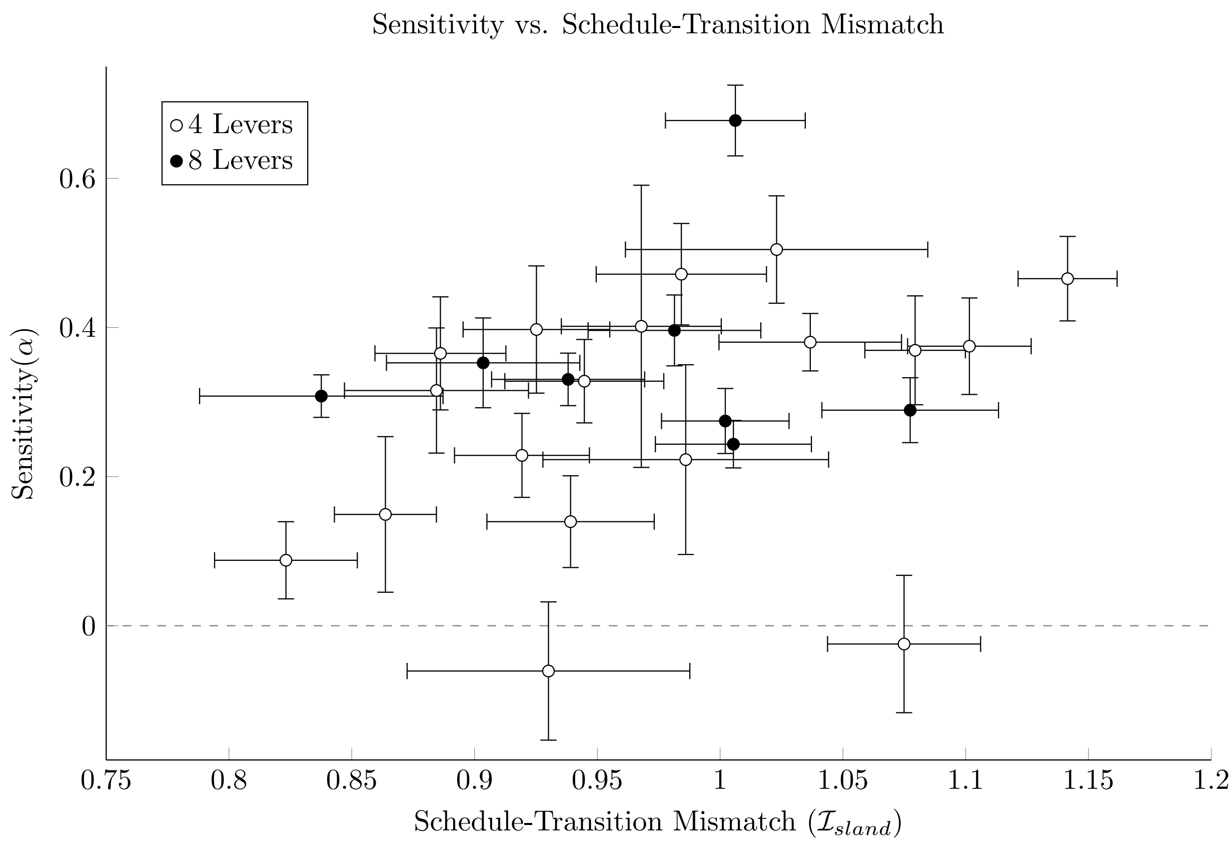

Figure 11.8: Degree of mismatch between transition matrices and schedules in Experiment 2 (black points) and Experiment 3 (white points). Error bars correspond to standard errors.

in parameter estimates can be obtained over shorter intervals, which is important because naturalistic scenarios may not give an organism hundreds of opportunities to adjust its behavior. On the other hand, however, parameter estimates were noisier than those obtained in Experiments 1 and 2, and it is unlikely that subjects achieved “steady state” responding of the same sort they did in those earlier experiments. Of particular concern is the estimation of the default transition matrices $\mathcal{Q}_0$ and $\mathcal{Q}_1$, which are essential to evaluating both the divergence rate and the degree of mismatch between the schedules and the behavior. In practice, these shortened phases were forced to some degree by the objectives of the experiment: if each of the 18 phases required two weeks of time to run, then Experiment 3 would take 36 weeks, in addition to the 16 weeks already elapsed in Experiment 2. This is a long enough period of time that aging effects necessarily contribute to some effects, but the possibility of subjects dying due to natural causes begins to also be a source of concern. This is one reason why many of long-term behavioral experiments have used pigeons instead of rats, since pigeons can easily live for over a decade in captivity. Pigeons are also relatively physically uniform throughout adulthood, as opposed to rats, who tend to continue to grow over the course of a lifetime unless food-restricted (Masoro et al., 1982).

Nevertheless, with these considerations in place, the overall story presented jointly by Experiments 2 and 3 is consistent with the hypothesis put forth to explain the results Experiment 1. During the early trials (in this case, those during the first sessions of Experiment 2), subjects engaged in an overall strategy that was more complex than the strategies observed in subsequent sessions, measured in terms of entropy. Over this period, a set of ‘transition biases’ materialized that explained the broad strokes of behavior, with only small adjustments required to exploit the experimental task in a consistent fashion. This approach continued into Experiment 3, although the physical re-arrangement of the chambers may have briefly forced the strategy into a more cognitively demanding mode. A consistent pattern was observed over this interval that, in general, subjects displayed higher sensitivities when their preferred patterns of transition yielded reasonable affordances with the arrangement of the schedules.

Several potential confounds remain. The most important of these is the order of experience which has been uniform across subjects in all experiments discussed up to this point. As previously mentioned, age effects are best understood as continuous over the life span in rats, and it is plausible that the very high sensitivity observed in first phases experienced by each subject reflects a more juvenile strategy. Order effects may also matter independent of aging, insofar as an animal's current strategy may be more or less suited to the subsequent schedule of rewards.

One final confound that has not received consideration thusfar is the reward rate. Power law models are, by their nature, scale-invariant, so if the purest forms of the generalized (Eq. 1.8) and barycentric (Eq. 1.12) matching equations are correct, then performance should yield similar baselines of sensitivity regardless of the schedule richness. There are several reasons to think, however, that schedule richness might yield shifts in behavior. Previous studies of concurrent choice in animals have reported observing impacts from reward rate (Fantino et al., 1972; Elliffe & Alsop, 1996). Though such results do not *in principle* invalidate the compositional approach, they suggest at the very least that the relationship between scheduled rewards and observed responses may not be scale-invariant.

On the basis of these considerations, another experiment was performed that provided counterbalancing between three factors: the number of alternatives (8 vs. 4), the order of experience, and the richness of the schedule. This experiment is described in the following chapter.

Chapter 12

Possible Confounding Effects of Subject Age and Schedule Richness

Although Experiments 2 and 3 examined the effects of more thorough counterbalancing, they failed to take into consideration one of the most important types of counterbalancing: Order of experience. The experiments described in previous chapters all made use of a uniform sequence of phases, rendering a rigorous test of the effects of time impossible to disambiguate from other effects of the manipulation.

Another consideration not adequately addressed by previous experiments is whether the overall frequency of rewards has an impact on the model parameters. The reward rate used in Experiments 1 through 3 was relatively high, as compared to results published elsewhere in the literature, so using a leaner schedule might reveal the effects observed thus far as peculiar to the rich schedules employed.

In order to address these concerns, an experiment was performed that counterbalanced both of these factors. All subjects were exposed to either four or eight levers, doing so under the auspices of either a relatively rich schedule (approx. 20% of responses rewarded) or a relatively lean one (approx. 10% of responses rewarded).

12.1 Methods

Experiment 4 followed an ABBA/AABB design for lever count and richness, spread across sixteen phases. In all cases, subjects began with one lever count (either four or eight levers available concurrently), then switched to the other count after four phases, and finally switched back again after a total of twelve phases. Additionally, every subject began with one schedule richness (rich or poor), switched to the other after eight phases. Between these two factors, the overall experiment changed every four phases. Given four different groupings of the subjects, this ensured that every subject experienced every condition, *and* that every condition was sampled during every quartile of the experiment.

12.1.1 Subjects & Apparatus

Subjects were a new set of 12 male albino Sprague-Dawley rats (Charles River, NY), weighing 450-650 g. Unless otherwise noted, subject care protocols, training prior to the beginning of the experiment, and the experimental apparatus were identical to those described in Experiment 1.

12.1.2 Procedure

Experiment 4 made use of 16 different 'conditions,' each of which is described in Table 12.1. Although every subject experienced every condition, they did not do so in the same order. Instead, each subject was assigned to one of four groups, experiencing the conditions according to their group's prescription. The order experienced by each group is described in Table 12.2

As in Experiment 3, phases consisted of only 5 sessions apiece. However, between phases 8 and 9, subjects experienced a four-week hiatus, during which time they continued to receive restricted food access according to the experimental protocol, but did not engage in any experimental activity. Consequently, the interval of time running from the beginning Phase 1 to the end of Phase 16 covered a span of 20 weeks.

12.2 Results

In examining the parameters of the matching equation, the most important consideration is consistency. In this spirit, the first analysis was a comparison of the bias parameters. Figure 12.1

Condition	Levers & Schedule	Lever 1	Lever 2	Lever 3	Lever 4	Lever 5	Lever 6	Lever 7	Lever 8
4PA	Four, Poor	–	–	.0209	.0387	.0089	.0566	–	–
4PB	Four, Poor	–	–	.0089	.0209	.0566	.0387	–	–
4PC	Four, Poor	–	–	.0387	.0566	.0209	.0089	–	–
4PD	Four, Poor	–	–	.0566	.0089	.0387	.0209	–	–
4RE	Four, Rich	–	–	.1131	.0774	.0178	.0417	–	–
4RF	Four, Rich	–	–	.0774	.0417	.1131	.0178	–	–
4RG	Four, Rich	–	–	.0178	.1131	.0417	.0774	–	–
4RH	Four, Rich	–	–	.0417	.0178	.0774	.1131	–	–
8PI	Eight, Poor	.0049	.0179	.0309	.0276	.0114	.0081	.0211	.0033
8PJ	Eight, Poor	.0211	.0033	.0114	.0081	.0309	.0276	.0049	.0179
8PK	Eight, Poor	.0179	.0309	.0081	.0049	.0276	.0211	.0033	.0114
8PL	Eight, Poor	.0033	.0114	.0276	.0211	.0081	.0049	.0179	.0309
8RM	Eight, Rich	.0162	.0422	.0065	.0617	.0357	.0227	.0552	.0097
8RN	Eight, Rich	.0552	.0097	.0357	.0227	.0065	.0617	.0162	.0422
8RO	Eight, Rich	.0617	.0162	.0422	.0357	.0097	.0065	.0227	.0552
8RP	Eight, Rich	.0227	.0552	.0097	.0065	.0422	.0357	.0617	.0162

Table 12.1: Schedule probabilities in Rat Experiment 4

shows the bias parameters obtained from molar analysis of responses made in each condition. This shows that bias was somewhat more extreme in the 4-lever condition, but was similar with respect to the richness of the schedule. This was formally tested using the 'absolute $\mathbf{u}_i$ contrast' approach described in Chapter 9. The orthonormal contrast $\mathbf{U}$ from Equation 9.2 was again used, with a focus on $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$. Table 12.3 shows the result of the mixed-model analysis comparing these factors.

Comparing the fixed effects in this analysis, $\mathbf{u}_1$ tended to show the greatest absolute contrast ($\mu = 0.5089$) compared to $\mathbf{u}_2$ ($\mu = 0.2311$) and $\mathbf{u}_3$ ($\mu = 0.1304$). The four-lever condition elicited greater contrasts ($\mu = 0.3921$) than the eight-lever condition ($\mu = 0.1885$). The additional significant interaction term arises because $\mathbf{u}_1$ is comparatively larger in the four-lever case ($\beta = \pm 0.1090$), while $\mathbf{u}_3$ is comparatively larger in the eight-lever case ($\beta = \pm 0.1179$).

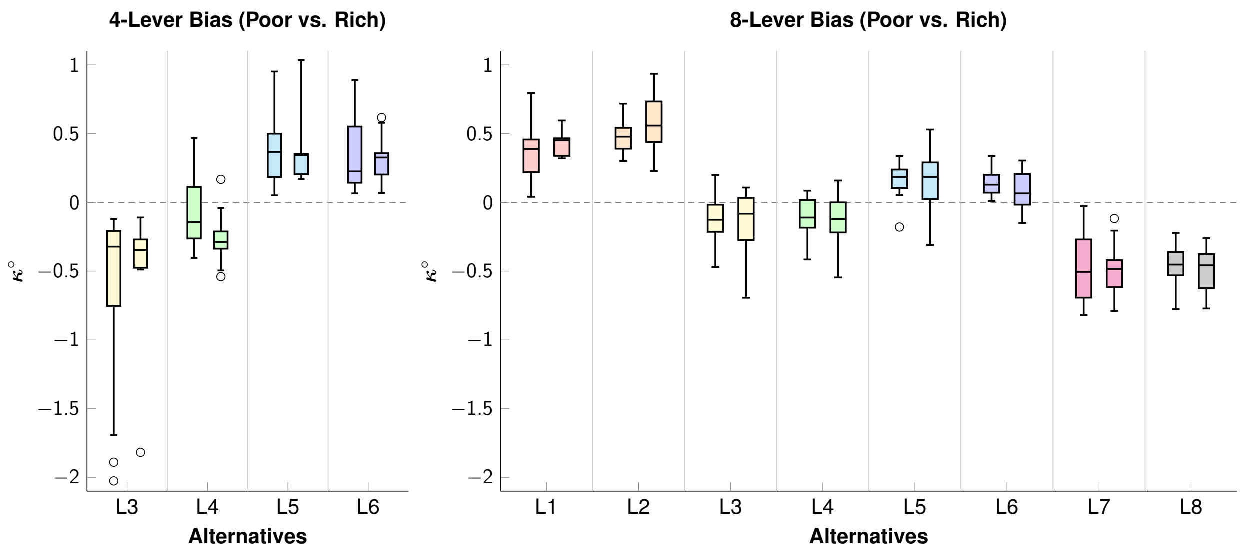

Figure 12.1: Molar CLR-transformed bias estimates (κ_i°) for Experiment 4. For each lever, the left-hand boxplot corresponds to the poor schedule, while the right-hand boxplot corresponds to the rich schedule. (Left) Bias parameter estimates in the 4-lever conditions, sorted by lever. (Right) Bias parameter estimates in the 8-lever conditions, sorted by lever.

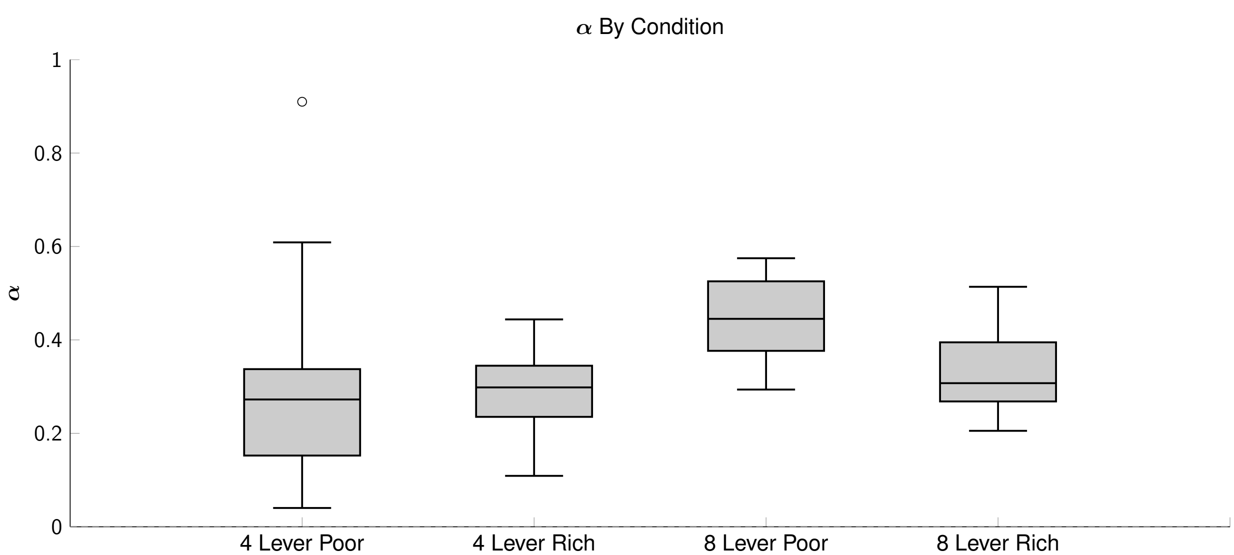

Figure 12.2: Molar sensitivity (α) estimates for each configuration in Experiment 4.

Order	Group One	Group Two	Group Three	Group Four
1	4PA	8PI	4RE	8RM
2	4PB	8PJ	4RF	8RN
3	4PC	8PK	4RG	8RO
4	4PD	8PL	4RH	8RP
5	8PL	4PD	8RP	4RH
6	8PK	4PC	8RO	4RG
7	8PJ	4PB	8RN	4RF
8	8PI	4PA	8RM	4RE
9	8RO	4RG	8PK	4PC
10	8RM	4RE	8PI	4PA
11	8RP	4RH	8PL	4PD
12	8RN	4RF	8PJ	4PB
13	4RF	8RN	4PB	8PJ
14	4RH	8RP	4PD	8PL
15	4RE	8RM	4PA	8PI
16	4RG	8RO	4PC	8PK

Table 12.2: Groupwise order of experience in Rat Experiment 4

Figure 12.2 shows the mean sensitivity per subject in each of the four conditions. Subjects in the 8-lever condition evidently showed elevated sensitivity compared to the four-lever condition, and possibly a further elevated sensitivity in the 8-lever poor schedule. To test this, a mixed model analysis of variance was performed, the results of which are presented in Table 12.4. In addition to number of levers and richness of the schedule, the order of experience was also included as a continuous factor. A significant effect of number of levers was detected, but it did not interact significantly with schedule richness, consistent with the results from previous experiments. Additionally, a substantial effect of time was observed.

Figure 12.3 shows the number of change-points detected in each phase chronologically, mixed across groups. This paints a picture consistent with the hypothesis that behavior becomes more stable over time, but glosses over effects of the number of alternatives. Because phase and config-

Factor	Sum Sq.	df	Mean Sq.	F	p
Contrast	3.6863	2	1.8431	20.76	<.0001
# of Levers	1.4919	1	1.4918	16.81	.0001
Richness	0.0001	1	0.0001	0.00	.9708
Contrast*Levers	1.2410	2	0.6205	6.99	.0013
Contrast*Richness	0.1145	2	0.0573	0.65	.5264
Levers*Richness	0.2044	1	0.2044	2.30	.1317
Subject(Random)	1.6060	11	0.1460	1.64	.0943
Error	10.9184	123	0.0888		
Total	19.2626	143			

Table 12.3: Mixed-model analysis of variance comparing absolute ILR-transformed bias contrast as a function of contrast, number of levers, and schedule richness (all fixed effect) and individual subjects (random effects).

uration are no longer nested, these data are considered analytically below.

Because Experiment 4 counterbalanced order of experience, this significant effect of time complicates subsequent visualization of the data. The data may either be sorted by condition (without regard to time), or sorted chronologically (without regard to condition). As a compromise, both methods are presented below. First, the analyses are presented chronologically. This provides insight into the consistent concern from previous experiments that some lifetime effect might be at work.

Figure 12.4 shows sensitivities estimated on a trial-by-trial basis, averaged across subjects chronologically without regard to condition (thus, intermixing both 4- and 8-lever conditions and also rich and poor schedules). Plotted in this fashion, a consistent within-phase effect is seen, embedded within a more general downward trend across the entire experiment. With each new schedule, subjects begin near $\alpha = 0$ but tend to rise to a lower mean level in each subsequent phase, as has consistently been observed in previous experiments. However, sensitivities tend to reach lower and lower maxima as trials persist. Because each phase only lasted for five sessions, it is unclear from this experiment whether subjects were merely improving more slowly, or approaching a lower asymptote; however, in light of previous experiments that made use of longer phases, it

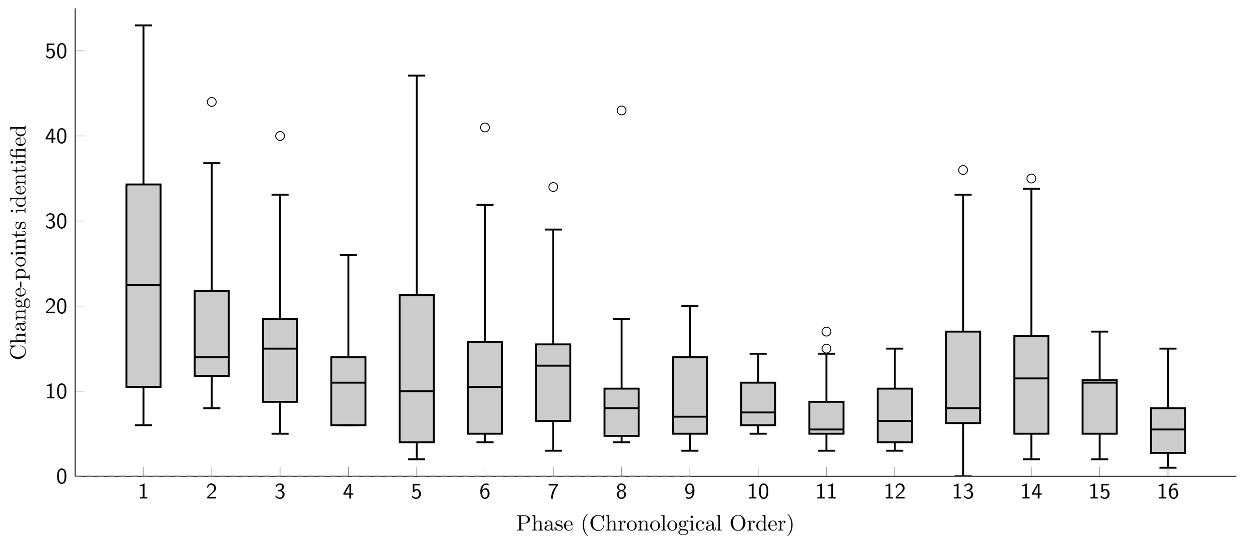

Figure 12.3: Number of change-points detected per chronological phase in Experiment 4. An outlier of 104 was also observed in the 5th phase.

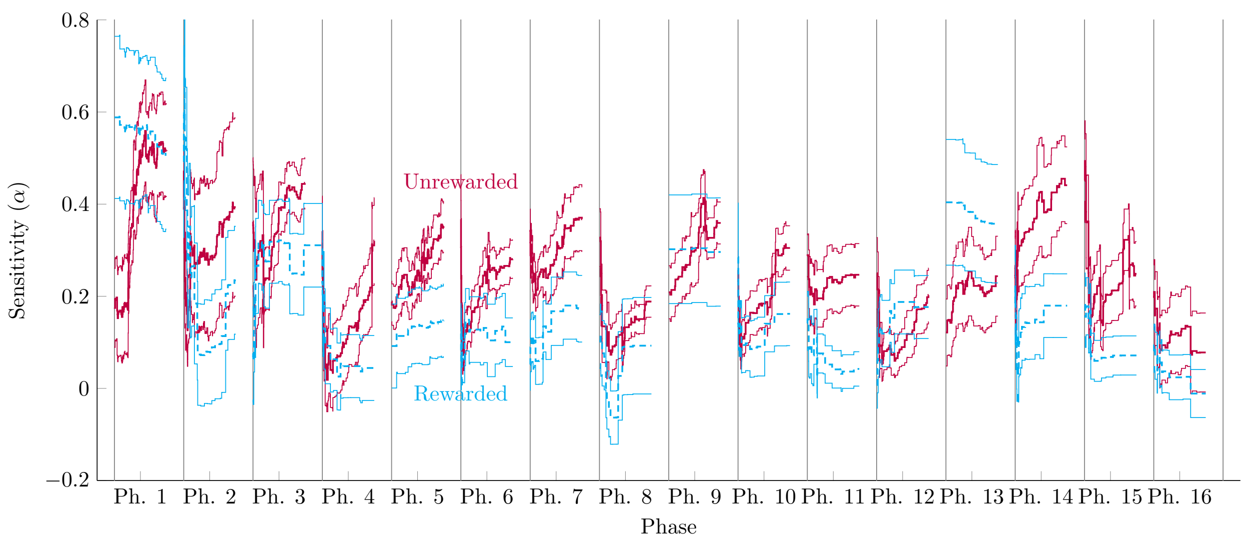

Figure 12.4: Mean sensitivity (α) estimate across subjects for each of the first 1,500 trials in each phase, averaged across groups in chronological order. The blue dashed line indicates sensitivity conditional on having just received a reward, while the red line indicates sensitivity otherwise. Thin lines indicate one standard error of the mean.

Factor	Sum Sq.	df	Mean Sq.	F	p
# of Levers	0.0906	1	0.0906	5.70	.0231
Richness	0.0469	1	0.0469	2.95	.0956
Levers*Richness	0.0189	1	0.0189	1.19	.2840
Time (Continuous)	0.2132	1	0.2132	13.41	<.0001
Subject(Random)	0.1644	11	0.0150	0.94	.5167
Error	0.5089	32			
Total	1.0429	47			

Table 12.4: Mixed-model analysis of variance comparing sensitivity estimates as a function of number of levers, schedule richness (fixed effects), dummy-coded time (continuous fixed effect) and individual subjects (random effects).

seems reasonable to assume the latter.

This gradual decrement in sensitivity did not have a substantive effect on *overall* rewards, as plotted in Figure 12.5, also plotted chronologically, across groups. Since these plots intermix the rich and poor schedules, they are plotted with particularly wide error bars; this is, of course, misleading, as the data are more properly described as bimodal (as will be shown in a subsequent figure).

A similar downward trajectory can be seen in mean entropy rates over time, plotted in Figure 12.6. In general, the complexity of observed behavior gradually decreased as a function of time. This confirms the earlier suggestion that subjects tend to grow more stereotyped as a result of either experience or aging.

A more pronounced effect is seen in the mean divergence rate, plotted in Figure 12.7. Following an early peak during their most juvenile sessions, the divergence rate quickly dropped to its standard floor value of around 0.2 bits per response, spiking upward at intervals that were marked by substantial shifts. In the transitions between phases 4 and 5 (on the one hand) and phases 12 and 13 (on the other), every animal either switched from a 4-lever to an 8-lever task, or visa versa. The smaller increase in divergence rates between phases 8 and 9 can be attributed to the month-long gap in the experiment that subjects experienced during that interval. These results are important, because they suggest that the divergence rates are not ‘burned in’ by a subject’s early experiences.

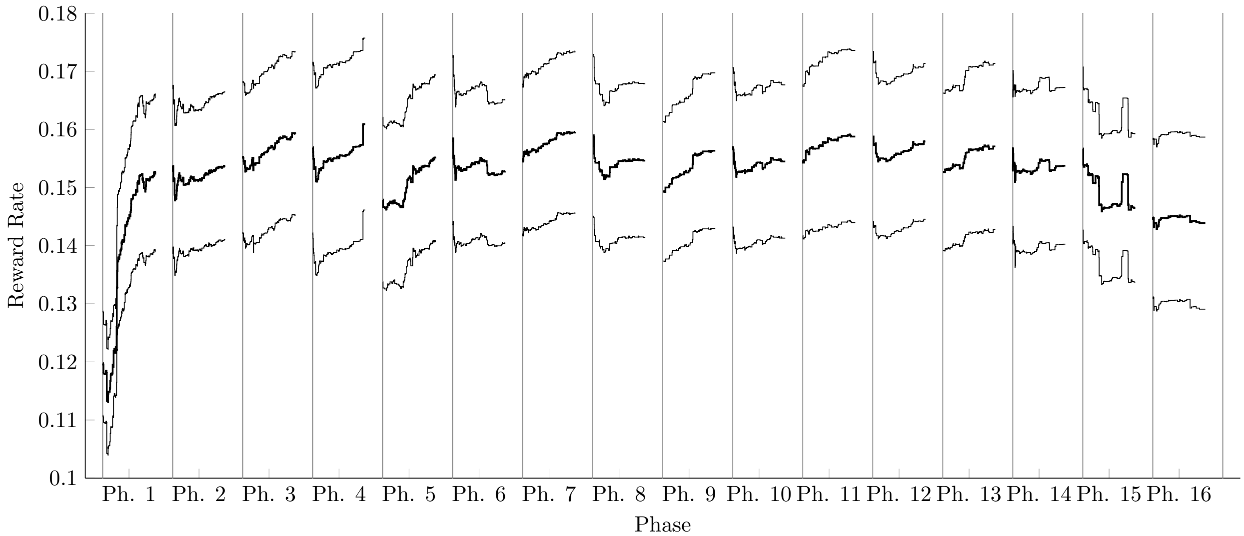

Figure 12.5: Mean rewards earned per response for each of the first 1,500 trials in each phase (as estimated based on 100,000 simulated responses using a subject's first-order contingency table at that time), averaged across groups in chronological order. Thin lines indicate one standard error of the mean.

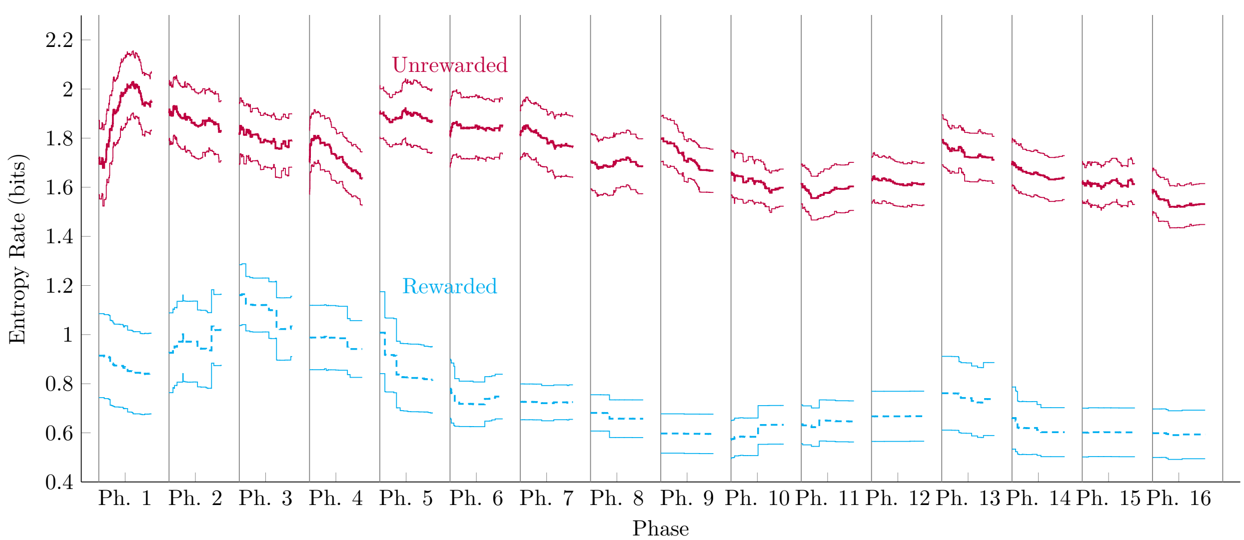

Figure 12.6: Mean entropy rate (Equation 9.5) estimated across subjects for each of the first 1,500 trials in each phase, averaged across groups in chronological order. The blue dashed line indicates sensitivity conditional on having just received a reward, while the red line indicates sensitivity otherwise. Thin lines indicate one standard error of the mean.

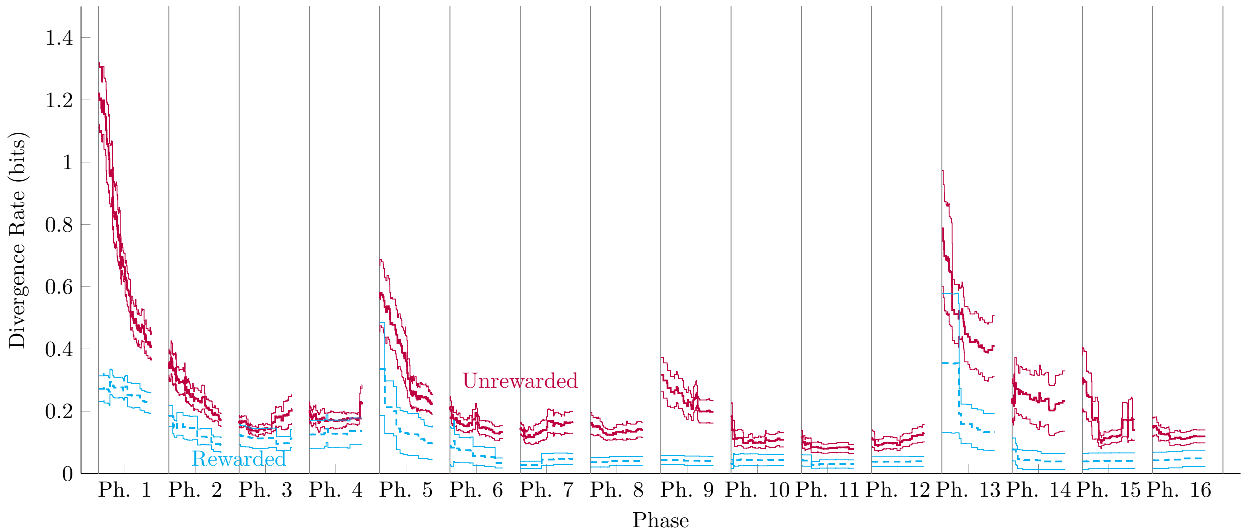

Figure 12.7: Mean divergence rate (Equation 9.7) estimate across subjects for each of the first 1,500 trials in each phase, averaged across groups in chronological order. The blue dashed line indicates sensitivity conditional on having just received a reward, while the red line indicates sensitivity otherwise. Thin lines indicate one standard error of the mean.

Instead, subjects appear to resume responding in a more sophisticated and flexible way when the physical configuration of their environment is changed. Implications of this finding are pursued in the discussion below.

Figure 12.8 plots the number of change-points identified for each phase type, disregarding chronological order. These plots are based on identical data to those in Figure 12.3, merely ordered with respect to a different variable.

The analysis of the number of detected change-points is quite complex, as the data continue to display heteroskedasticity (necessitating a non-parametric approach to the ANOVA), but at least three independent factors potentially contribute to the observed effect: The lever count, the schedule richness, and the chronological order. Thus, the data were rank-transformed, as in previous experiments, and a mixed-model ANOVA was performed that treated each of the above factors as fixed effects. Additionally, chronological order was treated as a continuous (rather than categorical) effect, as it was only partially counterbalanced. The additional effect of specific phase configuration was jointly nested within lever count and schedule richness. Lever count and schedule richness were allowed to interact. Finally, subject ID was included as a random effect, as were the interactions of

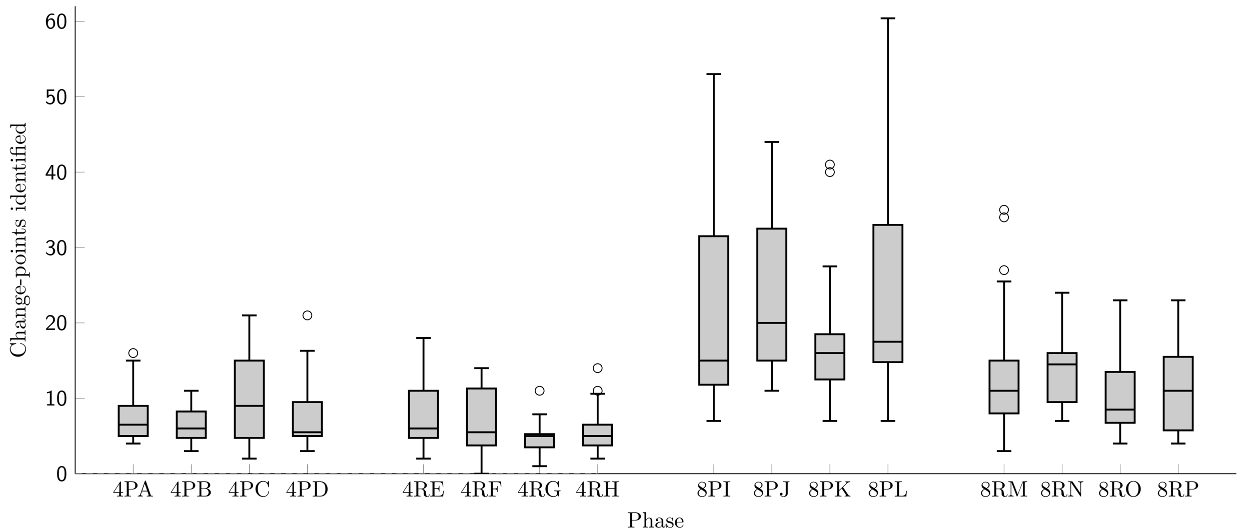

Figure 12.8: Number of change-points detected per phase type in Experiment 4. An outlier of 104 was observed in phase 8PL.

Subject ID with lever count and schedule richness. Lever count had the most substantial significant effect ($F(1,11) > 71.3, p < .0001$), followed by chronological order ($F(1,142) > 15.7, p < .0002$) and schedule richness ($F(1,10.3) > 13.8, p < .005$). The interaction of lever count and schedule richness was also significant ($F(1,142) > 5.9, p < .02$). The nested effects of phase were not significant ($F(13,142) = 1.23, p = .26$), nor was the main effect for subject ID ($F(11,13.1) = 0.53, p = .84$). However, the interactions of subject ID with lever count ($F(11,142) > 2.4, p < .01$) and richness ($F(11,142) > 2.1, p < .02$) were both significant.

Overall, this analysis paints a picture consistent with the impressions given by Figures 12.3 and 12.8: More change-points arose when more levers were available and when schedules of reward were poorer. The observed interaction of lever count and schedule richness suppressed these effects in the four-lever case while enhancing them in the eight-lever case. Throughout the experiment, change-points were detected more often early (rather than late) in training.

Averages were also computed with respect to particular phases, intermixed with respect to time. These tended to display considerably more variability, on account of the counterbalancing: Condition 4PA, for example, was experienced as the first, eight, tenth, or fifteenth phase, depending on which group a subject belonged to, so the values for any given condition are influenced to differing degrees by the effects of time.

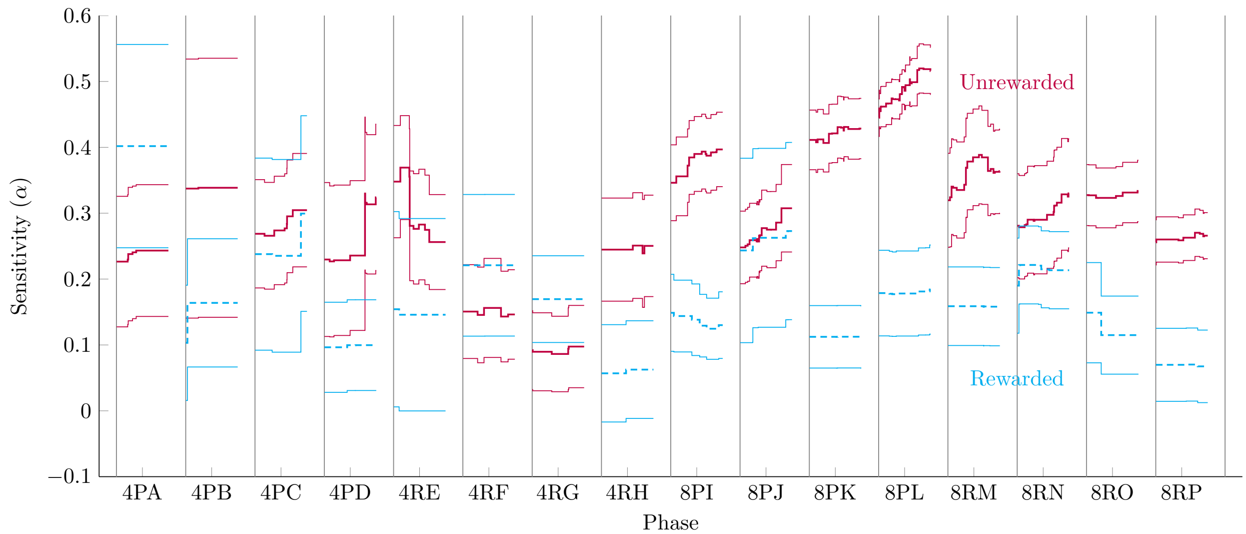

Figure 12.9: Mean sensitivity (α) estimate across subjects between trials 1,201 and 1,500 in each phase, averaged across groups in chronological order. The blue dashed line indicates sensitivity conditional on having just received a reward, while the red line indicates sensitivity otherwise. Thin lines indicate one standard error of the mean.

Figure 12.9 shows the mean sensitivity for all subjects in trials 1,201 to 1,500 of a particular condition, averaged across disparate time points. Because of the general decrease in sensitivity as a function of time, this yields much higher variability (and correspondingly larger standard errors). Nevertheless, the general pattern depicted in Figure 12.2 manifests in a similar fashion, with 8-lever sensitivity appearing generally higher than 4-lever sensitivity, and a hint (although far from definitive) that poor schedules may, under some circumstances, yield higher sensitivities.

A much more dramatic difference is observed when the mean reward rates are plotted in a similar fashion, as shown in Figure 12.10. As previously noted, the rich vs. poor schedules yielded rates of income that clustered around two different centers of mass. Incomes were highly consistent, yielding reward in a uniform range for each level of richness, regardless of the number of response alternatives.

The mean entropy rate (Figure 12.11) resembled earlier 4 vs. 8 lever comparisons: The 8-lever condition yielded higher entropy rates than the 4-lever condition, although not dramatically so. The divergence rates (Figure 12.12) in particular were fairly uniform across conditions, as the confound of order is at least somewhat counterbalanced in this plot. For both information-theoretic measures,

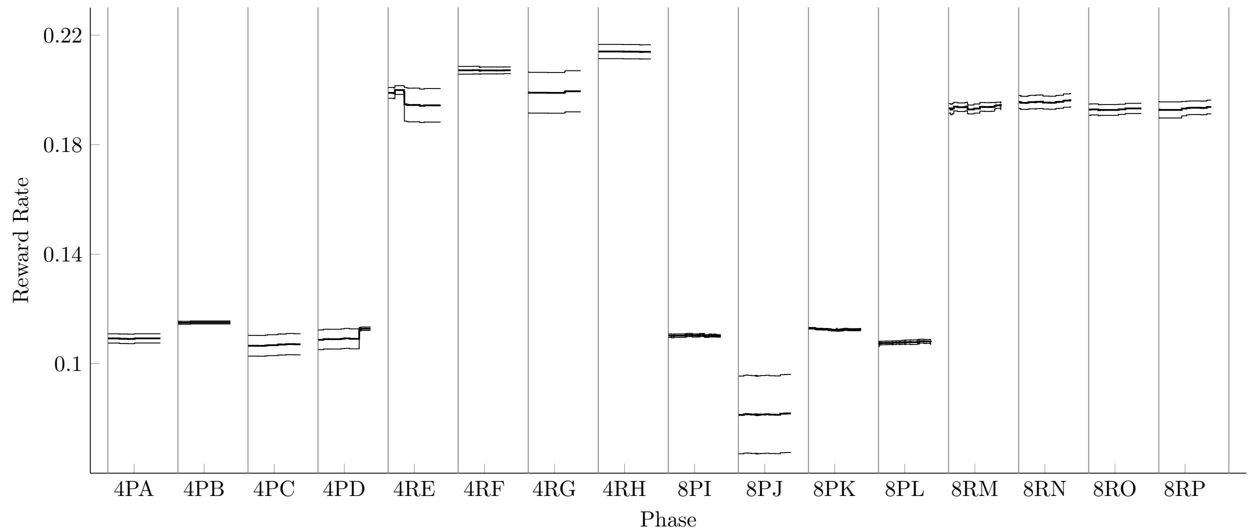

Figure 12.10: Mean rewards earned per response between trials 1,201 and 1,500 in each phase (as estimated based on 100,000 simulated responses using a subject's first-order contingency table at that time), averaged across groups in chronological order. Thin lines indicate one standard error of the mean.

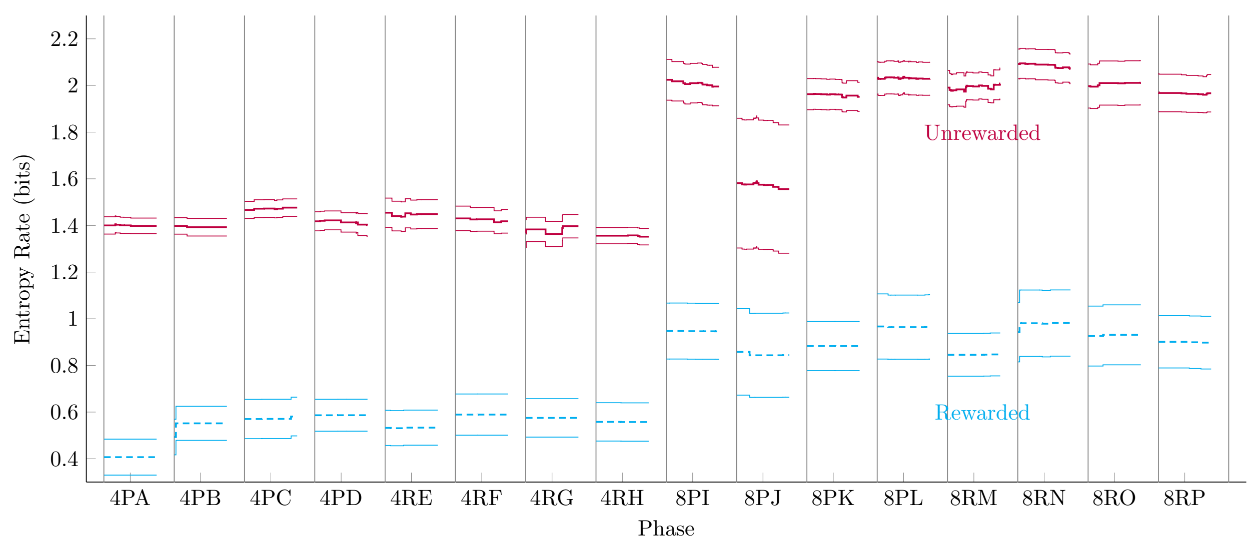

Figure 12.11: Mean entropy rate (Equation 9.5) estimated across subjects between trials 1,201 and 1,500 in each phase, averaged across groups in chronological order. The blue dashed line indicates sensitivity conditional on having just received a reward, while the red line indicates sensitivity otherwise. Thin lines indicate one standard error of the mean.

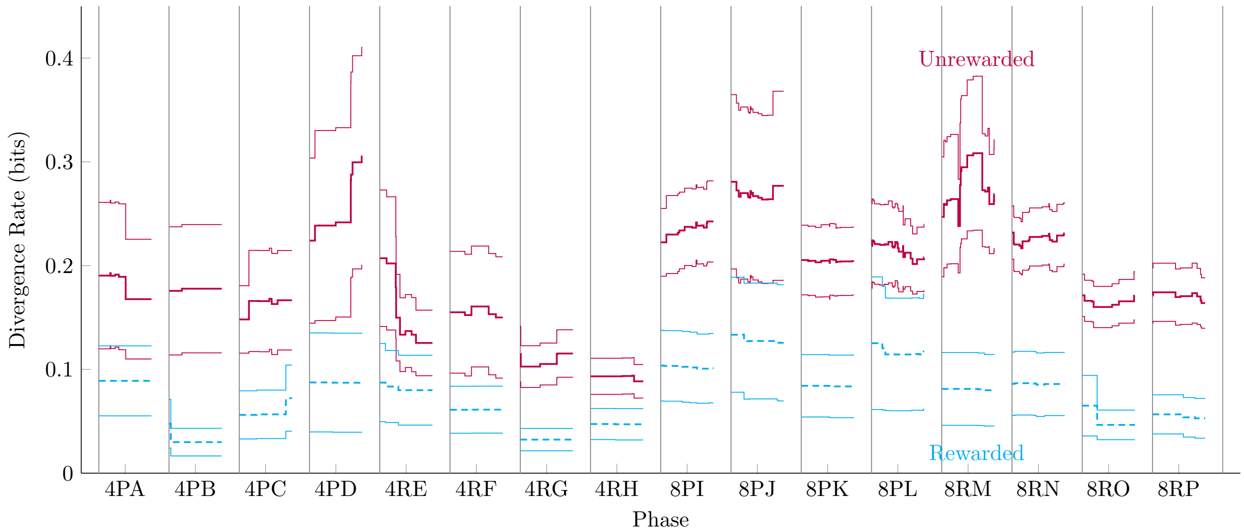

Figure 12.12: Mean divergence rate (Equation 9.7) estimate across subjects between trials 1,201 and 1,500 in each phase, averaged across groups in chronological order. The blue dashed line indicates sensitivity conditional on having just received a reward, while the red line indicates sensitivity otherwise. Thin lines indicate one standard error of the mean.

the manipulation of schedule richness did not yield results appreciably different from those observed in Experiment 1. In general, subjects could be expected to approach a divergence rate of between 0.2 and 0.3 bits per response, regardless of task complexity.

As in previous experiments, a relationship between the transition matrix $\mathcal{Q}_0$ was tested for its degree of mismatch with the schedules of reward using the $\mathcal{I}_{sland}$ metric. The means for these are plotted in Figure 12.13. For the most part, these follow a similar pattern to that reported in previous case, with the exception of two of the four-lever conditions (4PB and 4RH). As in Experiment 1, a mixed-effects regression model predicting sensitivity confirmed a continuous relationship with $\mathcal{I}_{sland}$ $(F(1,178) > 4.5, p < .04)$, as well as a fixed effect for number of levers $(F(1,178) > 5.7, p < .02)$. The effect size of these effects were $\eta^2 = .02$ and $\eta^2 = .03$, respectively, smaller than those reported in the previous experiment. The random effect for individual differences was not significant $(F(11,178) > 1.4, p = .29)$. Consequently, the previously-observed relationship appears to hold: A weak correlation between mismatch and sensitivity, and an only marginally stronger effect of number of alternatives.

In order to obtain more precise parameter estimates, an omnibus analysis of all four- and eight-

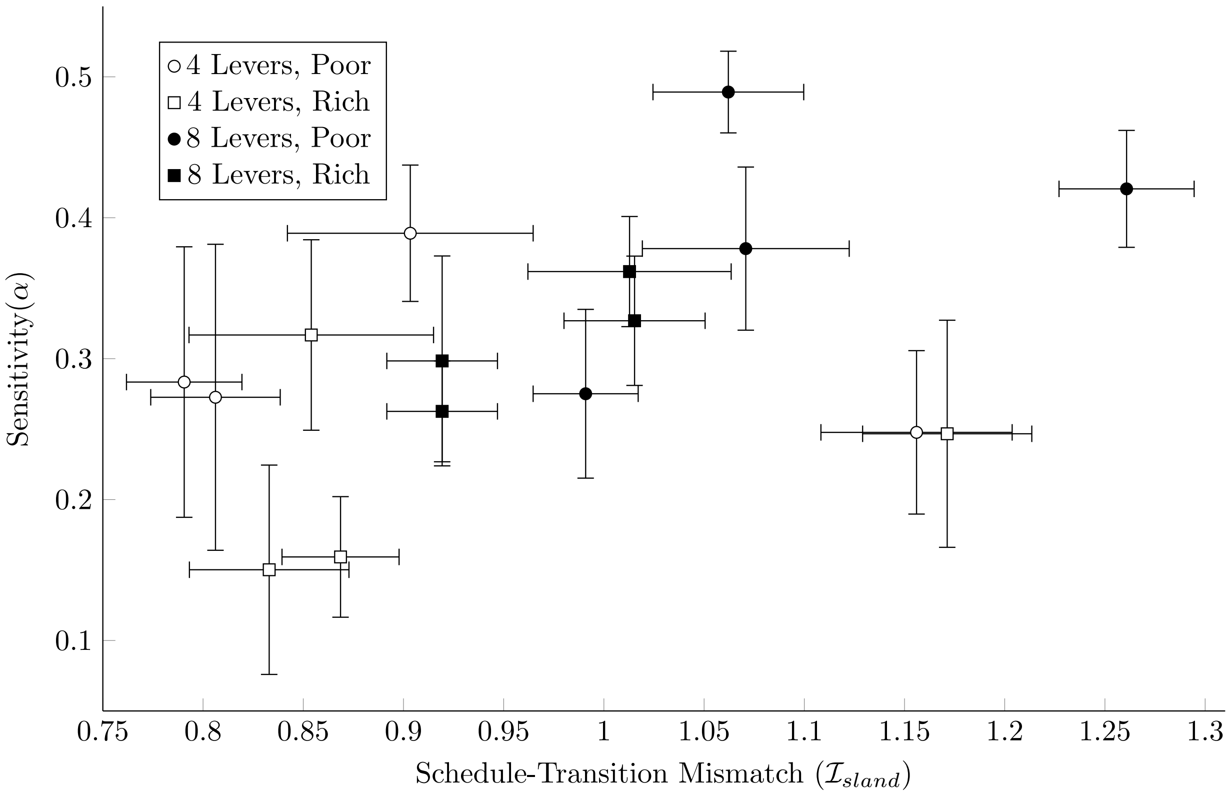

Figure 12.13: Degree of mismatch between transition matrices and schedules in Experiment 4. Four-lever phases are marked in white, whereas eight-lever phases are marked in black. Poor schedules are marked as circles, whereas rich schedules are marked as squares. Error bars correspond to standard errors.

lever phases in all four experiments was performed. The result was a pool of 594 observations (90 from Experiment 1, 312 from Experiments 2 and 3, and 192 from Experiment 4) drawn from 34 subjects. The means for each phase, as well as the best-fitting lines and confidence intervals, are plotted in Figure 12.14. The range of effects were similar to those reported in the individual analyses: A continuous relationship with $\mathcal{I}_{sland}$ ($F(1, 558) > 18.7, p < .001$), as well as a fixed effect for number of levers ($F(1, 558) > 14.1, p < .001$). The effect size of these effects were $\eta^2 = .03$ and $\eta^2 = .02$, respectively. The random effect for individual differences was not significant ($F(33, 178) = 0.99, p = .49$). Also plotted in Figure 12.14 are lines representing the best-fitting parameter estimates obtained by this mixed model analysis. These lines are estimated to have a slope of 0.294 ($SE = 0.068$) and intercepts of 0.005 for the four-lever case and 0.090 for the eight

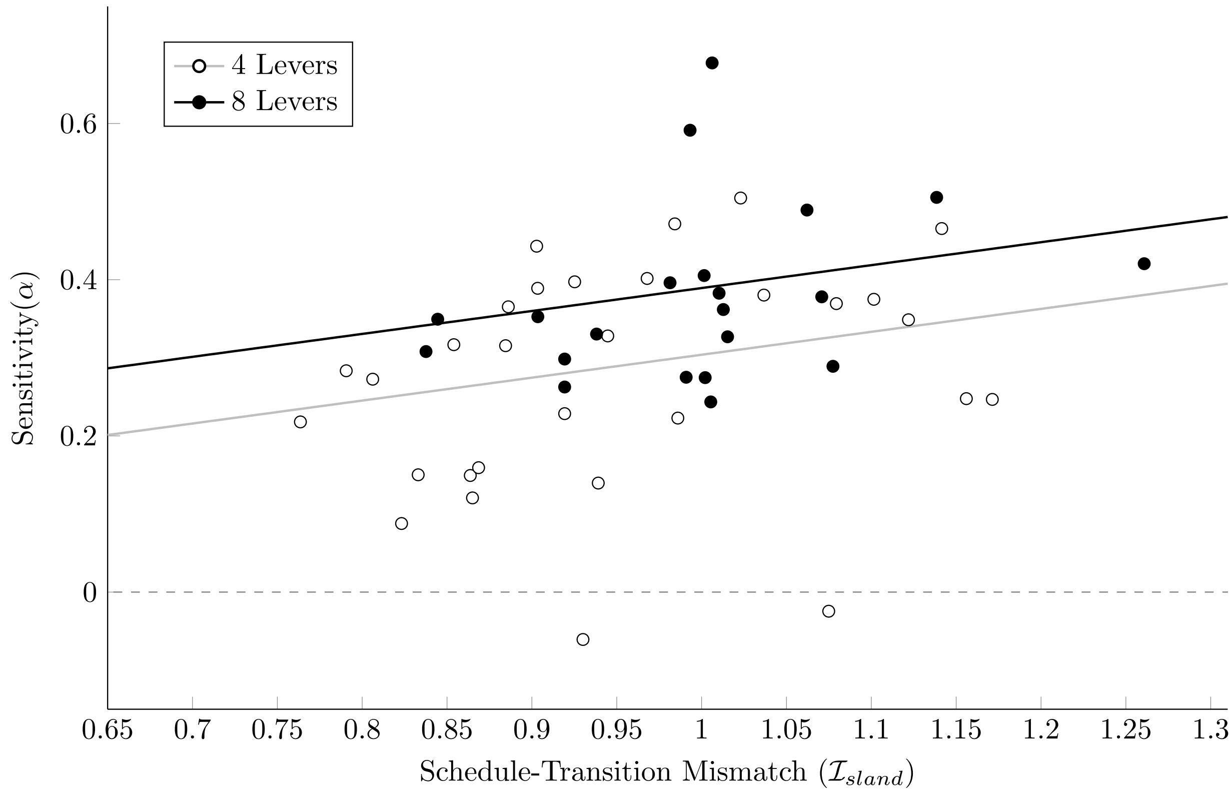

Figure 12.14: Degree of mismatch between transition matrices and schedules for four- and eight-lever phases across all experiments. Four-lever phases are marked in white, whereas eight-lever phases are marked in black. The lines represent the best-fitting parameter estimates of association for four levers (gray) and eight levers (black).

lever case ($SE = 0.11$ in both cases).

12.3 Discussion

Experiment 4 brought several potentially important confounds under direct examination. Twelve subjects were divided into four groups, who experienced both four- and eight-lever conditions in a counterbalanced order. In addition, the richness of the schedule was manipulated to compare reward rates of approximately 10% with those closer to 20%. Although these manipulations had substantive impacts on the behavior, these contributions were not enough to erase the effects observed in previous experiments.

Experiment 4 provides the study's first rigorous examination of the effect of subject age, counterbalanced for other experimental factors. In general, subjects showed a global decrease in sensitivity over time, despite also showing the more typical increase within a particular phase. As noted in the results, this is not especially surprising, given the ways in which a rodent's size and level of physical activity change as a function of normal aging. That said, it is not obvious from this experiment alone what the contributions of various possible correlates of aging might be.

A consistent effect in all experiments up to this point has been a period of adaptation during the earliest phases, as measured by the divergence rate of behavior from its eventual stationary baseline. This was seen clearly during the first phase of this experiment, but also appeared in cases where the number of alternatives changed, such as the transition from phase 4 to phase 5 (Figure 12.4). These later periods of adaptation were either *not* seen in in previous experiments, or were not as pronounced (depending on how generous one wishes to be with interpretation). Notably, the transition of phases 4 to 5 and 12 to 13 mark the only cases when a subset of subjects transitioned from 4 levers to 8 levers, and it is these six subjects who, in each of these cases is responsible for the momentary increase in the divergence rate. This can be seen by comparing the change in divergence rate immediately prior to and following the transition. In transition from phases 4 to 5, subjects going from 4 levers to 8 increased their divergence rate by a mean of 0.66 ($\sigma = 0.27$), while those going from 8 to 4 showed a mean increase of only 0.05 ($\sigma = 0.11$), a significant difference according to Wilcoxon's ranked sum test ($p < .005$). When the same subjects switched back in the transition from phases 12 to 13, it was the opposite group who went from 4 to 8, and this time it was they who show the larger difference ($\mu = 1.02, \sigma = 0.67$ going from 4 to 8, $\mu = 0.30, \sigma = 0.51$ going from 8 to 4, significantly different at $p < .04$).

This has the important implication that transitions from more complex scenarios to less complex ones do not necessarily result in the the same kind of 'representational remodeling' that switching in the other direction does, particularly when the scenario is simplified only by removing elements. Insofar as the steady-state transition matrix (from which current behavior is thought to diverge) can be seen as a kind of exploitation strategy, these transitions point to a very different sense of the explore/exploit tradeoff than is normally discussed. Because it is computationally inexpensive, steady-state transition biases should be considered to be exploitation strategies, even if they yield a wide-roving pattern of movement through the space. It is when subjects are forced to re-appraise the ways in which to make those transitions that the behavior can be seen as strategically exploratory.

Not many effects of schedule richness were seen, and these few can also be accounted for largely in terms of transitions. Since responses immediately following a reward are less sensitive to the schedule (a consistent effect across all experiments), then it stands to reason that a richer schedule will consist of a mixture of trials that skew more heavily toward those following a reward than in the case of the poorer schedule. However, while this is a plausible possibility, any net effect on behavior itself was sufficiently small that it failed to be statistically detectable in most cases. For the most part, variations in schedule richness over the range examined in these experiments did not appear to have a substantive impact on the character of behavior.

Finally, although the transition/schedule mismatch hypothesis continues to fit with the evidence obtained, the efficacy of the $\mathcal{I}_{sland}$ metric at providing a proper index for the degree of mismatch is questionable. Qualitatively, one can consider a strategy 'well-matched' to the environment if every response alternative transitions to at least one comparatively rich schedule. Characterizing this intuition quantitatively, however, is complicated both by operationalizing richness and by accommodating *all* transitions (and not merely the few that are rhetorically convenient). Producing a reasonable measure for a mismatched environment is even more problematic, because the metric must be sensitive to rich alternatives connected *only* to poor alternatives, as compared to rich alternatives connected to only a single rich alternative. Put another way, a single good alternative is 'sufficient' for the organism's next choice to be effective. A better justified metric would considerably clarify the situation in the eight-lever case, as the complexity of even first-order models are such that intuitions about them are difficult to form.

Part III

General Discussion

Chapter 13

General Discussion

Over the course of four experiments, three cohorts of rats made repeated choices among four, six, or eight levers. Each lever had a corresponding schedule of rewards, and each prepared rewards for collection in secret, only delivering the reward upon the next visit to that lever. This yielded a system of feedback, whereby the optimal strategy was not to reside exclusively at the highest-paying lever, but instead to visit each lever with a frequency proportional to the relative richness of its schedule.

In an extensive existing literature, the degree to which subjects have been able to adapt their responding to 'match' the proportions of rewards earned has been characterized using a power law, most commonly identified as Baum's 1974 *generalized matching law* (Equation 1.8). This log-odds approach provides an excellent description of overall proportions of responses made to two alternatives, and it can be extended to a broader sample space using the *barycentric matching model* (Equation 1.12, Jensen & Neuringer, 2009). However, for reasons described in the Introduction, obtaining unbiased parameter estimates for the barycentric model is not straightforward.

These parameters were obtained using compositional analysis, whose basic operations and assumptions are described in Part I. The resulting parameters for multiplicative bias ($\boldsymbol{\kappa}$) and exponential sensitivity (α) have the same interpretation as those reported in other power law models, and across both the present experiments and the re-analysis of previously published data, this methodology yields parameter estimates that are consistent (in that they display good convergence) and reasonable (in that they yield results congruent with the existing literature).

Among the most pressing questions that this approach allows us to entertain is the following:

"Do decisions among many alternatives resemble decisions among few alternatives?" Although the instinct to simplify in an experimental context is seductive, it is often the case that the jump to a multivariate problem space introduces new complications and tradeoffs that cannot be appreciated in the simple case, as is evident in arithmetic generalizations of log-odds that turn out to be inconsistent (e.g., Equation 1.10) or in ALR transformation (Equation 4.1). Using the tools of compositional analysis, this basic question was examined in detail across a series of experiments.

In every case, a surprising finding emerged: Subjects achieved higher levels of sensitivity to task demands in the eight-lever case than in the four- or six-lever cases. Two broad classes of theory have been developed to describe behavior under these conditions (i.e. cognitive vs. behavioral theories) and neither predicts this result. The central aim of this dissertation is to provide an account of why this result was obtained, and how it can be reconciled with what is already known about learning and decision making.

13.1 Response Structure in Repeated Choice

The adoption of compositional methods is not merely useful as a means of fitting the matching parameters. It also provides easy compatibility with the extensive set of tools available for analyzing stochastic systems that display the Markov property. Although the premise of "steady-state" processes has come under a degree of criticism in recent years (Fründ et al., 2011), using them to describe operant decision making has historically been deemed appropriate. This consensus is largely due to the apparent long-term behavioral stability of animals performing foraging tasks (as observed, for example, in Figure 9.6).

In this dissertation's experiments, two pieces of information proved crucial to understanding choice as an ongoing process. The first was the identity of the previous response, while the second was whether the previous response had been rewarded. The latter case is unsurprising because the food delivery trough was positioned in such a way that a behavioral bias resulted from its position. Somewhat more surprising was the level of response structure that was discovered in strings of consecutive non-rewarded responses. Subjects displayed very high levels of response structure, and these structures were very consistent, persisting throughout the multiple phases of each experiment associated with a given configuration of levers. This consistently favored set of transitions is encoded in the *default transition matrix*, denoted by $\mathcal{Q}$. As noted in the experiments, it was necessary to distinguish between $\mathcal{Q}_1$ (responses following a reward delivery) and $\mathcal{Q}_0$ (all other responses), because the delivery of food constitutes an 'interruption' that influences the behavior. However, for the sake of discussion, the fundamental premise of a default transition matrix need only concern itself with $\mathcal{Q}$.

Surprisingly, the general consistency of these patterns of transition did not correspond to insensitive behavior. Rather than being locked in a steady-state response script, subjects were able to calibrate their behavior in spite of these transition biases. If the first-order conditional probabilities of transition from Lever i to Lever j is represented as a matrix, these calibrations took the form of adjustments to only one row of the matrix at a time. Because the reward schedule is largely insensitive to the patterning of behavior (the current odds of reward on a given lever do not depend on some specific sequence of prior responses), a superficial treatment of the topic might suggest that subjects were needlessly encoding a contingency table of size $(D \times D)$ when only a single composition of size D need be represented. A subject maintaining the representation of an optimal steady-state stochastic strategy would need only 2.7 bits to encode the stationary distribution of behavior for the eight-lever schedules described in Experiment 1.

Of course, although a transition matrix may describe the outputs of a process, it does not reveal the nature of the process itself. Detailed description of behavior is, however, the first step toward characterizing the underlying process, and the results from the described experiments provide a number of clues. It is telling, for example, that the default transition matrix was highly consistent, and seemed to place considerable constraint on the resulting adaptation. It is also telling that change-point analysis determined that rows of the contingency tables changed at different times. These clues point to an underlying mechanism that engages in ongoing calibration, and does not require dramatic paradigm shifts to make adjustments.

There is no reason to suppose that an entire matrix is maintained in working memory during behavior. For example, almost every transition favored by any given animal consisted of movement to some adjacent lever along one of the chamber walls, a typical style of movement in rodents (Treit & Fundytus, 1988). In this way of thinking, an animal's 'default transitions' might result from simple decision rules, corresponding to very low levels of cognitive effort. Insofar as these transitions are governed by the physical configuration of the levers and the corresponding effort or discomfort associated with different transitions, they are not flexible in the face of changing feedback. Thus, successful adaptation requires making slight modifications to the original transition matrix. Based on the information-theoretic measure of the 'divergence rate' (Equation 9.7), which measured how much 'extra' information is needed, the ongoing cognitive load associated with this supplementary representation was very low, on the order of less than half a bit per response.

That said, it is important to distinguish between the observed transition matrices and 'fixed action patterns' in the traditional ethological sense. Although the default transition matrix may be to some extent hard-wired (insofar as it arises as a result of deeply-rooted physiological and instinctual constraints), it does not necessarily follow that the resulting behavior is strictly reflexive. Measures of α, whether they are estimated globally across large datasets (Figure 9.4) or estimated instantaneously (Figure 9.10), denote a degree of adaptation to feedback from the task. Only in cases where $\alpha \approx 0.0$ under all task conditions can the organism's behavior be characterized as reflexive. In all other cases, the factors that give rise to the default transition matrix are merely a part of a system of hierarchical control, with additional cognitive machinery shaping behavior within the confines of the organism's repertoire.

Although the heuristic of making slight modifications to an otherwise static set of transition preferences was broadly effective and cognitively inexpensive, it was vulnerable to failure states under those conditions in which relatively rich alternatives were 'surrounded' by relatively poor alternatives. Put another way, subjects did not reliably discover the most objectively efficient strategies for foraging in the operant context if those strategies involved making transitions that were not normally within their repertoire. Instead, subjects traveled between rich alternatives by way of the poor ones that lay along an indirect route, resulting in excessive visits made to these poorer alternatives.

In the present apparatus, this scenarios of 'isolation of rich alternatives' was more likely to arise when there were fewer available levers, and thus fewer candidate transitions. Consequently, it appears as though the observed differences in the global sensitivity parameter are the result of the phases whose reward schedules were severely mismatched with the default behavior of subjects. Note that even in two-alternative concurrent choice, however, severe schedule/transition mismatches are possible. For example, if subjects have a powerful bias toward switching (as rats in the present study did), then sensitivity will tend to be low if one schedule is consistently much richer than the other. Changeover delays are often used in the animal literature to reduce switching behavior (Fantino et al., 1972). Although changeover delays routinely increase sensitivity, they arguably do so artificially, because they impose specific constraints intended to shape the very behavior the procedure is intended to study. Furthermore, there is very little consistency in how changeover delays are implemented from one study to the next, making it difficult to draw any general conclusions about what their influence on the literature has been. In light of the present research, the use of changeover delays in the literature appears increasingly to paper over more complicated problems of choice as a process.

Of course, schedule/transition mismatch is not the *only* factor that impacts sensitivity. Throughout the reported results, other main effects have presented themselves (such as the systematic decrease in sensitivity seen as a result of the age of the rats). Consequently, low sensitivity should not automatically be taken as evidence of a schedule/transition mismatch. It does, however, appear to explain the systematic differences seen in the experiments in this study.

13.2 Temporal Structure in Reward Delivery

Throughout the present work, the Turn-Based Foraging paradigm (Algorithm 1) has been used to schedule rewards. As the name suggests, this approach to scheduling rewards in an uncertain manner proceeds one turn at a time, as if the subject is playing a kind of solitaire. Superficially, this appears to differentiate it from the more commonly-used Variable Interval (VI) approach, in which the uncertain process that governs reward availability is the passage of time quite separate from a subject's behavior. However, the Turn-Based Foraging paradigm has more in common with VI than it does with another common schedule, the Variable Ratio (VR). Indeed, subjects who are concurrently confronted with a Turn-Based Foraging schedule and a VI schedule are unable to distinguish the two so long as responding persists at a constant rate MacDonall (1988). Consequently, the optimal strategy in both of these cases is to engage in generalized matching (Equation 1.12).

The implication of these schedules is somewhat different, however, in the context of a process that unfolds over time. For example, if a single response is very effortful to make, but waiting is neither aversive nor costly, extended periods of waiting will be an effective strategy in the VI schedule but not in the Turn-Based Foraging schedule. Similarly, if the response alternatives are spaced far enough apart that travel time becomes a moderate investment of time (as, for example, in Aparicio & Cabrera, 2001), then the time needed to move from one alternative to another will count against the various elapsing reward intervals.

A similar scenario of some interest arises in the Turn-Based Foraging scenario when there is a schedule/transition mismatch. Consider a schedule with per-trial reward setup probabilities of [0.1, 0.0, 0.1, 0.0], with the two rich alternatives on opposing corners. Using Equation 9.1, we can conclude that alternating between the two richest options yields a continuous rate of 0.19 rewards per response, since each alternative has had two trials to set up alternatives and therefore $1 - (1 - 0.1)^2 = 0.19$. In the case where subjects instead travel in a circle around the four alternatives, each of the rich alternatives is visited half as often but is allowed two more trials to set up rewards. Thus, when selected, their probability of reward has grown to $1 - (1 - 0.1)^4 = 0.3439$, but being only selected half as often, the overall rate is 0.17195 rewards per response. The important point here is that although the 'insensitive' strategy yields a lower rate of reward, the rate is not *much* lower, because even the 'ineffective' responses to the non-rewarding alternatives serve to advance the schedule's implicit clock.

This raises an important line of theoretical inquiry: If the differences in reward rates differ so little between these two extremes of behavior, then what motive do subjects have to 'optimize' their responding, rather than merely maintaining a satisfactory income from engaging in behavior that is merely adequate? Put another way, why is any adaptation at all observed, given that the relationship between strategy and rewards seems tenuous?

There is no definitive answer to these questions (and indeed, they are well worth investigating empirically), but several possibilities nonetheless suggest themselves. One important point to recall is that subjects do not have privileged knowledge of task parameters, which might change at any time. A somewhat sensitive strategy has a reasonable chance of detecting and exploiting opportunities as they arise, whereas an insensitive one has the potential to be hugely wasteful.

The benefits of moderate sensitivity are most clearly seen when a very large number of alternatives are available *and* when considerable asymmetry exists in the distribution of the richness of reward schedules. Consider a scenario in which hundreds of response alternatives were available, all concurrently setting up rewards according to Turn-Based Foraging schedules. If most had no probability of reward, and a handful had a setup probability of 0.1 per trial, then repeatedly cycling through *all* the many choice alternatives would give the few non-zero alternatives many opportunities to make a reward available, but those few would only be visited once in a great while. The comparative benefit of confining most responses to the few alternatives that provide rewards is much greater in this case, and more appropriately resembles a naturalistic environment where most objects an animal encounters can't be converted into metabolic energy.

If the compositional approach is valid, then a symmetric model that retains a high sensitivity given an enormous range of responses is in principle possible, and such a model would then retain that sensitivity in all observed subsets of alternatives as well. This gives rise to a counterintuitive possibility: Rather than high sensitivity being a consequence of the simplicity of the two-alternative case (which, precisely because of its simplicity, does not *require* much sensitivity to obtain an effective result), it may instead be the case that high sensitivity is a characteristic of behavior ideally suited to large number of alternatives (because it permits subjects to minimize their investment in lost causes). Put another way, a behavior that is 'needlessly precise' in simple scenarios may rely on an algorithm much better suited to complex scenarios.

13.3 Bias, Sensitivity, and the Transition Matrix

It is important to distinguish the default transition matrix $\mathcal{Q}$ from the particular transitions observed in any given phase. The two are *related* but not identical. When keeping this distinction in mind, it is useful to consider the difference between an organism's 'preference' and its 'bias.'

Bias (as formally defined by $\boldsymbol{\kappa}$) corresponds to the behavior expected from the animal when all reward rates are equal. It also corresponds to the *barycenter* of all the strategies displayed by that animal. That is, if every variation in the schedule of rewards is imbalanced to an equivalent degree (as measured by the Aitchison distance in Equation 3.8), then the behaviors associated with those imbalances is expected to be centered on a compositional mean located at the coordinates $\boldsymbol{\kappa}$. However, *bias is not equivalent to preference*. The preference refers strictly to the observed proportion of behavior. This may be captured intuitively by first asking someone whether they would like to walk or to drive to a nearby location. If they reply that they would rather walk, then inform them that it is currently raining. It is quite likely that the same person would switch their answer to a preference for driving, and this is not seen as a sign of irrationality. Thus, preference is always a function of the specific conditions under which the behavior is exhibited, while bias is always an invariant quantity. The relation between the two, according to the barycentric matching model, is that preference is a consequence of three factors: the bias, the particulars of the schedule, and the sensitivity to those particulars.

In the same way, the default transition matrix $\mathcal{Q}$ is also presumed to be invariant over time. It is not merely analogous to the bias $\boldsymbol{\kappa}$, but in the present hypothesis, it represents the more fundamental process that gives rise to $\boldsymbol{\kappa}$. Interestingly, based on the reported results, $\mathcal{Q}$ also gives rise to the sensitivity parameter α, at least to some extent. Consequently, the present model is incomplete in an important way, because we have not yet identified the more fundamental process that corresponds to an organism's interpretation of feedback from the environment.

A candidate for such an approach would be a form of Bayesian updating. Under such an approach, the outcome of each response would be treated as evidence used to compute a posterior expectation of relative value. Low sensitivity might correspond to each observation yielding only a small reduction in uncertainty (a 'discrimination' problem), or might result from an especially strong prior (a 'stubbornness' problem). Specifying such a model formally, however, hinges entirely on how the matrix $\mathcal{Q}$ is represented by the organism, and whether $\mathcal{Q}$ changes over time as a function of experience. If $\mathcal{Q}$ is essentially static, then the organism must maintain some memory of the perturbation $\oplus\mathbf{q_i}$ appropriate for modifying each row i in $\mathcal{Q}$, presumably updating these perturbations when an error detect mechanism signals that the current modification is not effective.

On the other hand, if $\mathcal{Q}$ changes dynamically, it likely does so much more slowly, leading to the ongoing modification of both the short-term modifier $\mathbf{q_i}$ and a long-term conditional strategy $\mathcal{Q}_\mathbf{i}$. When both quantities are interpreted in terms of a subject's representations, correctly identifying changes in both $\mathbf{q_i}$ and $\mathcal{Q}_\mathbf{i}$ over time is a significant challenge. However, if $\mathcal{Q}_\mathbf{i}$ is taken to reflect physical constraints of the space (as opposed to the modification $\mathbf{q_i}$, which reflects the representation of the current schedule), then an experiment manipulating both the schedule and the physical environment in a controlled fashion (e.g. by changing the distance between levers, or changing the force required to depress them) could potentially track how each set of probabilities changes over time.

13.4 Adaptation & Stability in Conditional Behavior

Setting aside the appropriateness of the model in the abstract, a particularly acute problem facing the analyst is that of estimation. While there is usually little doubt as to the behavior of the organism (insofar as that behavior was operationalized), a theory reliant on $\mathcal{Q}$ for its description is only effective when a good estimate of $\mathcal{Q}$ can be obtained. The behavior, however, is in a state of constant flux, as a consequence of the organism's ongoing adaptation. Some forms of adaptation are relatively easy to track because they are connected to a measurable event in the world (as in the case of rewards delivered), while others are opaque and correspond to convenient constructs imposed on the organism that may not be well understood (such as hunger or motivation). The property that renders $\mathcal{Q}$ particularly difficult to estimate is its dimensionality: The matrix associated with n alternatives has $n(n-1)$ degrees of freedom. Given how often behavior changes, it is difficult to obtain samples of steady-state behavior that persist for long enough periods of time to be able to say with some certainty what the full matrix $\mathcal{Q}$ in that interval was.

A common approach for molar analyses is to average over many sessions, sometimes doing so over periods of weeks or months, in order to obtain estimates that converge on stable values. Such methods not only assume that behavior is entirely stable over those periods, but also make short-term changes (such as the transitions between schedules) impossible to study directly without engaging in other kinds of averaging. Thus, a method was sought that simultaneously allowed reasonable estimates to be obtained while also trying to limit such estimates to stable epochs of responding in order to avoid having to blindly assume stability.

One of the important observations in time-series analyses of the present experiment was that the conditional probabilities of response appeared stable over relatively long intervals (that is, long enough to obtain reasonable estimates), but that these conditional probabilities changed, often abruptly, at different times from one another. Each set of conditional probabilities can be thought of as a row in a transition matrix, so change-point analysis was used to identify when changes in conditional probabilities were observed, updating *only the row that changed*. This ensured that each row in the current transition matrix was estimated from as wide a window of time as possible, while also providing a principled demonstration that the behavior within that window could reasonably be considered 'stable.'

Such a method has intrinsic limits. All change-point analyses will display a tradeoff, such that smaller changes will take correspondingly more data to reliably detect. This tradeoff becomes more demanding as the number of alternatives increases: A change from $[0.4, 0.6]$ to $[0.5, 0.5]$ can be detected after a smaller number of trials than a change from $[0.1, 0.1, 0.1, 0.1, 0.6]$ to $[0.1, 0.1, 0.1, 0.2, 0.5]$. Consequently, the frequency with which change-points are detected for any given conditional probability will be lower for tasks with more alternatives. At the same time, however, more complex tasks also have more conditional probabilities to consider. Consequently, although the method is likely to break down at very large scales (e.g., dozens of response alternatives), it appears to be effective when the number of alternatives is less than ten.

Insofar as subjects are also considered to be obtaining parameter estimates about the relative rates of reward, they experience constraints very similar to those experienced by the analysts studying them. They too must use the sparse feedback of uncertain events as best they can. Under experimental conditions, the ability of both humans (Gallistel et al., 2014) and animals (Kheifets & Gallistel, 2012) have been compared to those of a 'statistically ideal observer.' Consistently, organisms in these studies display remarkably fine-tuned sensitivity of changing task demands, sometimes adjusting their behavior before having made even a single mistake. While it is safe to rule out the possibility that organisms have the ability to see into the future, even non-human organisms prove to be adept at detecting changes in their environment in the very recent past.

13.5 Cognition & Behavior: The Environment as Memory

An observer with a satirical sensibility might caricature 'cognitivism' and 'behaviorism' as lying at opposite ideological extremes. In the cognitive account, the world is remote from our experience, reaching us only through the intermediary form of 'representations.' A caricature of this position would propose that the world effectively does not exist, and that all we experience is instead a kind of 'virtual reality' constructed from whole cloth by crafty heuristics applied to unreliable sensory inputs. Meanwhile, the behaviorist perspective takes the opposite extreme: It posits that the 'self' and its corresponding feelings are the epiphenomena, and that what governs our behavior is the world. Put simply, cognitive theory explains psychological phenomena solely in terms of the contents of the mind's 'black box,' while behaviorism loudly speculates that the box itself might nearly be empty.

These caricatures are extreme and are easily dismissed as hyperbole. It is important to remember, however, that such extremity is remarkably easy to find in print. For example, the following quotation captures the common sensibility among members of the cognitive orthodoxy that the world has no 'reality' beyond the heavily processed form it takes as a representation:

> Before learning can take place, an organism must be able to perceive the world, and preferably take advantage of the fact that the information from the range of sensory channels is likely to be correlated. Objects have not only visual and spatial characteristics, but are likely also to have associated tactile features, and quite possibly to have a characteristic smell and taste. It seems likely that perceiving and integrating these various sources of information would benefit from at least a temporary form of storage, both to allow for extended processing, and also for the fact that the evidence from the various channels may not always be available simultaneously. Indeed in some cases, such as the subsequent taste of an orange, or the sound emitted by cat, information on one channel such as vision, may arrive substantially before that of others.
>
> –Baddeley (1992, p. 281)

The ultimate conclusion of Baddeley's logic is that any stimulus that elicits a behavior is encoded in some representational form, whether it be a part of the 'phonological loop' or the 'visuo-spatial sketchpad' (both of which are invoked in the article from which the above quotation is drawn).

Considered from a philosophical perspective, however, this 'representatively rich' approach to behavior clashes violently with behaviorism's extreme views about what a behavior's 'cause' might be. For example, writing on the subject of the Holocaust:

> ...it is important to emphasize that the real causes [of Hitler's order to exterminate the Jews] lay in the environment, because if we want to do anything about genocide, it is to the environment we must turn. We cannot make men stop killing each other by changing their feelings. Whatever UNESCO may say to the contrary, wars do not begin in the minds of men. The situation is much more hopeful. To prevent war we must change the environment.
>
> –B. F. Skinner, writing in Blanshard & Skinner (1967, p. 331)

Skinner's position is clearly meant to be provocative, and mainstream psychologists are quick to distance themselves from claims of this kind today. A hard-nosed analysis, however, reveals that *both* of these perspectives suffer from serious problems. Although Baddeley's statement seems the more reasonable of the two, it relies on its own long list of assumptions that, in practice, may not be tenable. Many cognitive constructs, being post-hoc inventions based mainly on observed behavior, may in reality arise from simple distributed algorithms.

The limitation of cognitive theorizing that neglects underlying mechanisms is elegantly demonstrated by 'Braitenberg vehicles' (Braitenberg, 1986). Braitenberg described a host of very simple machines composed only of sensors, actuators, and very simple wiring:

> We will talk only about machines with very simple internal structures, too simple in fact to be interesting from the point of view of mechanical or electrical engineering. Interest arises, rather, when we look at these machines or vehicles as if they were animals in a natural environment. We will be tempted, then, to use psychological language in describing their behavior. And yet we know very well that there is nothing in these vehicles that we have not put in ourselves.
>
> –Braitenberg (1986, p. 2)

The most famous of these vehicles is a box with two light sensors on the front, and a wheel on each side. Each sensor generates an analog signal linearly related to the intensity of light, and each is wired directly to one of the wheels (causing it to turn as a speed proportional to that intensity). No other interconnections are made. When such a vehicle is released into a natural environment with variable levels of light, it will engage in one of two behaviors: it will either move towards light sources or away from them, depending on which sensor is wired to which wheel. If these light sources are moving, the vehicle will immediately adjust its behavior. The temptation to attribute goals, temperaments, and even beliefs to these reactive devices is powerful. If each robot has a light source on its back, they will either gather together into chains and groups, or else scurry about trying to avoid one another. Nevertheless, all that is required for these vehicles to display complex (and even social) behavior is a suitably complex and dynamic environment. No representations or beliefs are required; indeed, in the case of Braitenberg's vehicles, not even a nervous system is necessary.

This lesson is not a new one. The same cautionary note can be seen in the original imprinting results reported by Lorenz: The 'follow' behavior of a group of goslings appears to require a sophisticated cognitive capacity, but can be described much more simply as a fixed action pattern that becomes inflexibly associated with certain kinds of moving stimuli. As in the case of the Braitenberg vehicles, appealing to the 'obvious' characteristics of an imagined mental representation may be inappropriate if a less sophisticated algorithm can yield the same behavior. Indeed, it is only their extreme simplicity that persuades us that Braitenberg vehicles are too primitive to have a 'psychology.' As a consequence, ethology was dominated by strict skepticism regarding the mental life of animals for much of the 20th century (reviewed in Klopfer, 2005).

Despite this intense skepticism, there is now ample evidence that organisms (human and otherwise) do indeed make use of cognitive machinery (reviewed in Zentall & Wasserman, 2012). However, although the radical portion of radical behaviorism has been discredited, many of the core lessons of *behaviorism* are nevertheless essential, particularly with respect to understanding the context in which behavior is observed. The heuristic proposed here, in which a default transition matrix $\mathcal{Q}$ is modified by additional feedback, can be interpreted as a hybrid of the cognitive and behaviorist viewpoints. This default matrix then acts as the foundation, and additional learning processes may provide continuous adjustments to behavior as a function of feedback from the environment.

It is not definite from the reported results, however, to what degree such a default matrix is a representation (as per Baddeley) or an emergent phenomenon (as per Braitenberg). For example, the early divergence rates plotted in Figures 9.14, 10.7, and 12.7 all suggest that each subject's initial behavior is not well represented by the default matrix, and the gradually diminishing values over the learning history correspond to better and better alignment with the default matrix. These early responses might capture a more dynamic, exploratory response that supplements the default matrix, and the diminishing divergence would reflect subjects discovering more and more efficient strategies. In this view, the default matrix is given *a priori* by the physical properties of the environment and the limitations of the subject, carrying no cognitive load beyond those of basic sensation. On the other hand, the early period may display little structure precisely because no default matrix has yet been identified. If such a matrix is instead more like a representation, then it must be built up and maintained by a gradual learning process working beneath the more rapid learning rules used to adapt to the changing schedules.

Several lines of evidence suggest that a more dynamic interpretation of $\mathcal{Q}$ is more likely to be correct. Any organism will necessarily have some physical limits, so the emergent interpretation must in some sense be true. This does not rule out the possibility of the cognitive interpretation, however, and does not signal to what degree each interpretation contributes to the final process. A clue towards resolving this ambiguity is provided by Experiment 4: Subjects making a transition from a 4-lever configuration to an 8-lever configuration show a brief rise in their divergence from the default matrix, *even when* they have an established history with the task and have encountered 8-lever scenarios previously. Subjects transitioning from 8 levers to 4 levers do not show a corresponding change in their divergence rates. These results lend some support to the cognitive interpretation, as it suggests that the 8-lever transition matrix is not 'stamped in' by the original exposure, and that a shift to a more complex task induces a brief period of exploratory behavior.

Another potentially relevant cognitive approach is that of the *mental set*, originally proposed by Jersild (1927). According to this framework, organisms not only learn particular sets of facts associated with current task demands, but also form 'mental sets' that encode how those facts relate to one another, and what the corresponding appropriate behaviors would be to exploit those facts. Although quiescent during the height of behaviorism, this approach has found new life in the context of 'task-switching' paradigms (Avdagic et al., 2013), in which subjects must make sudden transitions between tasks with different (or conflicting) demands. Task-switching paradigms have been invoked in the service of theories of hierarchical cognitive control (Kleinsorg & Heuer, 1999) and motor planning (Rogers & Monsell, 1995), as well as in studies of cognitive decline (Clapp et al., 2011) and impairment (Gu et al., 2008). The default transition matrix $\mathcal{Q}$ could provide a partial measure of a subject's mental set for the present task. Testing this hypothesis would entail training subjects with different initial conditions and then testing them with an ambiguous novel task. If $\mathcal{Q}$ primarily reflects the physical constraints of the task, then subjects should be relatively unaffected by prior training; if, however, their prior experience has resulted in procedural knowledge about the task, this would be expected to yield stable patterns of behavior heavily influenced by prior training.

13.6 Compositional Analysis in Applied Contexts

> PARADOX: A statement that reduces the matter at hand to complete obscurity while clarifying it.
>
> –Wolfe (1992, p. 237)

A substantial human literature has arisen around the idea of a ‘paradox of choice,’ wherein participants report less satisfaction and also reportedly make poorer choices when presented with large number of alternatives. Although meta-analysis suggests that the global effect size is nearly zero (Scheibehenne et al., 2010), the idea of ‘choice overload’ is a resilient trope in the choice literature. One of the seminal studies was reported by Iyengar & Lepper (2000), presented passing customers with free samples chosen either six flavors of jam or from 24 flavors. Participants were found to be less likely to approach the samples when fewer were offered, but were much more likely to purchase jam. Although other studies reporting similar results differ in their particulars, two dominant themes distinguish these experimental designs from the studies undertaken in basic research.

First, most experiments in this vein unfold without any opportunity for the participants to learn from their choices. Participants are assumed to be naïve when the experiment begins and to largely remain so when the experiment has concluded. Consequently, these are not studies of learning or of memory. If anything, they provide rough population measurements of the factors that contribute to momentary prejudices and impulsive action. Noticing that choice performance is poor when overwhelming options are coupled with no opportunity for feedback is hardly paradoxical.

Secondly, the priority of many of these experiments is to manipulate behavior, typically with a profit motive. This is not to say that all such studies have mercenary motives; indeed, a major application of this research is to identify policy approaches that save consumers money (e.g., in selecting a healthcare plan, Johnson et al., 2013). Nevertheless, the motives of applied work are not always aligned with those of basic research, as the degree of precision necessary for a result that is ‘good enough for government work’ often is not sufficient to build a robust theory describing the phenomenon.

Applying the present animal work to these applied domains, then, is less obvious than it might initially appear. Perhaps the most important rapprochement between the basic research and the applied work, then, is to understand the gap between naïve performance and expert performance. Rather than presenting participants with single choices and then leaving them to ponder the unexplored counterfactuals, such research should be of central interest to researchers who hope to discover how to train smarter decision makers (e.g., Milkman et al., 2009).

13.7 Future Directions

Further development of this broad class of problems suggests a range of different experimental possibilities. There is always a need for straightforward replication of the described effects given variation in the physical configuration, the scheduling of rewards, and the characteristics of subjects (age, gender, species), but most of these variations are obvious, and would serve mainly to validate or disconfirm the conclusions already presented. Thus, while replication and generalization experiments are important, I instead describe below several lines of inquiry that have the potential to build substantially on the existing conclusions, extending the model's generality and expanding the range of experimental preparations to which it can be applied.

13.7.1 Fixation and Symmetry

Throughout the experimental sciences, 'bias' has a bad reputation, and nearly all experimental designs implicitly seek to minimize it. This is understandable in physics and statistics (where the word usually corresponds to the misbehavior of a tool designed to provide precise measurement of some other process), but somewhat peculiar in psychology, where the 'misbehavior' is a part of the behavior being studied. Nevertheless, the study of choice and decision-making is replete with techniques for cleverly negating the effects of bias. These include the Findley switch key (Findley, 1958), the randomized telephone survey (Troldahl & Jr., 1964), the radial arm maze (Olton & Samuelson, 1976), and forced-choice adaptive procedures (Jesteadt, 1980).

The oldest and most widely used technique, however, is that of the visual fixation point (Dodge, 1900). Dating back to experimental psychology's early roots in perceptual psychophysics, fixation points remain widely used today as a way to eliminate confounding variables, particularly in the eye-tracking paradigms used in brain imaging and electrophysiology. By forcing the subject to return to some original starting point, fixation point paradigms ensure that the initial conditions of every trial are as similar as possible. Fixation procedures are not limited to eye-tracking paradigms; they may also be implemented as "start buttons" in touch-screen tasks (e.g., Basile & Hampton, 2010).

Importantly, however, no experimental procedure can account for *every* form of bias. For example, a fixation point does not prevent a subject from displaying a bias in one direction or another. Instead, it minimizes bias *arising from the prior location of gaze*. In practice, human eye movements show considerable bias independent of the content of the visual field, in a manner similar to a default transition matrix (Tatler & Vincent, 2009). Furthermore, even in cases of 'simple gaze' (were no choice is implied), participants display a bias that favors the center of the display screen (Tatler, 2007).

What we mean when we say that a fixation point task is unbiased is that it seeks to render trials *uniform in time* and *symmetrical in space*, with fixation acting as the point of reflection. Any biases displayed by subjects are then presumed to describe characteristics of the subject (memory, muscular asymmetry, etc.) and not characteristics of the experiment.

However, the present experiments suggest that not only does 'bias arising from process' contribute substantially to behavior, it may also (in the form of the default transition matrix) provide an efficient heuristic that subjects can exploit to achieve adaptive success. Because naturalistic scenarios are not governed by symmetries of experimental control, the role of process must be understood before a general theory of choice can be approached.

This observation yields two interesting, but competing hypotheses. On the one hand, because a heuristic that depends on process does poorly in cases where a mismatch between the default transition matrix and the particulars of the schedule is observed, it is reasonable to suppose that replicating the present experiments with a design with much more pronounced radial symmetry should cause this failure state to disappear. Under these circumstances, estimates of a global sensitivity parameter (as in Figure 9.4) may or may not show an effect of number of alternatives, while trial-by-trial estimates (as in Figure 9.10) may rise to more consistent asymptotic levels than were observed in the present experiments.

On the other hand, the remarkable informational efficiency of possessing a default transition matrix may break down if all conditional probabilities are forced by the task to have a similar character. We get a hint of this in the present experiments when considering the sensitivity on trials immediately following reward delivery, which is generally lower than sensitivity during the other trials. Collecting the reward effectively acts as biased fixation, as the conditional probabilities of response are much more consistent than those observed otherwise. If this reduced sensitivity is a general property of behavior (rather than being a local consequence of having just collected a reward), the sensitivity in a fixation-based paradigm might be substantially lower than in the free-responding paradigm presently used.

If using fixation to enforce response symmetry generally hurts performance, it would strengthen the claim that the process-based model of learning is the favored heuristic in the species under consideration. This in turn provides a different way in which to approach the question of what drives differences in performance between species. For example, as noted by Baum (1979), pigeons routinely display much higher sensitivities to reward than rats under similar schedules. To say that this seems reasonable for evolutionary reasons (given the pigeon's relative mobility) is vague and post-hoc, and it would be far more interesting if it were to transpire that rats favor a process-oriented heuristic while pigeons favor a more classical heuristic that weighs every alternative. While this is merely conjecture at the present time, tools have now been developed to begin to ask these questions in a more mechanism-based way.

13.7.2 Compositional Reinforcement Learning

As noted above, the working conclusion of the present research is incomplete. While a heuristic based on a default transition matrix appears to involve a minimal computational load (measured in information-theoretic terms), it remains to be determined how the learning component of such heuristic could be implemented. The challenges of working in compositional space become particularly acute in consideration of this question.

A promising possibility is afforded by renewed interest in drift diffusion models (Bogacz et al., 2006; Ratcliff & McKoon, 2008), often shortened to 'diffusion models' or called 'bounded accumulator models.' Although originally developed as physical models of particles undergoing Brownian motion, diffusion models have proved to be highly effective at modeling many other uncertain processes. At its most basic, such a model supposes that a hypothetical point lies on a number line and, following each unit of time t, is moved by an amount drawn from a Gaussian distribution. If this Gaussian is at all offset from zero, then the drift will favor one direction over the other. In all cases, the average distance from the starting point is always expected to increase over time, as a result of the random walk. At the moment the point has passed some threshold in either direction, a decision is triggered, the identity of which is determined by the direction in which the point ended up traveling. Remarkably, models of this kind not only provide good descriptions of proportions of choice, but also of reaction times and of subjective confidence ratings (Pleskac & Busemeyer, 2010; Drugowitsch et al., 2012).

Given the description above, it is not difficult to draw a parallel with conventional log-odds models. The line along which a point is thought to travel in the above description is, in most cases, directly analogous to a 1-simplex, and the additive movements of the diffusion Gaussian correspond exactly to perturbation operations (Equation 3.3) applied in log-transformed space. However, although a compositional generalization of a diffusion model appears to follow directly from this observation, the particulars are complex and have yet to be adequately studied (although see Churchland et al., 2008, for a 4-alternative model).

For example, in a two-alternative forced choice system, evidence *for* one alternative yields symmetric information relative to evidence *against* the other. This symmetry breaks down in the three-alternative case, instead yielding a simplex in which 'evidence against' is ambiguous, distributed in some fashion among the other alternatives. At least four models present themselves. On the one hand, a strictly additive model (using only the 'evidence for') would yield a pattern of diffusion resulting from the sum of many random vectors (one per alternative, radially symmetrical with respect to a common barycenter). A similar (but not identical) model would arise making use of only 'evidence against.' A more complex model would include both evidence for *and* evidence against, doing so in potentially uneven amounts. Finally, a model could be employed that simply encodes the drift as a multivariate normal distribution (which may or may not display sphericity).

Determining which of these possibilities is the 'correct' model is a substantial challenge, and depends critically on the process for which the model seeks to provide a description. For example, it is rather common for mathematically sophisticated researchers to point out that because certain statistical computations are computationally costly or intractable, their theoretical generality may be suspect (e.g., Courville et al., 2006). This is a naïve view of the brain, however, which ignores the massively parallel way in which information is processed. To take an extreme case, it is clearly impossible for a photon to 'perform computations,' because photons have no moving parts;

nevertheless, the path integrals followed by photons appear to calculate integrals of stupefying complexity at the speed of light. This capability is merely an emergent property of their interaction with innumerable other quantum fields. It is not the *photon* that does the computation, but rather that the computation is a *consequence* of the system. It is on this premise that 'quantum computing' could theoretically be performed (Kok et al., 2007). In a similar fashion, distributed computing by vast networks can represent intractable probability distributions to arbitrary precision, provided the desired distribution is the consequence of the relevant network. Put more succinctly: Neurons need not encode a value of π to display behavior well-described by a Gaussian distribution, even though an analyst must know π to high precision to evaluate the equation.

Although a variety of competing models of neuronal behavior have been put forward in recent years, it can be shown that many of them are reducible in one form or another to diffusion models (Roxin & Ledberg, 2008). Meta-analysis and simulation using these various models both suggest that although no clear winner has emerged as providing the best model of behavior, it is nevertheless likely that decisions arise from processes characterized by competition among alternatives (Teodorescu & Usher, 2013). Classical econometric model, in which the value of each alternative is evaluated independently and the results are compared, yield less precise predictions than those in which alternatives must in some sense fight among themselves, and many of the models that display this property can be described in terms of diffusion.

13.7.3 Fixed Transitions, Random Subjects

One of the analytic problems with the described method for estimating a transition matrix is that it relies on obtaining a representative sample of behavior across at least moderately counterbalanced conditions. This renders the approach untenable when fewer than several thousand responses collected over at least three (but ideally four or five) different schedules. While this is a reasonable approach in the context of a long-term behavioral experiment using animals, it isn't as feasible with (for example) a pool of human participants.

Fortunately, the default transition matrices observed in behavior throughout these experiments displayed a great deal of consistency. Although differing somewhat in the particulars, a general pattern of preferred transitions was observed, arising in unsurprising ways given a behavioral phenotype common among rodents. In cases where consistency of this kind is seen across subjects, mixed models can be used to pool data from multiple subjects to identify the 'fixed effects' common across the groups, while simultaneously taking the 'random effects' of individual differences into account (Fitzmaurice et al., 2004).

The fixed-effects default transition matrix that would result from such an analysis requires a slightly different theoretical stance than does a matrix estimated for a particular individual. Although a transition matrix pooled across subjects could still be used in a calculation of the divergence rate, this divergence can be expected to display inflated values, because its failure to account for the additional divergence between the fixed-effect matrix and the random effect of each subject's individual default transition matrix (which may not be estimable given a small behavioral sample).

Provided this caveat is properly taken into account, a mixed-model approach could be used to obtain effective population estimates from experiments with much shorter durations. Such a method could, in turn, permit a larger subject pool to be examined in a cost-effective way, and to achieve much higher levels of statistical power when making claims about population-level effect (Button et al., 2013).

13.7.4 Compositional Prospect Theory

Given the possibility of the mixed-model analytic method described above, one of the first domains in which the compositional model can be extended is that of human decision-making. As reviewed in the introduction, the compositional paradigm has extensive implications for the theories of utility and prospects that arose in the second half of the 20th century.

The first among these is that strict preference and stochastic dominance are neither theoretically reasonable nor empirically justified in the study of the actual behaviors emitted by organisms. A probabilistic model that permits 'preference reversals' is entirely within the reach of econometric theory, provided the analyst can endure a paradigm shift away from treating axioms of rationality as prescriptive. The driving force behind behavioral economics has been this recognition that decision making is heuristic rather than mathematically rational. What remains to be done is to achieve greater rapprochement between behavioral economics and ethology, recognizing both that animal decision making is likely more sophisticated than has been given credit by economists, and human decision-making less deliberative. However, it is not enough to attack the obvious deficiencies of classical economics, which are widely agreed upon. The contribution of a compositional methodology to this discussion is a set of constraints that change the manner in which feedback from the environment must be understood.

For example, a model built using compositional tools is almost obligated to consider sources of gain independently of sources of loss, with the integration of the two coming only at the end of the process. The reason for this is simple: The closure constraint is only consistently defined when all items share the same sign (that is, when they are either all positive or all negative). Closure, as defined, does not coherently accommodate components that make up a 'negative percentage' of the total. Furthermore, operations like perturbation and powering cannot ever move a coordinate within the simplicial geometry outside of its confines. Consequently, the most reasonable approach is to work in two isolated simplexes (one for gains and one for losses), working to yield a final composite value. The resulting composite value for overall gains can then be compared to the composite value for losses, these presumably lying on an unbounded ratio scale.

Additionally, a third class of options is compositionally distinct from gains and losses: Those whose value is precisely zero. Closure formally requires that all components in a composition be non-zero, as any structural zeros trap a point within a facet of any larger compositional structure. It therefore stands to reason that zeros are also dealt with during the final accounting stage.

This change in the order of operations has the potential to explain many 'irrational' behaviors in a fashion that is more mathematically consistent and analytically tractable than existing methods, and that scales to large numbers of alternatives more efficiently. Not only does the segregation of gains and losses appear discontinuous (as has been suggested by many experimental results), but that discontinuity may manifest as a considerable gulf in cases where outcomes with a value of zero are likely to occur.

The most direct tests of the compositional hypothesis take the form of 'multi-armed bandit' paradigms (Robbins, 1952), which have been studied both in strictly abstract terms (S. Zhang & Lee, 2010) and experimentally (Steyvers et al., 2009). These tasks are essentially identical to concurrent choice paradigms, and may be implemented using nearly identical schedules (e.g., Kangas et al., 2009). Of course, it is important to distinguish the temporal dynamics of the schedule involved (such as whether rewards are 'held' in secret or not; Jensen & Neuringer, 2008); in almost all studies of human decision-making, rewards are strictly probabilistic and, as such, the optimal strategy is exclusive selection of the best alternative. Although they are understudied in human paradigms, schedules that involve the continuous balancing of all response alternatives are very likely to yield more interesting results than the 'exclusive-choice' optimizations typical in the decision-making literature. The Turn-Based Foraging paradigm described by Algorithm 1, in particular, requires participants to make use of mixed strategies, while also giving rise to reward hazard functions that facilitate rigorous quantitative analysis.

13.7.5 The Psychophysics of Probability and Value

Although much of the literature on human decision making (particularly in econometrics) has confined itself to rationality axioms inspired by the von Neumann-Morgenstern theorems, there is a rich literature suggesting that, in addition to primary sensory modality, many of our other subjective experiences display 'psychophysical' properties. Brunswik was among the first to suggest that subjective probability might be experientially important in animal behavior (Brunswik, 1939) and in human psychophysics (Brunswik & Herma, 1951), whereas Attneave (1953) was among the first to examine human self-report of the 'feeling' of a probability in terms of a power law, consistent with contemporaneous work on power-law psychophysical functions (Stevens, 1957).

Despite this early work, there was little cross-talk between econometricians and psychometricians (although see Edwards, 1962; Galanter, 1962) prior to the advent of prospect theory. However, the rise of behavioral economics has yielded considerable work seeking to identify psychophysical functions of value and probability. These experimental investigations initially focused on econometric themes such as consumer spending (Christensen, 1989) and the characterization of utility functions (Tversky & Fox, 1995; Brandstätter & Brandstätter, 1996), but more recent work has drawn from a broader domain of experimental paradigms, including behavior analysis (Gallistel et al., 2014), motor coordination (Trommershäuser et al., 2008) and neuroscience (Vlaev et al., 2011). The sophistication of these results, as well as their inconsistency with rationality axioms, has led to something of a crisis of confidence in experimental econometrics, resulting in an increased emphasis on rigorous experimental design (Angrist & Pischke, 2010).

With respect to the problem of subjective value, the heart of the matter is discriminability. As Figure 1.3 makes clear, models that assume participants can perfectly discriminate between exact outcomes yield aberrant predictions, while models that can reasonably represent ambiguity do not.

Although cognitive models like prospect theory have proposed that gambles undergo an "editing" process prior to valuation, there is unlikely to be a universal algorithm across all individuals that governs all such editing. The dream of uncovering the one true editing algorithm is almost certainly a Quixotic one.

An alternative approach is to develop a more flexible descriptive model that permits the assessment of differing degrees of discriminability. Reiger & Wang (2008) describe a method by which continuous probability distributions (like that in Figure 1.3, bottom) can be interpreted, but also suggest a computationally straightforward process of converting traditional discrete gambles into continuous functions using a form of kernel density estimation. In their framework, the probability density function associated with each discrete outcome is converted from a mixture of degenerate point distributions into a set of overlapping step functions, each of which has a width parameter ϵ, as in kernel density estimation (Silverman, 1986). Their logic is that because similar outcomes cannot be distinguished from one another, they should be pooled into a common cluster whose density is characterized by the sum of their individual step functions. The closure constraint may then be applied with a normalizing factor equal not to the sum of the transformed probabilities, but instead to the integral of the resulting continuous distribution. Figure 13.1 (top) displays this process, using a uniform kernel with width $\epsilon = 1.5$, as applied to the subdivided gamble previously depicted in Figure 1.3.

This approach yields two problems. The first is that there is considerable evidence that numerical judgments generally fall on a ratio scale, such that the contrast between \$1 and \$10 is considerably more salient than the contrast between \$74 and \$83. The second is that because their domain of outcomes necessarily covers the range from $-\infty$ to ∞, their approach must either abandon the distinction between gains and losses, *or* predict paradoxical behavior. In the latter case, the paradox arises whenever an objective outcome has a value of \$0, as in Figure 13.1. An outcome of \$0 is converted by the kernel to cover the range from $-\epsilon$ to ϵ, leading to a mixture of both positive and negative expectation. If, consistent with the empirical evidence, losses are more salient than gains, then the expectation from a \$0 outcome should integrate to a negative value. Put another way, an objective outcome of \$0 will only have a subjective value of \$0 under this paradigm if gains and losses are exactly symmetrical.

A better approach reflects the psychophysical uncertainty of numerical judgments, as well as the

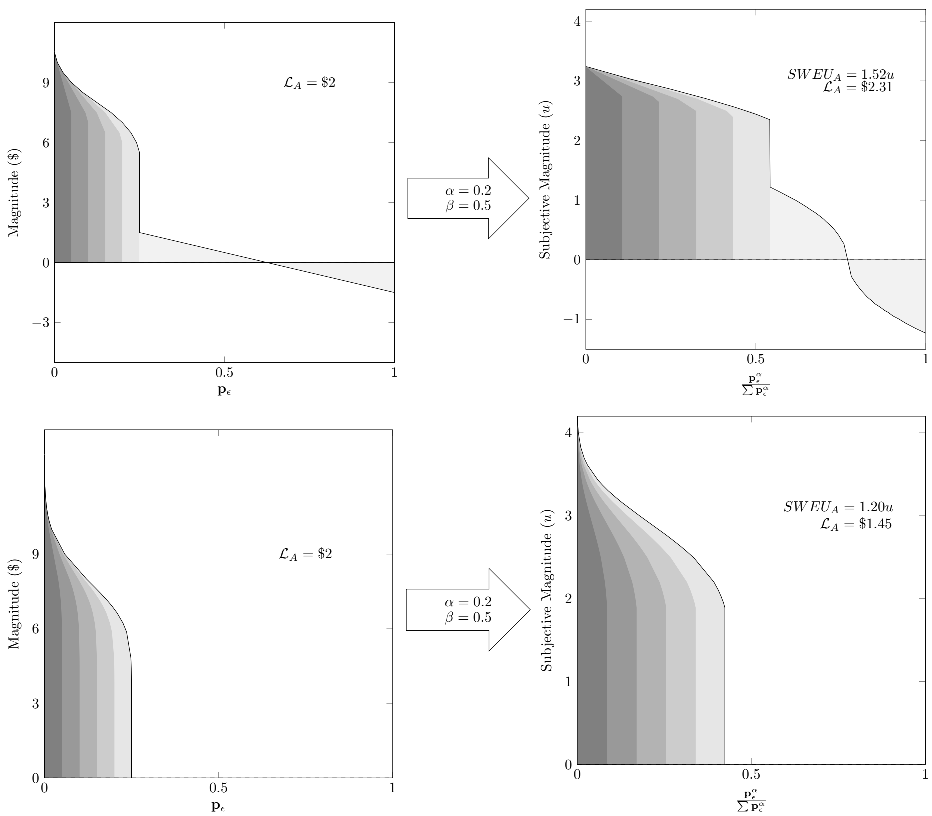

Figure 13.1: The subdivided lottery in Figure 1.3 is modified in two different ways that reflect the uncertainty of mental representation. In one case (top), a continuous distribution is build using the strategy outlined by Reiger & Wang (2008), using uniform kernels with a width of $\epsilon = 3$. This yields much less distorted subjective expectations. In a psychophysically-motivated account (bottom), the five non-\$0 lotteries are combined by summing log-normal distributions with a consistent coefficient of variation equal to 0.15 (generalization), and the resulting composite is treated as categorically distinct from the \$0 outcome (discrimination). The resulting influence of α and β yields a prediction that much more reasonably approximates the non-subdivided expectation in Figure 1.3 (top).

distinction between *generalization* and *discrimination*. In Figure 13.1, the five non-$0 outcomes are each converted to log-normal distributions all having a coefficient of variation $c_v = 0.15$, consistent with the psychophysics of numerosity reported by Whalen et al. (1999). These are then summed to produce a joint density estimate, in a manner similar to the strategy outlined by Reiger & Wang (2008). Because the resulting density function is a coherent cluster, it is generalized into its own discrete category, discriminated from the degenerate probability density at $0. Thus, when α and β are applied, they are applied to the clustered outcomes. This yields a very reasonable prediction of expected value, that is broadly consistent with Figure 1.3 (top).

While this general strategy goes a long way toward providing a template for characterizing the editing process of different individuals, it suffers from several procedural ambiguities. For example, what constitutes 'discriminably distinct' groupings of outcomes? One extreme would be to always use three distinct categories: gains, losses, and 'zeroes' (this last as a way to accommodate the interpretation of value belonging on a ratio scale). A compromise might be to set a threshold for shared variance by harkening to set theory (e.g. in the style of Bush & Mosteller, 1951).

A subtle consideration lies in the relationship between the coefficient of variation $c_v = 0.15$ and the model exponents α and β. Reiger & Wang take the position that uncertainty regarding the density function of expectation should be further modified by the distortion of probabilities; that is, that ϵ and α interact directly. It may be the case, however, that α (which governs the subdivision of distinct outcomes in Figures 1.2 and 1.3) acts independently of a coefficient of variation c_v (which governs the interpretation of ambiguity). Given the research on 'ambiguity aversion' (Fox & Tversky, 1995), treating α and c_v as parameters whose impact is orthogonal to one another is appealing.

Tentatively, the following model emerges:

$$\begin{aligned}
\mathcal{L}_A^+ &= \left(\sum_{i \in D^+} \frac{\left(p_{i,A}^+\right)^{\alpha^+}}{\sum_{j \in D^+} \left(p_{j,A}^+\right)^{\alpha^+}} \cdot \int_0^\infty Log\mathcal{N}\left(x | \mu = \log\left(M_{i,A}\right), \sigma = c_v\right) \cdot x^{\beta^+} dx \right)^{\frac{1}{\beta^+}} \\
\mathcal{L}_A^- &= -\left(\sum_{i \in D^-} \frac{\left(p_{i,A}^-\right)^{\alpha^-}}{\sum_{j \in D^-} \left(p_{j,A}^-\right)^{\alpha^-}} \cdot \int_0^\infty Log\mathcal{N}\left(x | \mu = \log\left(-M_{i,A}\right), \sigma = c_v\right) \cdot x^{\beta^-} dx \right)^{\frac{1}{\beta^-}} \\
\mathcal{L}_A &= \frac{\mathcal{L}_A^+ \cdot \left(\sum_{i \in D^+} p_i^+\right)^\alpha + \mathcal{L}_A^- \cdot \left(\sum_{i \in D^-} p_i^-\right)^\alpha}{\left(\sum_{i \in D^+} p_i^+\right)^\alpha + \left(\sum_{i \in D^-} p_i^-\right)^\alpha + \left(p^\emptyset\right)^\alpha}
\end{aligned} \tag{13.1}$$

Here, D^+ corresponds to the subset of outcomes with a positive expectation and D^- to those with a negative outcome. This leaves $p^\emptyset$, which corresponds to the probability of an outcome with zero value. At a minimum, this model has three free parameters: α, β, and c_v; in that case, $\alpha = \alpha^+ = \alpha^-$ and $\beta^+ = \beta^-$. In order for gains and losses to be accounted for differently, either α or β must be split into distinct parameters. In the interest of parsimony, β appears to be a better candidate, because it has no involvement with $p^\emptyset$.

Although Equation 13.1 appears unwieldy, its structure is both mathematically and psychologically justified. Because its treatment of probabilities consistently respects the closure constraint, its final calculated value is represented in the units of the objective problem. Furthermore, representing outcome uncertainty in terms of log-normal distributions with a uniform coefficient of variation c_v is consistent with reports that implicit judgments of numerical quantities display scale-invariant error.

13.7.6 Moving Away From Strict Preference Ordering

Despite emphasizing a violation of the rationality axioms of expected utility, the various forms of prospect theory are primarily concerned with the 'normative' behavior of a 'rational' agent (Tversky & Kahneman, 1981). Although this approach introduced a level of subjectivity not considered by classical economic models, it nevertheless assumed a strict preference ordering, such that 'rational' behavior consists of always choosing the alternative with the higher subjective value. In practice, choice behavior of both by human participants and in animal models appears to be fundamentally (and functionally) probabilistic (Neuringer & Jensen, 2012). Thus, in addition to parameters that characterize sensitivity to probability (α), value (β), and ambiguity (c_v), a decision function $\mathcal{D}()$ is needed to describe the odds structure for choice frequency of each lottery. Given a set of X lotteries, the decision to choose lottery A would depend on the function:

$$\mathcal{D}(A) = P(A|\mathcal{L}_A, \mathcal{L}_B, \cdots, \mathcal{L}_X) \tag{13.2}$$

Typically, $\mathcal{D}()$ takes the form of a probability distribution.

Among the most successful and enduring alternatives for $\mathcal{D}()$ is Luce's choice axiom (Luce, 1959, 1977), which takes the following form:

$$\mathcal{D}(A) = \frac{\mathcal{L}_A}{\sum_{i \in X} \mathcal{L}_i} \tag{13.3}$$

Of the decision functions in the literature, the above is the most psychologically motivated; despite this, it has received very little direct empirical testing (Stott, 2006). This function is trivially identical to the closure constraint (Equation 1.1), providing further justification for the compositional approach.

Since Equation 13.3 asserts that the decision function is effectively compositional, following that logic to its conclusion leads to the following straightforward formulation for weight and value:

$$\mathbf{B}^* = \mathbf{k}^* + \delta \cdot \mathbf{L}^* \tag{13.4}$$

Here, the probability of a behavior B_i in the set $\mathbf{B} = [B_1, \cdots, B_n]$ is a function of the set of subjective values $\mathbf{L} = [\mathcal{L}_A, \cdots, \mathcal{L}_n]$. This decision function has a vector of free parameters associated with biases not otherwise accounted for $\mathbf{k}$, and a sensitivity to lottery values δ.

13.7.7 Navigational Constraint as Compositional Choice

Although multi-armed bandit tasks have a long history of study, the simplicity that makes them analytically appealing also undermines their face validity. In particular, multi-armed bandit tasks are memoryless, resetting conditions to the same starting point on every trial in exactly the fashion described in the above section on fixation. Consequently, much could be gleaned from an experimental approach that explored choice into a less artificial context. At the same time, that very artificiality is what controls for extraneous factors and so permits parametric analysis.

In light of this, a broad class of video games provide a potent middle ground. Although games are often highly abstract and artificial in the degree to which they constrain choices, most also permit players to pursue the game's constrained objectives in a fashion that can be examined at multiple levels of analysis. Consider, for example, the classic arcade game Pac-Man (Iwatani, 1980). In each of the game's levels, the player has a high-level strategic goal of visiting each point in a maze, consuming 'pellets' to mark progress. However, because a number of antagonists ('ghosts') also patrol the maze, the player must also play the game in a tactical fashion, adapting to short-term cues. At its most global, Pac-Man is about adaptive strategies for tracing least-distance paths through complex environments while avoiding dynamic obstacles; at its most reductionistic, it's about one turn in the labyrinth after another.

A simplified version of this kind of labyrinthine game could be used to implement a version of the experiments presented to rats in the present study. Participants would be tasked with collecting as many rewards as possible by traveling through a labyrinth. Participants are made aware periodically that new rewards are available, but are not told where those rewards have been hidden. Furthermore, although participants can always see the layout of the labyrinth, they can only see rewards which lie within the line-of-sight of their in-game avatar. This forces the player to continuously explore and forage. Different unlabeled 'regions' of the maze would have differing probabilities of having rewards available, and the layout of the maze could be manipulated to vary the ease of transitions between those regions.

Such an experimental design would allow the process-based hypothesis put forward here to be directly tested in a human participant pool. The prediction is that humans, acting under time pressure, would fail to optimize their time spent exploring each region of the maze in the cases where islands of relatively rich reward are relatively isolated, having only poor regions for neighbors, but would perform considerably better in cases where every region had at least one profitable neighboring region.

Despite a flood of tools for making simple games, relatively few modern studies have made use of maze paradigms, unless the word 'maze' is reduced to the absolute minimal sense of a T-intersection or a figure 8 (although see Simon & Daw, 2011). Presumably, human decision making routinely operates at a much higher level of abstraction, and only tasks with enough complexity to allow decision-making processes to unfold will provide a reasonable test of whether a process-based hypothesis is plausible outside a narrow laboratory context.

Chapter 14

Conclusions

No quantitative problem in science can be satisfactorily resolved without understanding the sample space within which the data reside. Throughout this dissertation, I have argued that a great deal can be achieved by modeling repeated choices in compositional terms. An interdisciplinary review of repeated choice suggests that different lines of empirical evidence converge toward compositions as a possible unifying framework. If successful, this line of inquiry points toward heuristics for subjective value and subjective probability that underly basic decision making in multiple species.

The systematic treatment of compositional methods in Part I reveals them to be suitable to a wide range of problems. Rather than suggesting a single model, compositions permit families of models to be specified. The most fundamental of these models is the *barycentric matching model* (Equations 1.12, 5.1), which provided a highly successful model of the relationship between response proportions and scheduled rewards in the experiments performed for this project (Chapters 9 through 12). In the event that different response alternatives are difficult to discriminate from one another, the model's complexity need only be slightly increased to consider different contrasts between the alternatives, as described for the re-analysis of existing data in Chapter 8.

Although the barycentric matching model provides a good description of molar behavior, that description was surprising: Subjects adjusted their behavior in a manner that was more sensitive to the schedule parameters when there were *more* response alternatives. To explain this finding, the compositional framework was further extended to include change-point analysis over conditional time series.

From this analysis, a working hypothesis was developed that has substantial implications. Subjects familiar with the task display very substantial response structure regardless of the schedule parameters, a set of conditional probabilities called the 'default transition matrix.' Once this pattern is stable, subjects make only small deviations from it in response to feedback from the environment, incurring a very low computational load. This suggests that the so-called 'paradox of choice' (whereby complex tasks result in cognitive overload) does not apply to this kind of repeated choice. Indeed, rodents using this strategy displayed considerable adaptability.

At the same time, strategies that rely on computationally inexpensive deviations from a default transition matrix are fragile when the best alternatives become isolated from one anther. In the present study, this led to an effect quite reversed from the conventional idea of cognitive overload, in which it was the simpler tasks that engendered insensitive behavior. Understanding compositional choice as an ongoing process provides an explanation for this surprising result, and points to a variety of experimental designs that will further illuminate these effects.

In the end, both experimentalists and theorists are advised to take Einstein's immortal advice into careful consideration:

> It can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience.
>
> –Albert Einstein, the Herbert Spencer Lecture, Oxford University, June 10, 1933.

Crucially, the goal here is not to simplify the datum of experience! Although experimental control often demands that we begin with a simple case, it is only by gradually considering more and more complex data that we are able to determine whether our theory is *too* simple. In this respect, the study of choice is no different, and must build on existing methods to test theory against tasks of increasing complexity. Doing so will not only yield new and interesting phenomena, but it is the only way in which the disparate threads of choice and decision making can hope to be woven into a more integrated and complete theory.

Part IV

Bibliography

References

Abdellaoui, M., Vossmann, F., & Weber, M. (2005). Choice-based elicitation and decomposition of decision weights for gains and losses under uncertainty. *Management Science*, *51*, 1384–1399.

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, *117*, 288–318.

Aitchison, J. (1986). *The statistical analysis of compositional data.* Chapman & Hall. Reprinted in 2003 with additional material by The Blackburn Press.

Aitchison, J., & Egozcue, J. J. (2005). Compositional data analysis: Where are we and where should we be heading? *Mathematical Geology*, *37*, 829–850.

Aitchison, J., & Shen, S. M. (1980). Logistic-normal distributions: Some properties and uses. *Biometrika*, *67*, 261–272.

Allais, P. M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine. *Econometrica*, *21*, 503–546.

Alsop, B., & Elliffe, D. (1988). Concurrent-schedule performance: Effects of relative and overall reinforcer rate. *Journal of the Experimental Analysis of Behavior*, *49*, 21–36.

Anderson, G. (1996). Nonparametric tests of stochastic dominance in income distributions. *Econometrica*, *64*, 1183–1193.

Angrist, J. D., & Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, *24*, 3–30.

Aparicio, C. F., & Cabrera, F. (2001). Choice with multiple alternatives: The barrier choice paradigm. *Mexican Journal of Behavior Analysis*, *27*, 97–118.

Arrow, K. J. (1951). Alternative approaches to the theory of choice in risk-taking situations. *Econometrica*, *19*, 404–437.

Attneave, F. (1953). Psychological probability as a function of experience frequency. *Journal of Experimental Psychology*, *46*, 81–86.

Augustin, T. (2008). Stevens' power law and the problem of meaningfulness. *Acta Psychologica*, *128*, 176–185.

Avdagic, E., Jensen, G., Altschul, D., & Terrace, H. S. (2013). Rapid cognitive flexibility of rhesus macaques performing psychophysical task-switching. *Animal Cognition*, *October*.

Baddeley, A. (1992). Working memory: The interface between memory and cognition. *Journal of Cognitive Neuroscience*, *4*, 281–288.

Barnard, G. A. (1949). Statistical inference. *Journal of the Royal Statistical Society, Series B (Methodological)*, *11*, 115–149.

Basile, B. M., & Hampton, R. R. (2010). Rhesus monkeys (*macaca mulatta*) show robust primacy and recency in memory for lists from small, but not large, image sets. *Behavioural Processes*, *83*, 183–190.

Baum, W. M. (1974). On two types of deviation from the matching law: bias and undermatching. *Journal of the Experimental Analysis of Behavior*, *22*, 231–242.

Baum, W. M. (1979). Matching, undermatching, and overmatching in studies of choice. *Journal of the Experimental Analysis of Behavior*, *32*, 269–281.

Baum, W. M. (1993). Performance on ratio and interval schedules of reinforcement: Data and theory. *Journal of the Experimental Analysis of Behavior*, *59*, 245–265.

Baum, W. M. (2012). Rethinking reinforcement: Allocation, induction, and contingency. *Journal of the Experimental Analysis of Behavior*, *97*, 101–124.

Baum, W. M., & Aparicio, C. F. (1999). Optimality and concurrent variable-interval variable-ratio schedules. *Journal of the Experimental Analysis of Behavior*, *71*, 75–89.

Baum, W. M., & Davison, M. (2004). Choice in a variable environment: Visit patterns in the dynamics of choice. *Journal of the Experimental Analysis of Behavior*, *81*, 85–127.

Bauman, Y. (2003). Maniw's ten principles of economics, translated. *Annals of Improbable Research*, *9*, 4–8.

Berkson, J. (1944). Application of the logistic function to bio-assay. *Journal of the American Statistical Association*, *39*, 357–365.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception and Psychophysics*, *8*, 279–286.

Bernoulli, D. (1738|1954). Exposition of a new theory on the measurement of risk (L. Sommer, Trans.). *Econometrica*, *22*, 23–36.

Birnbaum, M. H. (2008). New paradoxes in risky decision making. *Psychological Review*, *115*, 463–501.

Birnbaum, M. H., & Navarrette, J. B. (1998). Testing descriptive utility theories: Violations of stochastic dominance and cumulative independence. *Journal of risk and uncertainty*, *17*, 49–78.

Birnbaum, M. H., Patton, J. N., & Lott, M. K. (1999). Evidence against rank-dependent utility theories: Tests of cumulative independence, interval independence, stochastic dominance, and transitivity. *Organizational behavior and human decision processes*, *77*, 44–83.

Blanshard, B., & Skinner, B. F. (1967). The problem of consciousness – a debate. *Philosophy and Phenomenological Research*, *27*, 317–337.

Blavatskyy, P. R. (2011). A model of probabilistic choice satisfying first-order stochastic dominance. *Management Science*, *57*, 542–548.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alterantive forced-choice tasks. *Psychological Review*, *113*, 700–765.

Braitenberg, V. (1986). *Vehicles: Experiments in synthetic psychology.* MIT Press.

Brandstätter, E., & Brandstätter, H. (1996). What's money worth? determinans of the subjective value of money. *Journal of Economic Psychology*, *17*, 443–464.

Brooks, R. R. W., & Purdie-Vaughns, V. (2007). The supermodular architecture of inclusion. *Harvard Journal of Law & Gender*, *30*, 379–387.

Brunswik, E. (1939). Probability as a determiner of rat behavior. *Journal of Experimental Psychology*, *25*, 175–197.

Brunswik, E., & Herma, H. (1951). Probability learning of perceptual cues in the establishment of a weight illusion. *Journal of Experimental Psychology*, *41*, 281–290.

Buhusi, C. B., & Meck, W. H. (2005). What makes us tick? functional and neura; mechanisms of interval timing. *Nature Reviews Neuroscience*, *6*, 755–765.

Bush, R. R., & Mosteller, F. (1951). A model for stimulus generalization and discrimination. *Psychological Review*, *58*, 413–423.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*, 365–376.

Butts, A. M. (1938). *Scrabble.* [board game].

Changizi, M. A. (2008). Harnessing vision for computation. *Perception*, *37*, 1131–1134.

Changizi, M. A. (2011). *Harnessed: How language and music mimicked nature and transformed ape to man.* BenBella Books.

Changizi, N., & Hamarneh, G. (2010). Probabilistic multi-shape representation using an isometric log-ratio mapping. In T. Jiang, N. Navab, J. Pluim, & M. Viergever (Eds.), *Medical image computing and computer-assisted intervention – miccai 2010* (pp. 563–570). Springer Berlin / Heidelberg.

Charness, G., Karni, E., & Levin, D. (2007). Individual and group decision making under risk: An experimental study of bayesian updating and violations of first-order stochastic dominance. *Journal of Risk and Uncertainty*, *35*, 129–148.

Chrisman, N. R. (1998). Rethinking levels of measurement for cartography. *Cartography and Geographic Information Science*, *25*, 231–242.

Christensen, C. (1989). The psychophysics of spending. *Journal of Behavioral Decision Making*, *2*, 69–80.

Churchland, A. K., Kiani, R., & Shadlen, M. N. (2008). Decision-making with multiple alternatives. *Nature Neuroscience*, *11*, 693–702.

Clapp, W. C., Rubens, M. T., Sabharwal, J., & Grazzaley, A. (2011). Deficit in switching between functional brain networks underlies the impact of multitasking on working memory in older adults. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 7212–7217.

Cobb, C. W., & Douglas, P. H. (1928). A theory of production. *American Economic Review*, *18*, 139–165.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *American Statistician*, *35*, 124–129.

Cording, J. R., McLean, A. P., & Grace, R. C. (2011). Testing the linearity and independence assumptions of the generalized matching law for reinforcer magnitude: A residual meta-analysis. *Behavioural Processes*, *87*, 64–70.

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences*, *10*, 294–300.

Davison, M. (1996). Stimulus effects on behavior allocation in three-alternative choice. *Journal of the Experimental Analysis of Behavior*, *66*, 149–168.

Davison, M., & Jenkins, P. E. (1985). Stimulus discriminability, contingency discriminability, and schedule performance. *Animal Learning & Behavior*, *13*, 77–84.

Davison, M., & McCarthy, D. (1988). *The matching law: A research review.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Davison, M., & McCarthy, D. (1994). Effects of the discriminability of alternatives in three-alternative concurrent-schedule performance. *Journal of the Experimental Analysis of Behavior*, *61*, 45–63.

Davison, M., & Nevin, J. A. (1999). Stimuli, reinforcers, and behavior: An integration. *Journal of the Experimental Analysis of Behavior*, *71*, 439–482.

de Villiers, P. A., & Herrnstein, R. J. (1976). Toward a law of response strength. *Psychological Bulletin*, *83*, 1131–1153.

Deicidue, E., Schmidt, U., & Wakker, P. P. (2004). The utility of gambling reconsidered. *Journal of Risk and Uncertainty*, *23*, 241–259.

Dodge, R. (1900). Visual perception during eye movement. *Psychological Review*, *7*, 454–465.

Drugowitsch, J., Moteno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *Journal of Neuroscience*, *32*, 3612–3628.

Edwards, W. (1962). Subjective probabilities inferred from decisions. *Psychological Review*, *69*, 109–135.

Egozcue, J. J., Barceló-Vidal, C., Martín-Fernández, J. A., Jarauta-Bragulat, E., Díaz-Barrero, J. L., & Mateu-Figueras, G. (2011). Elements of simplicial linear algebra and geometry. In V. Pawlowsky-Glahn & A. Buccianti (Eds.), *Compositional data analysis: Theory and applications* (pp. 141–157). Wiley.

Egozcue, J. J., Daunis-I-Estradella, J., Pawlowsky-Glahn, V., Hron, K., & Filzmoser, P. (2012). Simplicial regression. the normal model. *Journal of Applied Probability and Statistics*, *6*, 87–108.

Egozcue, J. J., & Pawlowsky-Glahn, V. (2005). Groups of parts and their balances in compositional data analysis. *Mathematical Geology*, *37*, 795–828.

Egozcue, J. J., & Pawlowsky-Glahn, V. (2006). Simplicial geometry for compositional data. *Geological Society, London, Special Publications*, *264*, 145–159.

Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., & Barceló-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, *35*, 279–300.

Elliffe, D., & Alsop, B. (1996). Concurrent choice: Effects of overall reinforcer rate and the temporal distribution of reinforcers. *Journal of the Experimental Analysis of Behavior*, *65*, 445–463.

Elliffe, D., Davison, N., & Landon, J. (2008). Relative reinforcer rates and magnitudes do not control concurrent choice independently. *Journal of the Experimental Analysis of Behavior*, *90*, 169–185.

Eppler, M. J., & Mengis, J. (2004). The concept of information overload: A review of literature from organization science, accounting, marketing, mis, and related disciplines. *The Information Society*, *20*, 325–344.

Fantino, E., Squires, N., Delbrück, N., & Peterson, C. (1972). Choice behavior and the accessibility of the reinforcer. *Journal of the Experimental Analysis of Behavior*, *18*, 35–43.

Findley, J. D. (1958). Preference and switching under concurrent scheduling. *Journal of the Experimental Analysis of Behavior*, *1*, 123–144.

Fiorina, M. P. (1971). A note on probability matching and rational choice. *Behavioral Science*, *16*, 158–166.

Fishburn, P. C. (1989). Retrospective on the utility theory of von neumann and morgenstern. *Journal of Risk and Uncertainty*, *2*, 127–158.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004). *Applied longitudinal analysis.* Hoboken, NJ: Wiley-Interscience.

Fleshler, M., & Hoffman, H. S. (1962). A progression for generating variable-interval schedules. *Journal of the Experimental Analysis of Behavior*, *5*, 529–530.

Fox, C. R., & Tversky, A. (1995). Ambiguity aversion and comparative ignorance. *The Quarterly Journal of Economics*, *110*, 585–603.

Fründ, I., Haenel, N. V., & Wichmann, F. A. (2011). Inference for psychometric functions in the presence of nonstationary behavior. *Journal of Vision*, *11*, 1–19.

Galanter, E. (1962). The direct measurement of utility and subjective probability. *American Journal of Psychology*, *75*, 208–220.

Gallistel, C. R., & King, A. P. (2010). *Memory and the computational brain.* Wiley-Blackwell.

Gallistel, C. R., Krishan, M., Liu, Y., Miller, R., & Latham, P. E. (2014). The perception of probability. *Psychological Review*, *121*, 96–123.

Goldstein, W. M., & Einhorn, H. J. (1987). Expression theory and the preference reversal phenomena. *Psychological Review*, *94*, 236–254.

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, *38*, 129–166.

Gu, B. M., Park, J. Y., Kang, D. H., Lee, S. J., Yoo, S. Y., Jo, H. J., ... Kwon, J. S. (2008). Neural correlates of cognitive flexibility during task-switching in obsessive-compulsive disorder. *Brain*, *131*, 155–164.

Hausman, J., & McFadden, D. (1984). Specification tests for the multinomial logit model. *Econometrica*, *52*, 1219–1240.

Hefferon, J. (2013). *Linear algebra.* (Available: `http://joshua.smcvt.edu/linearalgebra/`)

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of response. *Journal of the Experimental Analysis of Behavior*, *4*, 267–272.

Herrnstein, R. J. (1990). Rational choice theory. *American Psychologist*, *45*, 356–367.

Herrnstein, R. J. (1997). *The matching law: Papers in psychology and economics* (H. Rachlin & D. I. Laibson, Eds.). New York, NY: Russell Sage Foundation.

Hertwig, R., & Erev, I. (2009). The description-experience gap in risky choice. *Trends in Cognitive Sciences*, *13*, 517–523.

Hineline, P. N. (2001). Beyond the molar-molecular distinction: We need multiscaled analyses. *Journal of the Experimental Analysis of Behavior*, *75*, 342–347.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). Wiley.

Hurvich, C. M., & Tsai, C. L. (1989). Regression and times series model selection in small samples. *Biometrika*, *76*, 297–307.

Iwatani, T. (1980). *Pac-man.* [arcade game].

Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, *79*, 995–1006.

Jacquez, J. A., Mather, F. A., & Crawford, C. R. (1968). Linear regression with non-constant, unknown error variances: Sampling experiments with least squares, weighted least squares, and maximum likelihood estimators. *Biometrics*, *24*, 607–626.

Jeffreys, H. (1946). An invariant form for the prior probability in estimating problems. *Proceedings of the Royal Society of London, Series A, Mathematical and Physical Sciences*, *186*, 453–461.

Jensen, G. (2014). Compositions and their application to the analysis of choice. *Journal of the Experimental Analysis of Behavior*.

Jensen, G. (2014 (projected)). Closed-form estimation of multiple change-point models. *PeerJ PrePrints*, *1*, e90v3.

Jensen, G., & Neuringer, A. (2008). Choice as a function of reinforcer “hold”: From probability learning to concurrent reinforcement. *Journal of Experimental Psychology: Animal Behavior Processes*, *34*, 437–460.

Jensen, G., & Neuringer, A. (2009). Barycentric extension of generalized matching. *Journal of the Experimental Analysis of Behavior*, *92*, 139–159.

Jensen, G., Ward, R., & Balsam, P. (2013). Information: Theory, brain, and behavior. *Journal of the Experimental Analysis of Behavior*, *100*, 408–431.

Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology*, *89*, 1–82.

Jesteadt, W. (1980). An adaptive procedure for subjective judgments. *Perception and Psychophysics*, *28*, 85–88.

Johnson, E. J., Hassin, R., Baker, T., Bajger, A. T., & Treuer, G. (2013). Can consumers make affordable care affordable? the value of choice architecture. *PLOS ONE*, *8*, e81521.

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). Springer.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of risk under choice. *Econometrica*, *47*, 263–292.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*, 341–350.

Kangas, B. D., Berry, M. S., Cassidy, R. N., Dallery, J., Vaidya, M., & Hackenberg, T. D. (2009). Concurrent performance in a three-alternative choice situation: Response allocation in a rock/paper/scissors game. *Behavioural Processes*, *82*, 164–172.

Karmarkar, U. S. (1978). Subjectively weighted utility: A descriptive extension of the expected utility model. *Organizational behavior and human performance*, *21*, 61–72.

Kheifets, A., & Gallistel, C. R. (2012). Adapting without reinforcement. *Communicative & Integrative Biology*, *5*, 531–533.

Kleinsorg, T., & Heuer, H. (1999). Hierarchical switching in a multi-dimensional task space. *Psychological Research*, *62*, 300–312.

Klopfer, P. H. (2005). Animal cognition and the new anthropomorphism. *International Journal of Comparative Psychology*, *18*, 202–206.

Knetsch, J. L. (1989). The endowment effect and evidence of nonreversible indifference curves. *American Economic Review*, *79*, 1277–1284.

Knill, D. C., & Pouget, A. (2004). The bayesian brain: The role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, *27*, 712–719.

Kok, P., Munro, W. J., Nemoto, K., Ralph, T. C., Dowling, J. P., & Milburn, G. J. (2007). Linear optical quantum computing with photonic qubits. *Reviews of Modern Physics*, *79*, 135–174.

Kollins, S. H., Newland, M. C., & Critchfield, T. S. (1997). Human sensitivity to reinforcement in operant choice: How much do consequences matter? *Psychonomic Bulletin & Review*, *4*, 208–220.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, *22*, 79-86.

Kwak, C., & Clayton-Matthews, A. (2002). Multinomial logistic regression. *Nursing Research*, *51*, 404–410.

Lattimore, P. K., Baker, J. R., & Witte, A. D. (1992). The influence of probability on risky choice. *Journal of Economic Behavior and Organization*, *17*, 377–400.

Lau, B., & Glimcher, P. W. (2005). Dynamic response-by-response models of matching behavior in rhesus monkeys. *Journal of the Experimental Analysis of Behavior*, *84*, 555–579.

Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current Directions in Psychological Science*, *19*, 143–148.

Levy, H. (1992). Stochastic dominance and expected utility: Survey and analysis. *Management Science*, *38*, 555–593.

Loewenstein, Y., & Seung, H. S. (2009). Operant matching is a generic outcome of synaptic plasticity based on the covariance between reward and neural activity. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 15224–15229.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis.* Wiley.

Luce, R. D. (1969). Subjective expected utility theory. In *Coloóquio brasileiro de matemática, july 6-26* (pp. 5–15).

Luce, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology*, *15*, 215–233.

Luce, R. D. (1988). Rank-dependent, subjective expected-utility representations. *Journal of Risk and Uncertainty*, *1*, 305–332.

Luce, R. D. (2002). A psychophysical theory of intensity proportions, joint presentations, and matches. *Psychological Review*, *109*, 520–532.

Luce, R. D., & Fishburn, P. C. (1991). Rank- and sign-dependent linear utility models for finite first-order gambles. *Journal of Risk and Uncertainty*, *4*, 29–59.

Luce, R. D., & Krantz, D. H. (1971). Conditional expected utility. *Econometrica*, *39*, 253–271.

Luce, R. D., & Narens, L. (1985). Classification of concatenation measurement structures according to scale type. *Journal of Mathematical Psychology*, *29*, 1–72.

Ludbrook, J. (1998). Multiple comparison procedures updated. *Clinical and Experimental Pharmacology and Physiology*, *25*, 1032–1037.

MacDonall, J. S. (1988). Concurrent variable-ratio schedules – implications for the generalized matching law. *Journal of the Experimental Analysis of Behavior*, *50*, 55–64.

Marozeau, J., & Florentine, M. (2007). Loudness growth in individual listeners with hearing losses: A review. *Journal of the Acoustical Society of America*, *122*, EL81–EL87.

Martín-Fernández, J. A., Barceló-Vidal, C., & Pawlowsky-Glahn, V. (2003). Dealing with zeros and missing values in compositional data sets using nonparametric imputation. *Mathematical Geology*, *35*, 253–278.

Martín-Fernández, J. A., Palarea-Albaladejo, J., & Olea, R. A. (2011). Dealing with zeros. In V. Pawlowsky-Glahn & A. Buccianti (Eds.), *Compositional data analysis: Theory and applications* (p. 43-58). Wiley.

Martín-Fernández, J. A., & Thió-Henestrosa, S. (2006). Rounded zeros: Some practical aspects for compositional data. *Geological Society, London, Special Publications*, *264*, 191–201.

Masoro, E. J., Yu, B. P., & Bertrand, H. A. (1982). Action of food restriction in delaying the aging process. *Proceedings of the National Academy of Sciences of the United States of America*, *79*, 4239–4241.

McMillan, B. (1953). The basic theorems of information theory. *Annals of Mathematical Statistics*, *24*, 196–219.

Milkman, K. L., Chugh, D., & Bazerman, M. H. (2009). How can decision making be improved? *Perspectives on Psychological Science*, *4*, 379–383.

Montemurro, M. A., & Zanette, D. H. (2011). Universal entropy of word ordering across linguistic families. *PLOS ONE*, *6*, e19875.

Natapoff, A. (1970). How symmetry restricts symmetric choice. *Journal of Mathematical Psychology*, *7*, 444–465.

Navakatikyan, M. A., & Davison, M. (2010). The dynamics of the law of effect: A comparison of models. *Journal of the Experimental Analysis of Behavior*, *93*, 91–127.

Neuringer, A., & Jensen, G. (2010). Operant variability and voluntary action. *Psychological Review*, *117*, 972–993.

Neuringer, A., & Jensen, G. (2012). The predictably unpredictable operant. *Comparative cognition and behavior reviews*, *7*, 55–84.

Nevin, J. A. (2003). Mathematical principles of reinforcement and resistance to change. *Behavioural Processes*, *62*, 65–73.

Norvig, P. (2013, January). *English letter frequency counts: Mayzner revisited, or, ETAOIN SRHLDCU.* `http://norvig.com/mayzner.html`. (Accessed: 2014-01-15)

Olton, D. S., & Samuelson, R. J. (1976). Remeberance of places passed: Spatial memory in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, *2*, 97–116.

Ore, O. (1960). Pascal and the invention of probability theory. *American Mathematical Monthly*, *67*, 409–419.

Pawlowsky-Glahn, V., & Buccianti, A. (Eds.). (2011). *Compositional data analysis: Theory and applications.* Wiley.

Pearce, M. T., & Wiggins, G. A. (2004). Improved methods for statistical modeling of monophonic music. *Journal of New Music Research*, *33*, 367–385.

Pellegrino, F., Coupé, C., & Marsico, E. (2011). A cross-language perspective on speech information rate. *Language*, *87*, 539–558.

Pfanzagl, J. (1967). Subjective probability derived from the morgenstern-von neumann utility concept. In M. Shubik (Ed.), *Essays in mathematical economics in honor of oskar morgenstern* (pp. 237–251). Princeton University Press.

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, *117*, 864–901.

Polunchenko, A. S., & Tartakovsky, A. G. (2012). State-of-the-art in sequential change-point detection. *Methodology and Computing in Applied Probability*, *14*, 649–684.

Quade, D. (1984). Nonparametric methods in two-way layouts. In P. R. Krishnaiah & P. K. Sen (Eds.), *Handbook of statistics, vol. 4: Nonparametric methods* (pp. 185–228). Elsevier.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization*, *3*, 323–343.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural computation*, *20*, 873–922.

Reed, L. J., & Berkson, J. (1929). The application of the logistic function to experimental data. *Journal of Physical Chemistry*, *33*, 760–779.

Reiger, M. O., & Wang, M. (2008). Prospect theory for continuous distributions. *Journal of Risk and Uncertainty*, *36*, 83–102.

Rhyne, A. L., & Steel, R. G. D. (1965). Tables for a treatments versus control multiple comparisons sign test. *Technometrics*, *7*, 293–306.

Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, *58*, 527–535.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*, 207–231.

Roxin, A., & Ledberg, A. (2008). Neurobiological models of two-choice decision making can be reduced to a one-dimensional nonlinear diffusion equation. *PLoS Computational Biology*, *4*, e1000046.

Sakai, Y., & Fukai, T. (2008). The actor-critic learning is behind the matching law: Matching versus optimal behaviors. *Neural Computation*, *20*, 227–251.

Sanford, R. F., Pierson, C. T., & Crovelli, R. A. (1993). An objective replacement method for censored geochemical data. *Mathematical Geology*, *25*, 59–80.

Scheibehenne, B., Greifeneder, R., & Todd, P. M. (2010). Can there ever be too many options? a meta-analytic review of choice overload. *Journal of Consumer Research*, *37*, 409–425.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica*, *57*, 571–581.

Schneider, B. (1996). *Applied cryptography.* Wiley.

Schneider, S. M., & Davison, M. (2005). Demarcated response sequences and generalised matching. *Behavioural Processes*, *70*, 51–61.

Schwartz, B. (Ed.). (2004). *The paradox of choice: Why more is less.* Harper Perennial.

Shackle, G. L. S. (Ed.). (1949). *Expectation in economics.* Cambridge University Press.

Shackle, G. L. S. (1958). Decision in face of uncertainty: Some criticisms and extensions of a theory. *De Economist*, *106*, 673–686.

Shackle, G. L. S. (Ed.). (1970). *Expectation, enterprise, and profit: The theory of the firm.* George Allen & Unwin.

Silverman, B. W. (Ed.). (1986). *Density estimation for statistics and data analysis.* Chapman & Hall/CRC.

Simon, D. A., & Daw, N. D. (2011). Neural correlates of forward planning in a spatial decision task in humans. *Journal of Neuroscience*, *31*, 5526–5539.

Sol, D. (2009). The cognitive-buffer hypothesis for the evolution of large brains. In R. Dukas & J. M. Ratcliffe (Eds.), *Cognitive ecology ii* (pp. 111–134). University of Chicago Press.

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature*, *38*, 332–382.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*, 677–680.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, *64*, 153–181.

Steyvers, M., Lee, M. D., & Wagenmakers, E.-J. (2009). A bayesian analysis of human decision-making on bandit problems. *Journal of Mathematical Psychology*, *53*, 168–179.

Stott, H. P. (2006). Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty*, *32*, 101–130.

Stumpf, M. P. H., & Porter, M. A. (2012). Critical truths about power laws. *Science*, *335*, 665–666.

Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2004). Matching behavior and the representation of value in the parietal cortex. *Science*, *304*, 1782–1787.

Sun, J. Z., Wang, G. I., Goyal, V. K., & Varshney, L. R. (2012). A framework for bayesian optimality of psychophysical laws. *Journal of Mathematical Psychology*, *56*, 495–501.

Sutton, N. P., Grace, R. C., McLean, A. P., & Baum, W. M. (2008). Comparing the generalized matching law and contingency discriminability model as accounts of concurrent schedule performance using residual meta-analysis. *Behavioural Processes*, *78*, 224–230.

Sutton, R. S., & Barto, A. G. (Eds.). (1998). *Reinforcement learning: An introduction.* MIT Press.

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, *7*, 1–17.

Tatler, B. W., & Vincent, B. T. (2009). The prominence of behavioural biases in eye guidance. *Visual Cognition*, *17*, 1029–1054.

Taylor, R., & Davison, M. (1983). Sensitivity to reinforcement in concurrent arithmetic and exponential schedules. *Journal of the Experimental Analysis of Behavior*, *39*, 191–198.

Teodorescu, A. R., & Usher, M. (2013). Disentangling decision models: From independence to competition. *Psychological Review*, *120*, 1–38.

Treit, D., & Fundytus, M. (1988). Thigmotaxis as a test for anxiolytic activity in rats. *Pharmacology, Biochemistry, and Behavior*, *31*, 959–9620.

Trepel, C., Fox, C. R., & Poldrack, R. A. (2005). Prospect theory on the brain? toward a cognitive neuroscience of decision under risk. *Cognitive Brain Research*, *23*, 34–50.

Troldahl, V. C., & Jr., R. E. C. (1964). Random selection of respondents within households in phone surveys. *Journal of Marketing Research*, *1*, 71–76.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2008). Decision making, movement planning, and statistical decision theory. *Trends in Cognitive Sciences*, *12*, 291–297.

Tversky, A., & Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review*, *102*, 269–283.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *221*, 453–458.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representations of uncertainty. *Journal of Risk and Uncertainty*, *7*, 297–323.

Velleman, P. F., & Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. *The American statistician*, *47*, 65–72.

Vickery, C., & Neuringer, A. (2000). Pigeon reaction time, Hick's law, and intelligence. *Psychonomic Bulletin & Review*, *7*, 284–291.

Vilares, I., & Kording, K. (2011). Bayesian models: The structure of the world, uncertainty, behavior, and the brain. *Annals of the New York Academy of Sciences*, *1224*, 22–39.

Vlaev, I., Chater, N., Stewart, N., & Brown, G. D. A. (2011). Does the brain calculate value? *Trends in Cognitive Sciences*, *15*, 546–554.

von Neumann, J., & Morgenstern, O. (Eds.). (1947). *Theory of games and economic behavior* (2nd ed.). Princeton University Press.

Vulkan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys*, *14*, 101–118.

Ward, R. D., Gallistel, C. R., & Balsam, P. D. (2013). It's the information! *Behavioural Processes*, *95*, 3–7.

Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of Mathematical Psychology*, *44*, 92–107.

Weber, M., & Camerer, C. (1987). Recent developments in modelling preferences under risk. *OR Spektrum*, *9*, 129–151.

Wehrens, R., Putter, H., & Buydens, L. M. C. (2000). The bootstrap: A tutorial. *Chemometrics and Intelligent Laboratory Systems*, *54*, 35–52.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: the psychophysics of number representation. *Psychological Science*, *10*, 130–137.

Wolfe, G. (1992). Words weird and wonderful. In *Castle of days.* Orb Books.

Zentall, T. R., & Wasserman, E. A. (Eds.). (2012). *The oxford handbook of comparative cognition.* Oxford University Press.

Zhang, H., & Maloney, L. T. (2012). Ubiquitous log odds: A common representation of probability and frequency distortion in perception, action, and cognition. *Frontiers in Neuroscience*, *6*, 1–14.

Zhang, S., & Lee, M. D. (2010). Optimal experimental design for a class of bandit problems. *Journal of Mathematical Psychology*, *54*, 499–508.

Part V

Appendices

Appendix A

Glossary of Symbols and Equations

Although the compositional paradigm provides a powerful and consistent method for dealing with a sample space that is unfamiliar to many, the corresponding notation will also be unfamiliar to those readers. This appendix provides a glossary of the key symbols and operators needed to understand compositional data, as well as a digest of all numbered equations appearing in the text.

A.1 Symbol Glossary

Data

$\mathbf{x}$	Vector. If $x = \mathcal{C}(\mathbf{x})$ (Equation 1.1), then $\mathbf{x}$ is also a composition (Equation 3.1).
$\mathbf{x}^{@}$	An additive log-ratio (ALR) transformation of $\mathbf{x}$ (Equation 4.1).
$\mathbf{x}^{\circ}$	The centered log-ratio (CLR) transformation of $\mathbf{x}$ (Equation 4.3).
$\mathbf{x}^{*}$	An isometric log-ratio (ILR) transformation of $\mathbf{x}$ (Equation 4.5).

Compositional Operation

$\mathcal{C}(\cdot)$	The 'closure' operation, which converts a vector to a composition.
$\oplus$	The 'perturbation' operator (Equation 3.3), analogous to addition.
$\odot$	The 'powering' operator (Equation 3.3), analogous to multiplication.
$\ominus$	The 'anti-perturbation' operator (Equation 3.5), analogous to subtraction.
$\bigoplus$	The 'perturbation-linear combination' (Equation 3.10).

$\langle\cdot,\cdot\rangle_a$ The 'Aitchison inner product' (Equation 3.6), analogous to the inner product $\langle\cdot,\cdot\rangle$.

$\|\cdot\|_a$ The 'Aitchison norm' (Equation 3.7), analogous to the Euclidean norm $\|\cdot\|$.

$\mathfrak{d}_a(\cdot,\cdot)$ The 'Aitchison distance' (Equation 3.8), analogous to Euclidean distance $\mathfrak{d}(\cdot,\cdot)$.

Model Parameters

α Exponent used for powering, typically estimating 'sensitivity' (Equation 1.12).

$\boldsymbol{\kappa}$ Composition used for perturbation, typically estimating 'bias' (Equation 1.12).

$\mathcal{B}$ An estimated stationary Markov chain describing behavior over some interval.

$\mathcal{Q}$ A 'default transition matrix,' estimated across multiple behaviors $\mathcal{B}$.

Special Utility Matrices

$\mathbf{U}$ An orthonormal basis (Equation 4.6) for use in ILR transformation.

$\mathbf{B}$ A bifurcation matrix (Equation 4.10) used to construct a valid $\mathbf{U}$ matrix.

ϖ and ω Balancing terms (Equation 4.9) used in converting $\mathbf{B}$ to $\mathbf{U}$.

ϱ_j^d A vector of length d, in which element j equals 1 and all other elements equal zero.

Prospects & Lotteries

L_A The objective expected value of a lottery A (Equation 1.2).

$\mathcal{L}_A$ The subjective expected value of A, in the same units as L_A (Equation 1.5).

$\mathcal{L}_A^+$ and $\mathcal{L}_A^-$ $\mathcal{L}_A$ subdivided into gains (+) and losses(-) (Equation 13.1).

A.2 The Menagerie of Equations

Throughout this dissertation, the reader is referred to a great many equations, many of which are only fully clear in the context of the surrounding text. The following list redundantly lists each of the numbered equations in the text, providing a brief description and the page number of its original appearance. Rather than including a list of equations in the Table of Contents, this section is intended to provide the reader with a condensed reference. A reader who wishes to re-examine an equation in its original context may do so, but retracing that garden path is not required for those readers who merely need to refresh their memories or for whom that context is already familiar.

• **Equation 1.1**, P. 5 – The closure operation $\mathcal{C}\left(\right)$, which can be applied to any vector $\mathbf{B}$ of non-negative values(Aitchison, 1986):

$$\mathcal{C}\left(\mathbf{B}\right) = \left[\frac{B_1}{\sum_{j=1}^{n} B_j}, \cdots, \frac{B_n}{\sum_{j=1}^{n} B_j}\right] \tag{1.1}$$

• **Equation 1.2**, P. 8 – The expected value of a discrete outcome of a lottery A, as a function of the probability $p_{i,A}$ and reward magnitude $M_{i,A}$ of outcome i (Bernoulli, 1738|1954):

$$L_A = \sum_{i \in D} p_{i,A} \cdot M_{i,A} \tag{1.2}$$

• **Equation 1.3**, P. 9 – The expected utility of a discrete outcome A, modifying Equation 1.2 with a coefficient β of subjective reward value(Bernoulli, 1738|1954):

$$EU\left(L_A|\beta\right) = \sum_{i \in D} p_{i,A} \cdot M_{i,A}^{\beta} \tag{1.3}$$

• **Equation 1.4**, P. 10 – The subjected weighted expected utility of a discrete outcome A, modifying Equation 1.2 with a coefficient α of subjective probability and β of subjective reward value (Karmarkar, 1978):

$$SWEU\left(L_A|\alpha,\beta\right) = \sum_{i \in D} \frac{p_{i,A}^{\alpha}}{\sum_{j \in D} p_{j,A}^{\alpha}} \cdot M_{i,A}^{\beta} \tag{1.4}$$

• **Equation 1.5**, P. 10 – The subjected expected value, obtained by translating Equation 1.4 into the units originally employed by Equation 1.2.

$$\mathcal{L}_A = \left(\sum_{i \in D} \frac{p_{i,A}^{\alpha}}{\sum_{j \in D} p_{j,A}^{\alpha}} \cdot M_{i,A}^{\beta} \right)^{\frac{1}{\beta}} \tag{1.5}$$

• **Equation 1.6**, P. 13 – Herrnstein's matching law, comparing proportion of a response behavior B_i on alternative i to obtained reinforcers R_i on that alternative (Herrnstein, 1961):

$$\frac{B_1}{B_1 + B_2 + \cdots + B_n} = \frac{R_1}{R_1 + R_2 + \cdots + R_n} \tag{1.6}$$

• **Equation 1.7**, P. 13 – Herrnstein's hyperbolic model of asymptotic response rate, modifying Equation 1.6 to represent one response, while adding an error term r_e, a bias term k_i, and a rate parameter t (de Villiers & Herrnstein, 1976):

$$\frac{B_i}{t} = \frac{k_i}{t} \cdot \frac{R_i}{R_i + r_e} \tag{1.7}$$

• **Equation 1.8**, P. 13 – The Generalized Matching Law, relating the ratio response behavior B_i to obtained reinforcers R_i as a function of a ratio of bias terms k_i and a sensitivity parameter s (Baum, 1974):

$$\frac{B_1}{B_2} = \frac{k_1}{k_2} \cdot \left(\frac{R_1}{R_2} \right)^s \tag{1.8}$$

• **Equation 1.9**, P. 14 – The log-transformation of Equation 1.8, converting it to a linear equation:

$$\log \left(\frac{B_1}{B_2} \right) = \log \left(\frac{k_1}{k_2} \right) + s \cdot \log \left(\frac{R_1}{R_2} \right) \tag{1.9}$$

• **Equation 1.10**, P. 15 – A compositionally inconsistent attempt to extend Equation 1.8 to more than two alternatives using standard addition (Goldstein & Einhorn, 1987):

$$\frac{B_i}{\sum_{j=1}^{n} B_j} = \frac{k \cdot (R_i)^s}{k \cdot (R_i)^s + \left[\sum_{j \neq i} (R_j)^s\right]} \tag{1.10}$$

• **Equation 1.11**, P. 16 – A version of Equation 1.8 using the centered log-ratio transformation (Equation 4.3) (S. M. Schneider & Davison, 2005):

$$\mathbf{B}^\circ = \mathbf{k}^\circ + s \cdot \mathbf{R}^\circ \tag{1.11}$$

• **Equation 1.12**, P. 16 – The Barycentric Matching Model, another form of Equation 1.11 that emphasizes the distinct values of k_i (Jensen & Neuringer, 2009):

$$\frac{k_1 \cdot (R_1)^s}{B_1} = \cdots = \frac{k_n \cdot (R_n)^s}{B_n} \tag{1.12}$$

• **Equation 1.13**, P. 22 – Cumulative Prospect Theory, which predicts that gains (+) and losses (-) are processed using independent parameters and subsequently summed (Tversky & Kahneman, 1992):

$$SWEU(L_A) = \left(\sum_{i \in D} \mathcal{W}^+(\mathbf{p}_{A^+}) \cdot \mathcal{V}(\mathbf{M}_{A^+})\right) + \left(\sum_{i \in D} \mathcal{W}^-(\mathbf{p}_{A^-}) \cdot \mathcal{V}(\mathbf{M}_{A^-})\right) \tag{1.13}$$

• **Equation 3.1**, P. 31 – Formal definition of the conditions under which a vector qualifies as a composition; here, the closure constant ð is effectively arbitrary, and is taken to equal 1.0 unless otherwise specified (Aitchison, 1986):

$$\mathbf{x} = (x_1, \ldots, x_D) \Big| x_i > 0, \sum_{i=1}^{D} x_i = ð > 0 \tag{3.1}$$

• **Equation 3.2**, P. 31 – The principle of subcomposition, whereby any two vector **x** and **y** may be combined into a composition **z**, where $\eth_x$ and $\eth_y$ correspond to the overall proportion of **z** made up by each subcomposition (Aitchison, 1986):

$$\mathbf{z} = \mathcal{C}\left(\eth_x \cdot \mathcal{C}\left(\mathbf{x}\right), \eth_y \cdot \mathcal{C}\left(\mathbf{y}\right)\right) \tag{3.2}$$

• **Equation 3.3**, P. 35 – The perturbation operation, which is the analog of addition in the compositional sample space (Aitchison, 1986):

$$\mathbf{x} \oplus \mathbf{y} = \mathcal{C}\left(x_1 \cdot y_1, x_2 \cdot y_2, \ldots, x_D \cdot y_D\right) \tag{3.3}$$

• **Equation 3.4**, P. 35 – The powering operation, which is the analog of multiplication in the compositional sample space (Aitchison, 1986):

$$\alpha \odot \mathbf{x} = \mathcal{C}\left(x_1^{\alpha}, x_2^{\alpha}, \ldots, x_D^{\alpha}\right) \tag{3.4}$$

• **Equation 3.5**, P. 35 – The anti-perturbation operation, which is the analog of subtraction in the compositional sample space (Aitchison, 1986):

$$\mathbf{x} \ominus \mathbf{y} = \mathbf{x} \oplus \left((-1) \odot \mathbf{y}\right) \tag{3.5}$$

• **Equation 3.6**, P. 35 – The Aitchison inner product, which the analog of the inner product in the compositional sample space (Aitchison, 1986):

$$\langle \mathbf{x}, \mathbf{y} \rangle_a = \sum_{i=1}^{D} \log \frac{x_i}{\mathfrak{g}\left(\mathbf{x}\right)} \cdot \log \frac{y_i}{\mathfrak{g}\left(\mathbf{y}\right)} \tag{3.6}$$

• **Equation 3.7**, P. 36 – The Aitchison norm, which the analog of the norm in the compositional sample space (Aitchison, 1986):

$$\|\mathbf{x}\|_a = \langle \mathbf{x}, \mathbf{x} \rangle_a \tag{3.7}$$

• **Equation 3.8**, P. 36 – The Aitchison distance, which the analog of Euclidean distance in the compositional sample space (Aitchison, 1986):

$$\mathfrak{d}_a (\mathbf{x}, \mathbf{y}) = \|\mathbf{x} \ominus \mathbf{y}\|_a \tag{3.8}$$

• **Equation 3.9**, P. 36 – Formula for a 'simplicial line' in the compositional sample space (Aitchison, 1986):

$$\mathbf{y} = \boldsymbol{\kappa} \oplus (\alpha \odot \mathbf{x}) \tag{3.9}$$

• **Equation 3.10**, P. 37 – The perturbation-linear combination, analogous to linear combination as applied to compositions (Aitchison, 1986):

$$\mathbf{y} = \bigoplus_{j=1}^{D} (\alpha_j \odot \mathbf{x}_j) = (\alpha_1 \odot \mathbf{x}_1) \oplus \cdots \oplus (\alpha_D \odot \mathbf{x}_D) \tag{3.10}$$

• **Equation 4.1**, P. 39 – The additive log-ratio (ALR) transformation, denoted by a superscript @, which can be applied to any composition **x** (Aitchison, 1986):

$$\text{alr}(\mathbf{x}) = \mathbf{x}^{@} = \log \left(\frac{x_1}{x_D}, \frac{x_2}{x_D}, \ldots, \frac{x_d}{x_D} \right) \tag{4.1}$$

• **Equation 4.2**, P. 40 – Generalized Matching (Equation 1.8) recast in terms of the ALR transformation (Equation 4.1):

$$\mathbf{y}^{@} = \boldsymbol{\kappa}^{@} + \alpha \cdot \mathbf{x}^{@} \tag{4.2}$$

• **Equation 4.3**, P. 41 – The centered log-ratio (CLR) transformation, denoted by a superscript @, which can be applied to any composition $\mathbf{x}$ (Aitchison, 1986):

$$\operatorname{clr}(\mathbf{x}) = \mathbf{x}^{\circ} = \log\left(\frac{x_1}{\mathfrak{g}(\mathbf{x})}, \frac{x_2}{\mathfrak{g}(\mathbf{x})}, \ldots, \frac{x_D}{\mathfrak{g}(\mathbf{x})}\right) \tag{4.3}$$

• **Equation 4.4**, P. 41 – The anti-CLR transformation, which converts CLR data back into a compositional form:

$$\mathbf{x} = \mathcal{C}\left(\exp\left(\mathbf{x}^{\circ}\right)\right) \tag{4.4}$$

• **Equation 4.5**, P. 43 – The isometric log-ratio (ILR) transformation, denoted by a superscript $*$, which can be applied to any composition $\mathbf{x}$ through the joint use of the CLR transformation (Equation 4.3) and an appropriate orthonormal basis $\mathbf{U}$ (Equation 4.6) (Egozcue et al., 2003):

$$\mathbf{x}^{*} = [\mathbf{U}(\mathbf{x}^{\circ})^{\intercal}]^{\intercal} \qquad , \qquad \mathbf{x}^{\circ} = (\mathbf{x}^{*})\,\mathbf{U} \tag{4.5}$$

• **Equation 4.6**, P. 43 – An orthonormal basis of size (d by D) (Egozcue & Pawlowsky-Glahn, 2005):

$$\mathbf{U} = \begin{bmatrix} \mathbf{u_1} \\ \vdots \\ \mathbf{u_d} \end{bmatrix} = \begin{bmatrix} u_{1,1} & \cdots & u_{1,d} & u_{1,D} \\ \vdots & \ddots & \vdots & \vdots \\ u_{1,1} & \cdots & u_{1,d} & u_{1,D} \end{bmatrix} \quad \begin{matrix} ||\mathbf{u_1}|| = \cdots = ||\mathbf{u_d}|| = 1 \\ \langle \mathbf{u_i}, \mathbf{u_j} \rangle = 0, i \neq j \end{matrix} \tag{4.6}$$

• **Equation 4.7**, P. 44 – A sample orthonormal basis of size (2 by 3) (Egozcue & Pawlowsky-Glahn, 2005):

$$\mathbf{U} = \begin{bmatrix} \mathbf{u_1} \\ \mathbf{u_2} \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} & 0 \\ \sqrt{\frac{1}{2\cdot 3}} & \sqrt{\frac{1}{2\cdot 3}} & -\sqrt{\frac{2}{3}} \end{bmatrix} \tag{4.7}$$

• **Equation 4.8**, P. 45 – A 'default' orthonormal basis provided by the Gram-Schmidt process:

$$\mathbf{U} = \begin{bmatrix} \mathbf{u_1} \\ \mathbf{u_2} \\ \mathbf{u_3} \\ \vdots \\ \mathbf{u_d} \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} & 0 & 0 & \cdots & 0 & 0 \\ \sqrt{\frac{1}{2\cdot 3}} & \sqrt{\frac{1}{2\cdot 3}} & -\sqrt{\frac{2}{3}} & 0 & \cdots & 0 & 0 \\ \sqrt{\frac{1}{3\cdot 4}} & \sqrt{\frac{1}{3\cdot 4}} & \sqrt{\frac{1}{3\cdot 4}} & -\sqrt{\frac{3}{4}} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \sqrt{\frac{1}{d\cdot D}} & \sqrt{\frac{1}{d\cdot D}} & \sqrt{\frac{1}{d\cdot D}} & \sqrt{\frac{1}{d\cdot D}} & \cdots & \sqrt{\frac{1}{d\cdot D}} & -\sqrt{\frac{d}{D}} \end{bmatrix} \tag{4.8}$$

• **Equations 4.9 and 4.10**, P. 47 – Formula for converting a bifucation matrix **B** (of size d by D) into an orthonormal basis (Equation 4.6) (Egozcue & Pawlowsky-Glahn, 2006):

$$\begin{aligned} \varpi_i &= \sum \{j \in \mathbf{b_i} | j = 1\}, \varpi_i > 0 \\ \omega_i &= -\sum \{j \in \mathbf{b_i} | j = -1\}, \omega_i > 0 \\ \mathbf{u_{i}}_{,j} &= \begin{cases} \sqrt{\frac{\omega_i}{\varpi_i(\varpi_i+\omega_i)}} & \text{if } \mathbf{b_{i}}_{,j} = 1 \\ -\sqrt{\frac{\varpi_i}{\omega_i(\varpi_i+\omega_i)}} & \text{if } \mathbf{b_{i}}_{,j} = -1 \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{4.9 and 4.10}$$

• **Equation 4.11**, P. 47 – Alternative ILR transformation using the eigenvectors **W** obtained from a principal component analysis.

$$\mathbf{X}^* = [\mathbf{W}\,(\mathbf{X}^\circ)^\intercal]^\intercal, \mathbf{W} = \begin{bmatrix} \mathbf{w_1} \\ \vdots \\ \mathbf{w_d} \end{bmatrix} \tag{4.11}$$

• **Equation 5.1**, P. 49 – A compositional formulation of the Barycentric Matching Model (Equation 1.12):

$$\mathbf{y}^* = \boldsymbol{\kappa}^* + \alpha \cdot \mathbf{x}^* \tag{5.1}$$

• **Equation 5.2**, P. 50 – Matrix form for Equation 5.1 to facilitate construction of dummy variables. Here, ϱ_j^d denotes a vector of length d in which element j is a 1 and all others are zero:

$$\left[\mathbf{y}^*_{\mathbf{i},j}\right] = \left[\varrho_j^d, \mathbf{x}^*_{\mathbf{i},j}\right] \left[\boldsymbol{\kappa}^*, \alpha\right]^\intercal \tag{5.2}$$

• **Equation 5.3**, P. 51 – A compositional extension of the Barycentric Matching Model (Equation 1.12) to incorporate both probabilities of reward $\mathbf{x}_R$ and magnitude of reward $\mathbf{x}_M$:

$$\mathbf{y}^* = \boldsymbol{\kappa}^* + \alpha_R \cdot \mathbf{x}^*_R + \alpha_M \cdot \mathbf{x}^*_M \tag{5.3}$$

• **Equation 5.4**, P. 56 – A compositional extension of the Barycentric Matching Model (Equation 1.12) in which different points of indifference $\boldsymbol{\kappa}^*_i$ arise for c different subsets of data:

$$\left[\mathbf{y}^*_{\mathbf{i},j}\right] = \left[\underbrace{\varrho_j^d, \ldots, \varrho_j^d}_{c \text{ vectors}}, \mathbf{x}^*_{\mathbf{i},j}\right] \left[\boldsymbol{\kappa}^*_1, \ldots, \boldsymbol{\kappa}^*_c, \alpha\right]^\intercal \tag{5.4}$$

• **Equation 5.5**, P. 56 – A compositional extension of the Barycentric Matching Model (Equation 1.12) in which different sensitivities
$alpha_i$ arise for c different subsets of data, encoded in a vector $\boldsymbol{\alpha}$ of length c:

$$\left[\mathbf{y}^*_{\mathbf{i},j}\right] = \left[\varrho_j^d, \left(\varrho_i^c \cdot \mathbf{x}^*_{\mathbf{i},j}\right)\right] \left[\boldsymbol{\kappa}^*, \boldsymbol{\alpha}\right]^\intercal \tag{5.5}$$

• **Equation 5.6**, P. 58 – A compositional extension of the Barycentric Matching Model (Equation 1.12) in which each orthonormal contrast is assigned a different sensitivity
$alpha_i$, such that the vector $\boldsymbol{\alpha}$ is of length d:

$$\left[\mathbf{y}^*_{\mathbf{i},j}\right] = \left[\varrho^d_j, \left(\varrho^d_j \cdot \mathbf{x}^*_{\mathbf{i},j}\right)\right] \left[\boldsymbol{\kappa}^*, \boldsymbol{\alpha}\right]^\intercal \tag{5.6}$$

• **Equation 5.7**, P. 59 – A compositional extension of the Barycentric Matching Model (Equation 1.12) to systematic multivariate regression:

$$\mathbf{Y}^* = \mathbf{h}^\intercal \left[\boldsymbol{\kappa}^*\right] + \mathbf{X}^* \left[\boldsymbol{\aleph}\right]^\intercal, \quad \boldsymbol{\aleph} = \begin{bmatrix} \boldsymbol{\alpha}_\mathbf{1} \\ \vdots \\ \boldsymbol{\alpha}_\mathbf{d} \end{bmatrix}$$
$$\mathbf{h} = \underbrace{(1, \ldots, 1)}_{n \text{ ones}}, \quad \boldsymbol{\alpha}_\mathbf{i} = (\alpha_1, \ldots, \alpha_d) \tag{5.7}$$

• **Equation 6.1**, P. 61 – A rule-of-thumb for replacing rounded zeros in order to avoid marginal compositions (Martín-Fernández et al., 2011). Note that $\delta_i = 0.55$ is considered suitable in cases where the raw data are discrete counts, as per Sanford et al. (1993):

$$\mathbf{r} = \mathcal{C}\left(\mathbf{B}^r\right), \mathbf{B}^r_i = \begin{cases} \delta_i & \text{if } B_i = 0 \\ B_i & \text{if } B_i > 0 \end{cases} \tag{6.1}$$

• **Equation 6.2**, P. 63 – Explanatory factor for centered multinomial logistic regression:

$$\varkappa_{n,i} = \kappa_n + (\alpha_{1,n} \cdot x_{1,i}) + \ldots + (\alpha_{M,n} \cdot x_{M,i}) \tag{6.2}$$

• **Equation 6.3**, P. 63 – The Softmax function (R. S. Sutton & Barto, 1998):

$$\mathbf{y}_{\mathbf{i},j} = \frac{\exp\left(\varkappa_{j,i}\right)}{\sum_{n=1}^{N} \exp\left(\varkappa_{n,i}\right)} \Big| \sum_{n=1}^{N} \mathbf{y}_{\mathbf{i},n} = 1 \tag{6.3}$$

• **Equation 6.4**, P. 64 – Simplification of the Softmax using the ILR transformation (Equation 4.5) to eliminate the computationally problematic denominator:

$$\mathbf{y}^{*}_{\mathbf{i},j} = \sum_{n=1}^{N} \left(\mathbf{u}_{\mathbf{i},n} \cdot \varkappa_{n,i}\right) \tag{6.4}$$

• **Equation 8.1**, P. 71 – Matching model predicting the proportions of responses **B** in terms of the frequency of reinforcers **R** and the magnitude of reinforcers **M**, extending Baum's matching model (Equation 1.8) and representing it in terms of the ALR-transformation (Equation 4.1) (Elliffe et al., 2008):

$$\mathbf{B}^{@} = \boldsymbol{\kappa}^{@} + \alpha\mathbf{R}^{@} + \beta\mathbf{M}^{@} \tag{8.1}$$

• **Equation 8.2**, P. 74 – The contingency discriminability model for two choice alternatives. Here, λ acts as a bias factor and ρ corresponds to a factor of the confusability of the two scheduled of reinforcement. ω is a correction term that enhances the subjective contrast between the schedules (Davison & Jenkins, 1985):

$$\frac{B_1}{B_2} = \lambda\frac{R_1 - \rho R_1 + \rho R_2 - \omega}{R_2 - \rho R_2 + \rho R_1 - \omega} \tag{8.2}$$

• **Equation 8.3**, P. 76 – Two orthonormal bases (Equation 4.6) used for ILR transformation (Equation 4.5) in the re-analysis of data collected by Davison & McCarthy:

$$\mathbf{U}_{xz|y} = \begin{bmatrix} \sqrt{\frac{1}{2}} & 0 & -\sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{6}} & -\sqrt{\frac{2}{3}} & \sqrt{\frac{1}{6}} \end{bmatrix} \quad , \quad \mathbf{U}_{yz|x} = \begin{bmatrix} 0 & \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} \\ \sqrt{\frac{2}{3}} & -\sqrt{\frac{1}{6}} & -\sqrt{\frac{1}{6}} \end{bmatrix} \tag{8.3}$$

• **Equation 9.1**, P. 80 – Probability of reward delivery in the Turn-Based Foraging paradigm (Algorithm 1, p. 81), predicted according to a geometric distribution as a function of trials since last visit t (Jensen & Neuringer, 2008):

$$\Pr\left(\text{Reward}|C_i\right) = 1 - (1 - P_i)^{t_i} \tag{9.1}$$

• **Equation 9.2**, P. 84 – Orthonormal basis (Equation 4.6) used for ILR transformation (Equation 4.5) in the analysis of Experiment 1:

$$\mathbf{U}_8 = \begin{bmatrix} \mathbf{u_1} \\ \mathbf{u_2} \\ \mathbf{u_3} \\ \mathbf{u_4} \\ \mathbf{u_5} \\ \mathbf{u_6} \\ \mathbf{u_7} \end{bmatrix} = \begin{bmatrix} 0 & 0 & \sqrt{\frac{2}{8}} & \sqrt{\frac{2}{8}} & -\sqrt{\frac{2}{8}} & -\sqrt{\frac{2}{8}} & 0 & 0 \\ 0 & 0 & \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} & 0 & 0 \\ 0 & 0 & \sqrt{\frac{2}{24}} & \sqrt{\frac{2}{24}} & \sqrt{\frac{2}{24}} & \sqrt{\frac{2}{24}} & -\sqrt{\frac{4}{12}} & -\sqrt{\frac{4}{12}} \\ 0 & 0 & 0 & 0 & 0 & 0 & \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} \\ \sqrt{\frac{6}{16}} & \sqrt{\frac{6}{16}} & -\sqrt{\frac{2}{48}} & -\sqrt{\frac{2}{48}} & -\sqrt{\frac{2}{48}} & -\sqrt{\frac{2}{48}} & -\sqrt{\frac{2}{48}} & -\sqrt{\frac{2}{48}} \\ \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{9.2}$$

• **Equation 9.3**, P. 87 – Silverman's 'rule of thumb' for bandwidth selection in Gaussian kernel density estimation (Silverman, 1986):

$$\sigma_{\text{KDE}} = \left(\frac{4}{3n}\right)^{\frac{1}{5}} \hat{\sigma} \tag{9.3}$$

• **Equation 9.4**, P. 94 – Formula for the stationary distribution of a positive recurrent Markov chain:

$$\pi_j = \lim_{n\to\infty} p_{i,j}^{(n)} \quad \text{where } p_{i,j}^{(n)} = \sum_{r\in D} p_{i,r}^{(k)} \cdot p_{r,j}^{(n-k)} \quad , \quad \boldsymbol{\pi} = \boldsymbol{\pi}\mathcal{B} \tag{9.4}$$

- **Equation 9.5**, P. 98 – Entropy rate in one step of a stationary Markov chain (McMillan, 1953):

$$H(\mathcal{B}) = -\sum_{i,j \in D} \pi_i \cdot \Pr(i|j) \cdot \log_2 (\Pr(i|j)) \tag{9.5}$$

- **Equation 9.6**, P. 101 – Kullback-Leibler divergence between an observed distribution P and a theoretical distribution Q (Kullback & Leibler, 1951):

$$KLD(P\|Q) = \sum_i P(i) \cdot \log_2 \left(\frac{P(i)}{Q(i)} \right) \tag{9.6}$$

- **Equation 9.7**, P. 102 – The divergence rate per step in a stationary Markov chain, achieved as a hybrid of Equation 9.5 and Equation 9.6:

$$KLD_T(\mathcal{B}) = \sum_{i,j \in D} \pi_i \cdot \mathcal{B}_{(i|j)} \cdot \log_2 \left(\frac{\mathcal{B}_{(i|j)}}{\mathcal{Q}_{(i|j)}} \right) \quad \text{where } \mathcal{Q}_{\mathbf{i}} = \mathcal{C}\left(\prod_{t=1}^{T} \left(\mathcal{B}_{\mathbf{i}|t}\right)^{-T} \right) \tag{9.7}$$

- **Equation 9.8**, P. 104 – Metric for the degree of mismatch between a default transition matrix $\mathcal{Q}$ (a stationary Markov chain) and the relative schedule richness among the alternatives $\mathbf{R}$:

$$\mathcal{I}_{sland} = D^2 \cdot \sum_{i,j}^{D} \mathcal{Q}_{i,j} \cdot \pi_i \cdot R_i \cdot R_j \quad \text{given } \mathcal{C}(\mathbf{R}) \tag{9.8}$$

• **Equation 13.1**, P. 176 – Compositional Prospect Theory, rendered in terms of subjective expected value, given a compositionally consistent weighting function and an integrated ratio scale for the value function:

$$
\begin{aligned}
\mathcal{L}_A^+ &= \left(\sum_{i \in D^+} \frac{\left(p_{i,A}^+\right)^{\alpha^+}}{\sum_{j \in D^+} \left(p_{j,A}^+\right)^{\alpha^+}} \cdot \int_0^\infty Log\mathcal{N}\left(x | \mu = \log\left(M_{i,A}\right), \sigma = c_v\right) \cdot x^{\beta^+} dx \right)^{\frac{1}{\beta^+}} \\
\mathcal{L}_A^- &= -\left(\sum_{i \in D^-} \frac{\left(p_{i,A}^-\right)^{\alpha^-}}{\sum_{j \in D^-} \left(p_{j,A}^-\right)^{\alpha^-}} \cdot \int_0^\infty Log\mathcal{N}\left(x | \mu = \log\left(-M_{i,A}\right), \sigma = c_v\right) \cdot x^{\beta^-} dx \right)^{\frac{1}{\beta^-}} \\
\mathcal{L}_A &= \frac{\mathcal{L}_A^+ \cdot \left(\sum_{i \in D^+} p_i^+\right)^\alpha + \mathcal{L}_A^- \cdot \left(\sum_{i \in D^-} p_i^-\right)^\alpha}{\left(\sum_{i \in D^+} p_i^+\right)^\alpha + \left(\sum_{i \in D^-} p_i^-\right)^\alpha + \left(p^\emptyset\right)^\alpha}
\end{aligned}
\tag{13.1}
$$

Appendix B

Change-Point Analysis: The CPR Algorithm

One of the consistent challenges in working with times series is that summary statistics (such as mean response rate) are calculated across blocks of time during which those summary statistics are presumed to be consistent. For example, most studies examining choice under the influence of concurrent schedules omit the trials (or sessions) that immediately follow a change in the schedule, in order to perform an analysis on 'steady-state' behavior. Typically, the steadiness of behavior is assumed (e.g. when a study uses 'the last five sessions in each phase') without being examined empirically.

There are, however, many quantitatively rigorous ways to determine when and how behaviors change over time. One such class of methods is 'change-point analysis,' which divides a time series into segments that are determined to be internally consistent, but also to differ measurably from one another (Polunchenko & Tartakovsky, 2012).

The change-point analyses in this dissertation are performed using the 'Conjugate Partitioned Recursion' (or CPR) algorithm (Jensen, 2014 (projected)). This method uses basic concepts in Bayesian statistics to identify an unknown number of abrupt discontinuities in a time series.

B.1 Bayes' Theorem

Bayes' Theorem describes one of the fundamental operations in probability theory: How an existing model M (whose parameters θ are described by a probability distribution) is updated on the basis of new observations x. Typically, the theorem takes the following form:

$$\Pr(\theta|x,M) = \frac{f(x|\theta,M)\Pr(\theta,M)}{m(x,M)}$$

where

$$m(x,M) = \int_\theta f(x|\theta,M)\Pr(\theta,M)\,d\theta \tag{B.1}$$

This equation color-codes the four essential parts of the equation. A *prior probability distribution* $\Pr(\theta,M)$ (in green) reflects the range of plausible values for the model parameters θ in the model M. This distribution is the convolved with a *likelihood function* $f(x|\theta,M)$ (in yellow), which reflects the odds associated with the observations x given the prior parameters. This convolution changes the shape of the distribution, reflecting the new information provided by x. Because the convolved distribution rarely integrates to 1.0, it must be divided by a *normalizing constant* $m(x,M)$ (in blue). As indicated, the normalizing constant is simply the integral of all possible values for θ. When these three elements are combined in this manner, the result is a *posterior probability distribution* $\Pr(\theta|x,M)$ (in red), which is the new distribution of possible values for θ once x has been taken into account.

B.2 Marginal Model Likelihood and Conjugate Priors

For many Bayesian methods, the normalizing constant $m(x,M)$ is treated as a nuisance variable and is omitted entirely. For example, because $m(x,M)$ does not depend on θ at all, it is not needed to determine the value for θ that is most likely given the likelihood function (the 'maximum likelihood' estimator) or given the posterior distribution (the 'maximum a posteriori' estimator).

However, $m(x,M)$ is an effective metric for parsimonious model selection. Given two models, M_1 and M_2, the model that provides a better overall fit to the data will have a larger normalizing constant. However, if the model M_1 has more parameters than the model M_2, its normalizing constant will be smaller (all else being equal). For this reason, $m(x,M)$ is also called the *marginal* *model likelihood* (or MML). Because it captures the principle of parsimony in a quantitatively rigorous way, it is reasonable to favor a model whose marginal model likelihood is higher than its competitors (Wasserman, 2000)

Unfortunately, $m(x, M)$ is often difficult to calculate. Even in cases where the prior distribution and the likelihood function are known, the MML often has no closed-form solution. While sophisticated numerical methods have arisen that yield precise approximations, these can be prohibitively intensive to compute.

If an analyst is willing to constrain the posterior distribution to a particular family of distributions, it can be shown that both the prior and posterior distributions are described by distributions whose sufficient statistics are closely linked. In these cases, the prior distribution is known as a *conjugate prior*. Conjugate priors can (under the right circumstances) make computing the values of each element in Equation B.1 a matter of simple arithmetic.

In the cases described in this dissertation, proportions of response are described by multinomial distributions. For example, given n observations made across 4 alternatives (such that the data $\mathbf{x}$ consist of the set (x_1, x_2, x_3, x_4)), the density associated with the parameters $\mathbf{p}$ is described as:

$$f(\mathbf{x}, \mathbf{p}) = \frac{n!}{x_1! x_2! x_3! x_4!} p_1^{x_1} p_2^{x_2} p_3^{x_3} p_4^{x_4} \quad \text{when } n = \sum_{i=1}^{4} x_i \tag{B.2}$$

The conjugate prior of the multinomial distribution is the Dirichlet distribution, whose normalizing constant is the multinomial beta function. It therefore follows that:

$$m(x, \text{multinomial}_D) = \frac{\prod_{i=1}^{D} \Gamma(x_i)}{\Gamma\left(\sum_{i=1}^{D} x_i\right)} \tag{B.3}$$

Like all marginal model likelihoods, Equation B.3 does no depend on the particular values of the model parameters θ. Instead, it depends on *hyperparameters*, which are the sufficient statistics necessary to specify the prior and posterior sampling distributions. In the case of a multinomial model, the counts of observations x_i are the appropriate hyperparameters. Since the values for x_i can be determined trivially from examining the data, computing the value for $m(x, \text{multinomial}_D)$ is equally straightforward.

B.3 Change-Point Detection As Model Selection

In change-point detection problems, the objective is to identify plausible discontinuities in a time series without either missing too many real discontinuities or making too many false positives. Parsimony is an essential criterion in this process, because every change-point effectively increases the complexity of the model. This is captured automatically by marginal model likelihoods: All else being equal, the marginal model likelihood of one uninterrupted time-series $(m(x, M))$ will be higher than the product of two marginal model likelihoods, each calculated for their own segment of the data $\left(m\left(x_{(1:c)}, M\right) \cdot m\left(x_{(c+1:n)}, M\right)\right)$.

The CPR algorithm uses marginal model likelihoods as a means of evaluating whether the current evidence supports adding change-points. This is done recursively, using a divide-and-conquer method. Given some segment of data, if the average MML for all possible positions of a change-point is greater than that of a no-change model, a change-point is added at the point which displays the maximum likelihood of being a change. This partitioning process is then repeated recursively on each of the resulting sub-segments, until no new change-points are identified.

Consider the following hypothetical data:

Lever 1:	0 1 2 0 2 3 2 2 6 4 5 5 4 8 1 2 1 1 1 0
Lever 2:	2 2 0 2 2 6 4 4 3 5 3 4 5 2 5 2 1 2 4 3
Lever 3:	2 4 4 3 3 1 3 4 0 0 0 0 1 0 1 2 5 3 4 4
Lever 4:	6 3 4 5 3 0 1 0 1 1 2 1 0 0 3 4 3 4 1 3

Here, each row corresponds to a number of observations, while each column corresponds to a block of ten responses[1]. The prior hyperparameters $\mathbf{x}$ are set to $(0.5, 0.5, 0.5, 0.5)$, as these have been known to be relatively uninformative. Then, to calculate the MML for the no-change model M_0, the posterior hyperparameters $\mathbf{x}'$ are obtained by adding the sums of all observations for each alternative. This yields $\mathbf{x}' = (50.5, 61.5, 44.5, 45.5)$. Thus, the MML for the no-change model is as follows:

$$m(x, M_0) = \frac{\Gamma(50.5) \cdot \Gamma(61.5) \cdot \Gamma(44.5) \cdot \Gamma(45.5)}{\Gamma(202)} = \exp(-280.69)$$

[1]The compression of the data into block of ten is done here merely for the sake of brevity; under normal circumstances, every trial would receive its own column, allowing for a more granular analysis.

The MML associated with placing a change-point at a given interval in this time-series can be computed by calculating the summary statistics before and after that interval and multiplying the resulting MMLs. For example, for a change-point between the 5th and 6th blocks of data, $\mathbf{x_{1:5}}' = (5.5, 8.5, 16.5, 21.5)$ and $\mathbf{x_{6:20}}' = (45.5, 53.5, 28.5, 24.5)$. Thus:

$$\begin{aligned} m\left(x, M_{1:5:20}\right) &= m\left(x_{1:5}, M_0\right) \cdot m\left(x_{6:20}, M_0\right) \\ &= \frac{\Gamma(5.5)\cdot\Gamma(8.5)\cdot\Gamma(16.5)\cdot\Gamma(21.5)}{\Gamma(52)} \cdot \frac{\Gamma(45.5)\cdot\Gamma(53.5)\cdot\Gamma(28.5)\cdot\Gamma(24.5)}{\Gamma(152)} \\ &= \exp(-65.77)\cdot\exp(-205.07) = \exp(-270.84) \\ &= \exp(-270.84) \end{aligned}$$

The ratio between two MMLs constitutes a *Bayes Factor* in favor of one hypothesis over the other. Thus, since $\frac{\exp(-270.84)}{\exp(-280.69)} = 19{,}057$ these data appear to overwhelmingly support a change-point between block 5 and block 6. However, the determination of whether to add a change-point depends on *average* ratio over all possible change-points. This is because the position of the change-point is itself a free parameter that the computation must integrate over. Thus, the Bayes Factor K for the first change-point is determined by:

$$K = \frac{1}{19}\sum_{c=1}^{19}\frac{m\left(x, M_{1:c:20}\right)}{m\left(x, M_0\right)} = \sum_{c=1}^{19}\frac{m\left(x_{1:c}, M_0\right)\cdot m\left(x_{c+1:20}, M_0\right)}{19\cdot m\left(x, M_0\right)} = 2{,}880$$

In order to make a decision whether to add a change-point, the Bayes Factor K must be considered with respect to the *prior odds ratio* with respect to a change, denoted by $\frac{p_1}{p_0}$. The CPR algorithm begins with the assumption that the prior odds for a change in any given interval is $p_c = \frac{1}{n-1}$, and it can be shown that the prior odds ratio for any interval of length n is $p_c \cdot (n-1)$. As additional change-points are identified, the value of p_c is adjusted to equal $p_c = \frac{\#\text{ of CPs}}{n-1}$. This yields the following general criterion[2] for whether to add a change-point to the data in the interval from i to j:

$$\frac{p_1'}{p_0'} = K \cdot \frac{p_1}{p_0} = \left[\frac{1}{j-i}\sum_{c=i}^{j-i}\frac{m(x_{i:j}, M_{i:c:j})}{m(x_{i:j}, M_0)}\right] \cdot \left[p_c \cdot (j-i)\right] \tag{B.4}$$

Given some decision criterion τ (which, by default, was set to 10 for all change-point analyses in this

[2]Equation B.4 requires an additional small-sample correction, without which the contributions of data near the edges of the segment are unduly inflated. This additional calculation is omitted here for the sake of brevity, but is taken into consideration in the calculations that follow. See Jensen (2014 (projected)) for further details.

dissertation), a change-point is added to any segment if $\frac{p_1'}{p_0'} > \tau$. The position of that change-point is selected based on which MML ratio is largest within the segment.

Returning to the example data introduced above, the evidence overwhelmingly favors a change-point, and the interval between block 5 and block 6 had the largest ratio. Thus, the first subdivision of the data yields:

Lever 1:	0 1 2 0 2	3 2 2 6 4 5 5 4 8 1 2 1 1 1 0
Lever 2:	2 2 0 2 2	6 4 4 3 5 3 4 5 2 5 2 1 2 4 3
Lever 3:	2 4 4 3 3	1 3 4 0 0 0 0 1 0 1 2 5 3 4 4
Lever 4:	6 3 4 5 3	0 1 0 1 1 2 1 0 0 3 4 3 4 1 3

The algorithm is then repeated on the two resulting segments, $\mathbf{x}_{1:5}$ and $\mathbf{x}_{6:20}$. When the algorithm is applied to the first of these segments, no additional change-points are identified, but when applied to the second, the evidence favors adding a change-point between blocks 14 and 15:

Lever 1:	0 1 2 0 2	3 2 2 6 4 5 5 4 8	1 2 1 1 1 0
Lever 2:	2 2 0 2 2	6 4 4 3 5 3 4 5 2	5 2 1 2 4 3
Lever 3:	2 4 4 3 3	1 3 4 0 0 0 0 1 0	1 2 5 3 4 4
Lever 4:	6 3 4 5 3	0 1 0 1 1 2 1 0 0	3 4 3 4 1 3

Because $\mathbf{x}_{1:5}$ has already been ruled out, the algorithm now recursively examines the segments $\mathbf{x}_{6:14}$ and $\mathbf{x}_{15:20}$. There is insufficient evidence to subdivide the latter of these two, but the former yields yet another change-point:

Lever 1:	0 1 2 0 2	3 2 2	6 4 5 5 4 8	1 2 1 1 1 0
Lever 2:	2 2 0 2 2	6 4 4	3 5 3 4 5 2	5 2 1 2 4 3
Lever 3:	2 4 4 3 3	1 3 4	0 0 0 0 1 0	1 2 5 3 4 4
Lever 4:	6 3 4 5 3	0 1 0	1 1 2 1 0 0	3 4 3 4 1 3

These are the last change-points identified.

For further details regarding the implementation of the CPR algorithm, see Jensen (2014 (projected)).

Appendix C

Default Transition Matrices

In each case, the default transition matrices are estimated by the procedure described in Chapter 9. Because each subject's behavior consisted of two matrices ($\mathcal{Q}_1$ for responses following reward delivery, and $\mathcal{Q}_0$ for all other responses), the matrices listed below always present $\mathcal{Q}_0$ on the left and $\mathcal{Q}_1$ on the right. In order to facilitate visual inspection of the matrices, each element is color-coded with an intensity of red identical to its value. So, for example, the proportion of 0.40 would have a background color RGB(255,153,153), the proportion 0.00 would have a color RGB(255,255,255), and the proportion 1.00 would have a color RGB(255,0,0).

Experiment 1 (Chapter 9)

Subject 1

0.12	0.58	0.01	0.29	0.02	0.01	0.43	0.55
0.53	0.02	0.37	0.08	0.02	0.02	0.63	0.33
0.02	0.28	0.22	0.48	0.02	0.02	0.67	0.29
0.45	0.01	0.41	0.14	0.02	0.01	0.85	0.12

Subject 2

0.01	0.47	0.45	0.07	0.00	0.00	0.07	0.92
0.36	0.03	0.58	0.02	0.01	0.00	0.12	0.87
0.05	0.44	0.01	0.50	0.00	0.02	0.16	0.82
0.10	0.07	0.80	0.02	0.00	0.00	0.15	0.84

Subject 3

0.02	0.66	0.19	0.13	0.00	0.00	0.04	0.96
0.26	0.05	0.54	0.15	0.00	0.00	0.13	0.86
0.15	0.21	0.24	0.40	0.00	0.00	0.07	0.93
0.16	0.07	0.73	0.04	0.00	0.00	0.39	0.61

Subject 4

0.01	0.50	0.02	0.47	0.01	0.01	0.03	0.95
0.60	0.01	0.36	0.03	0.02	0.01	0.15	0.81
0.04	0.62	0.01	0.34	0.02	0.02	0.39	0.56
0.33	0.02	0.60	0.04	0.01	0.01	0.17	0.82

Subject 5

0.00	0.72	0.14	0.13	0.01	0.01	0.56	0.42
0.11	0.01	0.63	0.25	0.01	0.02	0.58	0.39
0.01	0.29	0.14	0.55	0.01	0.07	0.43	0.49
0.49	0.04	0.37	0.10	0.03	0.03	0.56	0.37

Subject 6

0.04	0.38	0.23	0.36	0.01	0.00	0.76	0.22
0.42	0.02	0.29	0.27	0.02	0.01	0.94	0.04
0.05	0.48	0.03	0.44	0.00	0.01	0.25	0.74
0.38	0.05	0.54	0.03	0.00	0.00	0.70	0.29

Subject 7

0.01	0.39	0.10	0.50	0.01	0.01	0.11	0.87
0.64	0.04	0.26	0.06	0.02	0.02	0.09	0.88
0.01	0.50	0.02	0.48	0.01	0.02	0.47	0.50
0.32	0.04	0.61	0.02	0.01	0.03	0.58	0.39

Subject 8

0.01	0.56	0.06	0.37	0.00	0.01	0.57	0.42
0.59	0.01	0.20	0.20	0.00	0.00	0.42	0.57
0.01	0.45	0.06	0.48	0.00	0.00	0.70	0.30
0.36	0.01	0.60	0.03	0.00	0.00	0.42	0.58

Subject 9

0.09	0.51	0.02	0.38	0.01	0.02	0.07	0.89
0.60	0.22	0.13	0.05	0.01	0.03	0.08	0.89
0.01	0.29	0.21	0.50	0.00	0.01	0.45	0.54
0.16	0.05	0.69	0.10	0.01	0.01	0.15	0.84

Subject 10

0.04	0.52	0.05	0.38	0.00	0.01	0.50	0.49
0.58	0.08	0.27	0.06	0.01	0.01	0.87	0.11
0.08	0.09	0.07	0.77	0.01	0.01	0.39	0.60
0.10	0.12	0.57	0.22	0.01	0.00	0.28	0.71

Subject 1	0.10	0.09	0.01	0.14	0.63	0.03	0.01	0.00	0.84	0.14	0.00	0.00
	0.06	0.01	0.09	0.01	0.12	0.71	0.01	0.01	0.77	0.20	0.00	0.01
	0.01	0.32	0.17	0.49	0.00	0.01	0.03	0.02	0.48	0.47	0.00	0.01
	0.62	0.01	0.32	0.05	0.00	0.00	0.02	0.00	0.87	0.10	0.01	0.00
	0.24	0.47	0.01	0.10	0.09	0.09	0.01	0.01	0.74	0.23	0.00	0.00
	0.24	0.04	0.35	0.02	0.30	0.05	0.01	0.00	0.80	0.18	0.01	0.01

Subject 2	0.01	0.09	0.11	0.10	0.54	0.16	0.00	0.00	0.05	0.95	0.00	0.00
	0.16	0.01	0.34	0.01	0.36	0.13	0.02	0.00	0.11	0.86	0.01	0.00
	0.13	0.40	0.01	0.43	0.01	0.02	0.05	0.00	0.26	0.69	0.00	0.00
	0.35	0.01	0.62	0.01	0.01	0.00	0.07	0.00	0.14	0.78	0.00	0.00
	0.00	0.33	0.14	0.05	0.01	0.46	0.00	0.00	0.04	0.95	0.00	0.00
	0.02	0.00	0.86	0.03	0.07	0.02	0.03	0.00	0.08	0.89	0.00	0.00

Subject 3	0.01	0.39	0.25	0.08	0.18	0.10	0.00	0.00	0.05	0.94	0.00	0.00
	0.45	0.01	0.29	0.04	0.05	0.16	0.01	0.00	0.11	0.87	0.01	0.00
	0.09	0.26	0.13	0.28	0.01	0.23	0.01	0.00	0.25	0.72	0.01	0.00
	0.30	0.02	0.64	0.02	0.02	0.00	0.00	0.00	0.51	0.48	0.00	0.00
	0.01	0.53	0.08	0.22	0.08	0.08	0.00	0.01	0.02	0.95	0.01	0.00
	0.54	0.02	0.30	0.04	0.08	0.01	0.01	0.00	0.12	0.86	0.01	0.00

Subject 4	0.01	0.18	0.01	0.21	0.59	0.01	0.01	0.00	0.04	0.94	0.00	0.00
	0.16	0.01	0.14	0.02	0.00	0.66	0.01	0.04	0.35	0.59	0.00	0.01
	0.01	0.65	0.02	0.29	0.00	0.02	0.03	0.07	0.36	0.52	0.01	0.01
	0.41	0.02	0.54	0.02	0.01	0.00	0.01	0.02	0.23	0.74	0.01	0.00
	0.00	0.67	0.02	0.26	0.01	0.03	0.01	0.01	0.04	0.94	0.00	0.00
	0.69	0.00	0.27	0.01	0.01	0.01	0.01	0.03	0.31	0.65	0.00	0.01

Subject 5	0.00	0.03	0.02	0.09	0.82	0.05	0.01	0.00	0.67	0.32	0.00	0.00
	0.22	0.00	0.07	0.03	0.12	0.55	0.01	0.00	0.48	0.49	0.01	0.01
	0.01	0.36	0.16	0.45	0.01	0.01	0.01	0.03	0.51	0.43	0.00	0.01
	0.43	0.03	0.40	0.05	0.09	0.00	0.01	0.01	0.90	0.08	0.00	0.00
	0.00	0.08	0.06	0.46	0.03	0.36	0.00	0.00	0.65	0.34	0.00	0.00
	0.24	0.00	0.58	0.05	0.11	0.01	0.01	0.00	0.68	0.29	0.01	0.01

Subject 6	0.05	0.11	0.14	0.21	0.16	0.33	0.01	0.00	0.70	0.28	0.00	0.01
	0.44	0.01	0.13	0.11	0.12	0.20	0.01	0.01	0.91	0.06	0.00	0.01
	0.12	0.46	0.01	0.33	0.01	0.07	0.00	0.02	0.29	0.67	0.00	0.01
	0.29	0.04	0.59	0.01	0.04	0.01	0.00	0.00	0.85	0.14	0.00	0.00
	0.02	0.01	0.12	0.55	0.04	0.26	0.00	0.01	0.60	0.37	0.01	0.00
	0.21	0.01	0.24	0.19	0.34	0.01	0.01	0.00	0.90	0.08	0.00	0.01
Subject 7	0.02	0.18	0.01	0.12	0.51	0.16	0.01	0.01	0.05	0.91	0.01	0.01
	0.60	0.02	0.08	0.02	0.05	0.23	0.02	0.01	0.09	0.87	0.00	0.01
	0.04	0.50	0.03	0.42	0.00	0.00	0.02	0.03	0.46	0.48	0.00	0.00
	0.29	0.02	0.65	0.01	0.03	0.00	0.01	0.01	0.25	0.72	0.00	0.00
	0.01	0.26	0.01	0.49	0.02	0.21	0.01	0.01	0.02	0.96	0.00	0.00
	0.18	0.07	0.58	0.05	0.10	0.02	0.01	0.01	0.06	0.89	0.01	0.01
Subject 8	0.02	0.37	0.04	0.31	0.25	0.02	0.00	0.00	0.58	0.41	0.01	0.00
	0.31	0.01	0.05	0.04	0.08	0.51	0.00	0.00	0.52	0.46	0.01	0.01
	0.00	0.34	0.12	0.51	0.00	0.02	0.00	0.00	0.76	0.23	0.00	0.00
	0.35	0.01	0.62	0.01	0.01	0.00	0.00	0.01	0.51	0.48	0.00	0.01
	0.03	0.38	0.05	0.47	0.03	0.04	0.00	0.00	0.39	0.61	0.00	0.00
	0.56	0.00	0.11	0.03	0.29	0.01	0.00	0.01	0.50	0.48	0.00	0.01
Subject 9	0.02	0.18	0.01	0.19	0.43	0.17	0.02	0.01	0.08	0.88	0.01	0.00
	0.29	0.04	0.07	0.01	0.05	0.54	0.01	0.02	0.10	0.86	0.00	0.00
	0.02	0.34	0.20	0.44	0.00	0.00	0.00	0.01	0.55	0.43	0.00	0.01
	0.23	0.04	0.63	0.09	0.00	0.00	0.01	0.01	0.27	0.70	0.00	0.00
	0.02	0.37	0.02	0.20	0.03	0.36	0.01	0.01	0.08	0.89	0.01	0.01
	0.35	0.09	0.15	0.07	0.17	0.17	0.01	0.02	0.12	0.85	0.00	0.00
Subject 10	0.01	0.23	0.01	0.20	0.52	0.02	0.00	0.02	0.56	0.41	0.01	0.00
	0.22	0.02	0.08	0.02	0.03	0.62	0.01	0.01	0.90	0.07	0.01	0.01
	0.02	0.32	0.07	0.56	0.00	0.02	0.00	0.02	0.71	0.27	0.00	0.00
	0.29	0.07	0.49	0.12	0.03	0.01	0.01	0.01	0.34	0.64	0.00	0.00
	0.00	0.52	0.02	0.34	0.02	0.09	0.00	0.03	0.59	0.36	0.00	0.01
	0.63	0.01	0.11	0.03	0.15	0.07	0.01	0.01	0.86	0.11	0.00	0.01

Subject 1	0.14	0.11	0.01	0.19	0.53	0.01	0.00	0.01	0.63	0.26	0.00	0.01	0.09	0.01	0.00	0.00
	0.20	0.05	0.17	0.00	0.01	0.55	0.00	0.00	0.34	0.59	0.01	0.01	0.02	0.03	0.00	0.00
	0.02	0.14	0.07	0.12	0.00	0.00	0.61	0.04	0.12	0.81	0.01	0.01	0.01	0.03	0.01	0.01
	0.15	0.01	0.25	0.04	0.00	0.00	0.03	0.52	0.20	0.76	0.02	0.00	0.01	0.00	0.00	0.01
	0.22	0.29	0.02	0.29	0.12	0.05	0.00	0.01	0.41	0.39	0.01	0.03	0.15	0.00	0.00	0.01
	0.51	0.05	0.34	0.01	0.02	0.05	0.01	0.01	0.67	0.26	0.01	0.01	0.01	0.02	0.01	0.01
	0.03	0.18	0.07	0.53	0.01	0.01	0.07	0.10	0.09	0.86	0.01	0.01	0.00	0.02	0.01	0.00
	0.32	0.03	0.41	0.05	0.01	0.01	0.09	0.08	0.09	0.82	0.01	0.01	0.03	0.01	0.00	0.02

Subject 2	0.02	0.44	0.02	0.10	0.28	0.08	0.03	0.04	0.17	0.65	0.01	0.02	0.06	0.08	0.01	0.00
	0.01	0.02	0.00	0.02	0.36	0.38	0.20	0.01	0.06	0.55	0.00	0.01	0.16	0.20	0.01	0.01
	0.26	0.17	0.01	0.06	0.02	0.04	0.10	0.33	0.02	0.60	0.01	0.01	0.01	0.32	0.02	0.01
	0.28	0.05	0.12	0.01	0.04	0.01	0.36	0.12	0.04	0.75	0.00	0.01	0.02	0.16	0.01	0.01
	0.01	0.22	0.04	0.42	0.02	0.09	0.05	0.15	0.46	0.28	0.01	0.03	0.09	0.10	0.01	0.02
	0.03	0.01	0.01	0.01	0.49	0.02	0.41	0.01	0.10	0.22	0.01	0.01	0.50	0.15	0.01	0.02
	0.28	0.06	0.00	0.28	0.04	0.04	0.02	0.29	0.01	0.71	0.00	0.01	0.03	0.23	0.01	0.01
	0.58	0.09	0.06	0.01	0.06	0.03	0.14	0.03	0.03	0.64	0.00	0.00	0.06	0.23	0.01	0.02

Subject 3	0.09	0.29	0.03	0.01	0.27	0.04	0.03	0.24	0.09	0.48	0.01	0.01	0.29	0.10	0.01	0.02
	0.20	0.05	0.01	0.03	0.08	0.28	0.34	0.01	0.17	0.39	0.00	0.00	0.16	0.26	0.01	0.00
	0.16	0.14	0.10	0.15	0.10	0.10	0.14	0.10	0.04	0.85	0.01	0.03	0.01	0.04	0.01	0.01
	0.51	0.05	0.02	0.05	0.09	0.01	0.01	0.27	0.35	0.50	0.01	0.02	0.05	0.06	0.01	0.01
	0.19	0.32	0.02	0.01	0.10	0.07	0.04	0.25	0.06	0.51	0.00	0.00	0.31	0.11	0.00	0.00
	0.26	0.11	0.01	0.03	0.10	0.05	0.42	0.02	0.20	0.08	0.00	0.00	0.44	0.26	0.01	0.01
	0.13	0.02	0.01	0.67	0.03	0.03	0.02	0.09	0.11	0.73	0.00	0.02	0.01	0.12	0.01	0.00
	0.20	0.02	0.64	0.03	0.07	0.01	0.04	0.01	0.18	0.63	0.04	0.02	0.07	0.05	0.01	0.01

Subject 4	0.01	0.11	0.01	0.06	0.74	0.01	0.00	0.06	0.45	0.41	0.01	0.01	0.09	0.02	0.00	0.01
	0.31	0.02	0.12	0.01	0.02	0.50	0.02	0.00	0.07	0.82	0.00	0.00	0.01	0.09	0.00	0.00
	0.01	0.21	0.01	0.27	0.00	0.01	0.47	0.01	0.03	0.89	0.00	0.00	0.00	0.07	0.01	0.01
	0.26	0.01	0.23	0.02	0.02	0.00	0.02	0.44	0.10	0.87	0.01	0.01	0.00	0.02	0.00	0.00
	0.01	0.33	0.01	0.35	0.02	0.02	0.01	0.25	0.05	0.83	0.00	0.01	0.06	0.03	0.00	0.01
	0.55	0.03	0.25	0.02	0.06	0.03	0.05	0.02	0.28	0.35	0.00	0.01	0.10	0.25	0.00	0.01
	0.02	0.42	0.00	0.47	0.00	0.04	0.00	0.03	0.01	0.91	0.00	0.00	0.00	0.07	0.00	0.00
	0.17	0.03	0.72	0.01	0.01	0.01	0.04	0.02	0.03	0.90	0.01	0.00	0.03	0.03	0.00	0.01

Subject 5	0.05	0.13	0.02	0.22	0.54	0.02	0.01	0.01	0.54	0.35	0.00	0.02	0.06	0.00	0.01	0.01
	0.23	0.08	0.21	0.01	0.05	0.38	0.03	0.00	0.29	0.66	0.01	0.01	0.01	0.01	0.01	0.00
	0.04	0.09	0.02	0.13	0.00	0.00	0.70	0.01	0.49	0.43	0.01	0.02	0.01	0.02	0.01	0.01
	0.28	0.02	0.12	0.03	0.02	0.00	0.12	0.42	0.40	0.55	0.01	0.02	0.02	0.00	0.00	0.01
	0.07	0.40	0.02	0.33	0.10	0.06	0.01	0.02	0.24	0.65	0.01	0.04	0.03	0.01	0.01	0.01
	0.37	0.05	0.31	0.02	0.14	0.04	0.06	0.01	0.51	0.34	0.02	0.03	0.07	0.01	0.01	0.02
	0.06	0.27	0.01	0.53	0.01	0.03	0.04	0.06	0.48	0.44	0.01	0.02	0.02	0.02	0.01	0.00
	0.44	0.06	0.16	0.01	0.04	0.02	0.23	0.05	0.59	0.34	0.01	0.01	0.02	0.01	0.01	0.02

Subject 6

0.04	0.14	0.05	0.32	0.29	0.08	0.02	0.07	0.60	0.32	0.00	0.01	0.05	0.01	0.00	0.00
0.22	0.03	0.25	0.05	0.10	0.30	0.03	0.02	0.18	0.73	0.01	0.01	0.02	0.03	0.00	0.01
0.04	0.19	0.01	0.04	0.05	0.07	0.28	0.32	0.76	0.19	0.02	0.01	0.01	0.01	0.00	0.00
0.07	0.15	0.11	0.01	0.06	0.02	0.32	0.26	0.59	0.25	0.01	0.01	0.11	0.01	0.01	0.01
0.03	0.33	0.04	0.36	0.05	0.10	0.02	0.07	0.10	0.77	0.00	0.01	0.06	0.06	0.00	0.00
0.45	0.03	0.26	0.04	0.14	0.01	0.04	0.03	0.64	0.16	0.01	0.01	0.06	0.11	0.01	0.01
0.05	0.48	0.01	0.06	0.07	0.04	0.02	0.27	0.65	0.28	0.01	0.00	0.02	0.03	0.00	0.00
0.11	0.15	0.21	0.01	0.27	0.04	0.17	0.05	0.80	0.12	0.01	0.00	0.05	0.01	0.00	0.00

Subject 7

0.03	0.06	0.01	0.20	0.58	0.11	0.00	0.02	0.74	0.10	0.00	0.01	0.12	0.03	0.00	0.01
0.38	0.03	0.13	0.01	0.05	0.39	0.01	0.00	0.19	0.63	0.02	0.01	0.01	0.13	0.01	0.00
0.03	0.14	0.04	0.18	0.01	0.02	0.51	0.07	0.23	0.56	0.01	0.01	0.04	0.13	0.02	0.01
0.11	0.02	0.09	0.04	0.02	0.01	0.11	0.61	0.09	0.59	0.04	0.03	0.09	0.10	0.03	0.04
0.03	0.15	0.02	0.42	0.09	0.25	0.00	0.04	0.47	0.14	0.01	0.03	0.22	0.11	0.01	0.01
0.42	0.04	0.30	0.02	0.11	0.06	0.04	0.01	0.60	0.16	0.01	0.02	0.08	0.14	0.00	0.01
0.03	0.49	0.03	0.18	0.01	0.04	0.09	0.12	0.07	0.79	0.01	0.01	0.01	0.09	0.01	0.01
0.30	0.05	0.16	0.04	0.04	0.03	0.31	0.07	0.05	0.69	0.02	0.01	0.07	0.13	0.01	0.02

Subject 8

0.02	0.40	0.01	0.15	0.28	0.05	0.00	0.09	0.54	0.09	0.00	0.01	0.33	0.01	0.01	0.01
0.10	0.03	0.19	0.02	0.14	0.52	0.01	0.01	0.21	0.67	0.00	0.00	0.10	0.02	0.00	0.00
0.02	0.03	0.01	0.62	0.01	0.01	0.18	0.12	0.41	0.14	0.01	0.01	0.39	0.02	0.01	0.02
0.10	0.03	0.01	0.01	0.02	0.00	0.46	0.36	0.29	0.42	0.01	0.01	0.21	0.05	0.01	0.01
0.05	0.50	0.01	0.23	0.05	0.05	0.00	0.12	0.32	0.24	0.01	0.01	0.38	0.03	0.01	0.01
0.21	0.03	0.40	0.02	0.29	0.02	0.02	0.02	0.31	0.26	0.00	0.01	0.37	0.04	0.00	0.01
0.01	0.74	0.01	0.10	0.01	0.07	0.02	0.04	0.05	0.85	0.00	0.00	0.06	0.03	0.00	0.00
0.19	0.04	0.01	0.00	0.05	0.02	0.68	0.01	0.24	0.31	0.01	0.01	0.36	0.04	0.01	0.02

Subject 9

0.03	0.17	0.01	0.29	0.44	0.05	0.00	0.01	0.26	0.69	0.00	0.00	0.03	0.01	0.00	0.00
0.38	0.10	0.06	0.03	0.09	0.33	0.01	0.01	0.03	0.88	0.00	0.00	0.03	0.05	0.00	0.00
0.01	0.16	0.02	0.17	0.01	0.00	0.47	0.15	0.03	0.91	0.00	0.01	0.01	0.03	0.01	0.00
0.06	0.06	0.26	0.02	0.01	0.00	0.17	0.42	0.04	0.89	0.00	0.03	0.02	0.01	0.00	0.01
0.09	0.35	0.01	0.29	0.06	0.16	0.01	0.02	0.07	0.85	0.00	0.01	0.05	0.02	0.00	0.00
0.41	0.08	0.19	0.06	0.18	0.04	0.03	0.01	0.08	0.81	0.01	0.00	0.05	0.03	0.01	0.01
0.04	0.30	0.02	0.32	0.01	0.01	0.04	0.25	0.02	0.93	0.00	0.00	0.01	0.03	0.00	0.00
0.14	0.27	0.27	0.02	0.03	0.01	0.17	0.10	0.04	0.89	0.00	0.01	0.01	0.04	0.00	0.01

Subject 10

0.04	0.23	0.02	0.01	0.40	0.05	0.02	0.23	0.30	0.47	0.00	0.01	0.17	0.05	0.01	0.00
0.14	0.03	0.00	0.02	0.07	0.54	0.20	0.01	0.17	0.63	0.00	0.00	0.12	0.06	0.00	0.00
0.10	0.07	0.08	0.14	0.07	0.23	0.21	0.10	0.05	0.82	0.00	0.01	0.02	0.08	0.02	0.00
0.01	0.04	0.13	0.03	0.14	0.03	0.03	0.57	0.14	0.35	0.01	0.01	0.47	0.02	0.00	0.01
0.02	0.45	0.03	0.01	0.06	0.09	0.02	0.33	0.08	0.68	0.00	0.00	0.12	0.10	0.00	0.01
0.38	0.02	0.00	0.03	0.19	0.04	0.32	0.01	0.48	0.15	0.01	0.01	0.26	0.08	0.01	0.01
0.02	0.01	0.00	0.67	0.04	0.11	0.03	0.12	0.42	0.41	0.01	0.01	0.09	0.04	0.01	0.01
0.01	0.02	0.54	0.01	0.13	0.03	0.23	0.03	0.05	0.35	0.02	0.00	0.48	0.07	0.01	0.01

Experiment 2 (Chapter 10)

Subject 11

0.02	0.17	0.01	0.13	0.65	0.02	0.01	0.01	0.84	0.06	0.00	0.00	0.09	0.01	0.00	0.00
0.22	0.02	0.06	0.01	0.05	0.56	0.07	0.00	0.04	0.81	0.01	0.00	0.02	0.12	0.00	0.00
0.05	0.40	0.02	0.26	0.01	0.02	0.08	0.16	0.02	0.90	0.00	0.00	0.00	0.04	0.01	0.01
0.16	0.10	0.20	0.04	0.01	0.01	0.08	0.40	0.85	0.12	0.00	0.00	0.02	0.00	0.00	0.00
0.01	0.42	0.01	0.44	0.02	0.06	0.01	0.02	0.39	0.45	0.00	0.01	0.11	0.02	0.00	0.00
0.30	0.01	0.14	0.05	0.14	0.10	0.24	0.01	0.22	0.29	0.01	0.01	0.08	0.37	0.01	0.01
0.02	0.15	0.01	0.55	0.01	0.01	0.06	0.20	0.01	0.94	0.01	0.00	0.01	0.02	0.01	0.00
0.25	0.27	0.21	0.02	0.03	0.02	0.09	0.12	0.75	0.20	0.00	0.00	0.02	0.01	0.01	0.01

Subject 12

0.06	0.30	0.02	0.04	0.49	0.04	0.01	0.04	0.64	0.29	0.00	0.00	0.03	0.02	0.01	0.00
0.21	0.06	0.12	0.01	0.05	0.47	0.06	0.01	0.54	0.37	0.01	0.00	0.02	0.05	0.01	0.00
0.05	0.03	0.02	0.47	0.02	0.02	0.28	0.11	0.81	0.12	0.01	0.01	0.02	0.03	0.01	0.01
0.27	0.06	0.19	0.04	0.10	0.01	0.07	0.26	0.27	0.67	0.01	0.01	0.02	0.02	0.00	0.00
0.01	0.59	0.04	0.07	0.04	0.11	0.03	0.11	0.26	0.66	0.01	0.00	0.02	0.04	0.01	0.00
0.24	0.01	0.32	0.04	0.15	0.03	0.19	0.03	0.76	0.09	0.01	0.00	0.07	0.05	0.01	0.01
0.04	0.02	0.01	0.69	0.01	0.01	0.09	0.13	0.82	0.07	0.00	0.00	0.05	0.02	0.01	0.01
0.30	0.09	0.20	0.03	0.14	0.03	0.15	0.06	0.23	0.67	0.01	0.00	0.02	0.04	0.01	0.01

Subject 13

0.03	0.34	0.01	0.20	0.41	0.01	0.00	0.01	0.78	0.19	0.00	0.00	0.02	0.00	0.00	0.00
0.22	0.06	0.14	0.01	0.03	0.53	0.00	0.00	0.12	0.87	0.00	0.00	0.01	0.01	0.00	0.00
0.04	0.16	0.04	0.25	0.00	0.00	0.48	0.02	0.70	0.24	0.01	0.01	0.01	0.00	0.01	0.01
0.06	0.03	0.12	0.04	0.00	0.00	0.09	0.66	0.73	0.13	0.02	0.02	0.01	0.01	0.01	0.07
0.03	0.47	0.02	0.42	0.04	0.02	0.00	0.00	0.35	0.57	0.00	0.01	0.05	0.01	0.00	0.00
0.40	0.04	0.43	0.02	0.05	0.04	0.01	0.00	0.44	0.51	0.01	0.00	0.02	0.03	0.00	0.00
0.10	0.38	0.02	0.36	0.01	0.01	0.06	0.07	0.63	0.27	0.01	0.05	0.00	0.01	0.01	0.01
0.31	0.04	0.30	0.01	0.01	0.00	0.28	0.04	0.78	0.10	0.07	0.00	0.01	0.01	0.01	0.01

Subject 14

0.03	0.13	0.02	0.13	0.65	0.02	0.01	0.02	0.72	0.15	0.01	0.01	0.10	0.01	0.01	0.01
0.32	0.02	0.03	0.01	0.02	0.57	0.02	0.00	0.13	0.72	0.00	0.00	0.03	0.10	0.01	0.00
0.02	0.23	0.04	0.20	0.00	0.01	0.43	0.06	0.19	0.73	0.01	0.01	0.02	0.03	0.01	0.00
0.04	0.04	0.30	0.03	0.01	0.01	0.04	0.54	0.46	0.46	0.01	0.01	0.04	0.01	0.01	0.01
0.01	0.34	0.06	0.45	0.01	0.06	0.01	0.06	0.33	0.45	0.00	0.01	0.14	0.05	0.01	0.01
0.49	0.01	0.31	0.02	0.03	0.01	0.12	0.01	0.58	0.20	0.01	0.01	0.09	0.09	0.01	0.00
0.03	0.31	0.01	0.44	0.02	0.03	0.03	0.14	0.17	0.74	0.00	0.01	0.02	0.03	0.01	0.01
0.28	0.06	0.49	0.02	0.03	0.01	0.08	0.03	0.61	0.29	0.01	0.00	0.07	0.01	0.01	0.00

Subject 15

0.03	0.19	0.01	0.10	0.66	0.01	0.00	0.00	0.94	0.04	0.00	0.00	0.02	0.00	0.00	0.00
0.07	0.03	0.21	0.01	0.01	0.64	0.03	0.01	0.48	0.49	0.00	0.00	0.01	0.01	0.00	0.00
0.10	0.09	0.02	0.26	0.01	0.02	0.31	0.19	0.93	0.05	0.00	0.01	0.00	0.01	0.00	0.00
0.36	0.12	0.10	0.01	0.04	0.02	0.09	0.25	0.95	0.02	0.00	0.00	0.01	0.00	0.00	0.01
0.04	0.73	0.01	0.15	0.02	0.03	0.01	0.01	0.47	0.49	0.00	0.00	0.02	0.00	0.00	0.00
0.10	0.02	0.62	0.04	0.04	0.07	0.09	0.02	0.72	0.24	0.00	0.00	0.01	0.02	0.00	0.01
0.03	0.04	0.01	0.51	0.01	0.01	0.08	0.31	0.90	0.06	0.00	0.01	0.01	0.00	0.00	0.01
0.57	0.08	0.09	0.02	0.08	0.02	0.04	0.09	0.96	0.02	0.00	0.00	0.00	0.01	0.00	0.00

Subject 16

0.01	0.25	0.02	0.14	0.48	0.07	0.01	0.01	0.90	0.06	0.00	0.00	0.02	0.01	0.00	0.00
0.08	0.01	0.23	0.02	0.05	0.56	0.03	0.01	0.51	0.41	0.00	0.00	0.01	0.07	0.01	0.00
0.16	0.04	0.01	0.19	0.01	0.01	0.46	0.12	0.91	0.06	0.00	0.00	0.01	0.01	0.00	0.00
0.29	0.05	0.05	0.03	0.04	0.01	0.18	0.36	0.93	0.03	0.00	0.00	0.02	0.01	0.00	0.00
0.02	0.50	0.03	0.25	0.03	0.13	0.02	0.02	0.70	0.23	0.00	0.00	0.02	0.03	0.00	0.00
0.12	0.01	0.47	0.07	0.21	0.02	0.06	0.02	0.66	0.16	0.01	0.00	0.02	0.14	0.00	0.00
0.19	0.08	0.01	0.29	0.02	0.01	0.01	0.39	0.86	0.09	0.00	0.01	0.02	0.02	0.00	0.00
0.67	0.08	0.01	0.01	0.07	0.03	0.07	0.06	0.93	0.02	0.00	0.00	0.02	0.01	0.00	0.00

Subject 17	0.01	0.04	0.03	0.16	0.20	0.34	0.07	0.15	0.10	0.83	0.00	0.00	0.02	0.04	0.01	0.00
	0.22	0.01	0.04	0.07	0.31	0.12	0.16	0.07	0.66	0.28	0.00	0.00	0.03	0.01	0.01	0.01
	0.04	0.03	0.02	0.04	0.04	0.35	0.44	0.05	0.28	0.67	0.01	0.00	0.01	0.01	0.01	0.00
	0.04	0.02	0.20	0.05	0.23	0.02	0.14	0.30	0.43	0.53	0.00	0.00	0.01	0.03	0.00	0.00
	0.05	0.21	0.03	0.29	0.03	0.15	0.04	0.22	0.02	0.94	0.00	0.00	0.02	0.01	0.00	0.00
	0.22	0.03	0.03	0.02	0.08	0.03	0.56	0.03	0.37	0.54	0.00	0.00	0.03	0.04	0.01	0.00
	0.03	0.04	0.01	0.46	0.08	0.17	0.03	0.18	0.08	0.88	0.00	0.00	0.01	0.02	0.01	0.00
	0.02	0.01	0.46	0.12	0.14	0.02	0.16	0.06	0.60	0.32	0.00	0.01	0.04	0.01	0.01	0.00

Subject 18	0.04	0.20	0.03	0.19	0.46	0.06	0.01	0.01	0.28	0.61	0.00	0.00	0.09	0.00	0.00	0.00
	0.23	0.03	0.15	0.02	0.12	0.45	0.01	0.00	0.05	0.90	0.00	0.00	0.03	0.01	0.00	0.00
	0.10	0.06	0.04	0.25	0.01	0.01	0.36	0.18	0.20	0.75	0.00	0.01	0.02	0.01	0.01	0.01
	0.22	0.04	0.29	0.06	0.02	0.00	0.12	0.25	0.06	0.90	0.00	0.01	0.02	0.00	0.00	0.01
	0.05	0.24	0.04	0.51	0.02	0.10	0.02	0.01	0.12	0.81	0.00	0.01	0.05	0.01	0.00	0.00
	0.16	0.03	0.58	0.01	0.15	0.04	0.01	0.01	0.08	0.79	0.00	0.00	0.10	0.01	0.00	0.01
	0.09	0.15	0.03	0.31	0.02	0.01	0.03	0.37	0.18	0.76	0.00	0.01	0.02	0.01	0.00	0.01
	0.62	0.06	0.08	0.05	0.03	0.01	0.02	0.13	0.14	0.75	0.00	0.01	0.09	0.00	0.00	0.00

Subject 19	0.02	0.22	0.02	0.05	0.56	0.12	0.01	0.01	0.82	0.16	0.00	0.00	0.01	0.00	0.00	0.00
	0.39	0.01	0.10	0.01	0.02	0.46	0.02	0.00	0.07	0.87	0.00	0.00	0.01	0.04	0.00	0.00
	0.06	0.12	0.03	0.15	0.00	0.01	0.47	0.16	0.49	0.48	0.00	0.00	0.00	0.01	0.00	0.00
	0.23	0.08	0.31	0.00	0.01	0.01	0.22	0.13	0.63	0.34	0.01	0.01	0.01	0.01	0.00	0.00
	0.01	0.19	0.08	0.35	0.15	0.20	0.01	0.02	0.47	0.46	0.00	0.00	0.03	0.03	0.00	0.01
	0.35	0.01	0.35	0.05	0.14	0.03	0.06	0.01	0.26	0.65	0.00	0.00	0.01	0.07	0.00	0.00
	0.06	0.25	0.01	0.33	0.01	0.01	0.09	0.24	0.46	0.50	0.00	0.00	0.01	0.02	0.00	0.00
	0.45	0.18	0.13	0.02	0.04	0.01	0.13	0.04	0.68	0.26	0.01	0.00	0.03	0.01	0.00	0.01

Subject 20	0.02	0.32	0.01	0.07	0.47	0.10	0.00	0.00	0.94	0.02	0.00	0.00	0.02	0.01	0.00	0.01
	0.17	0.01	0.25	0.05	0.20	0.31	0.01	0.00	0.63	0.28	0.00	0.01	0.03	0.04	0.01	0.01
	0.04	0.13	0.01	0.27	0.01	0.00	0.35	0.17	0.94	0.01	0.00	0.00	0.02	0.01	0.00	0.00
	0.03	0.08	0.16	0.01	0.00	0.01	0.42	0.29	0.93	0.03	0.00	0.00	0.01	0.01	0.00	0.01
	0.05	0.29	0.01	0.30	0.06	0.29	0.00	0.01	0.87	0.03	0.00	0.00	0.06	0.03	0.00	0.00
	0.06	0.01	0.40	0.20	0.24	0.06	0.02	0.01	0.70	0.13	0.01	0.01	0.08	0.06	0.01	0.01
	0.02	0.35	0.04	0.19	0.01	0.01	0.08	0.30	0.91	0.03	0.00	0.01	0.02	0.01	0.01	0.01
	0.15	0.33	0.08	0.01	0.02	0.02	0.30	0.09	0.95	0.01	0.00	0.00	0.02	0.01	0.01	0.01

Subject 21	0.02	0.07	0.04	0.19	0.52	0.09	0.02	0.05	0.56	0.26	0.00	0.00	0.14	0.02	0.00	0.00
	0.18	0.01	0.19	0.02	0.06	0.51	0.02	0.01	0.36	0.56	0.00	0.00	0.01	0.06	0.00	0.00
	0.13	0.16	0.01	0.15	0.01	0.03	0.31	0.19	0.47	0.48	0.00	0.00	0.00	0.04	0.00	0.00
	0.15	0.03	0.16	0.01	0.01	0.00	0.31	0.32	0.91	0.06	0.00	0.00	0.01	0.00	0.00	0.01
	0.02	0.27	0.07	0.29	0.05	0.18	0.04	0.07	0.17	0.61	0.00	0.00	0.11	0.10	0.00	0.00
	0.23	0.01	0.41	0.04	0.20	0.05	0.05	0.01	0.61	0.28	0.00	0.00	0.02	0.08	0.00	0.00
	0.15	0.41	0.01	0.06	0.01	0.04	0.11	0.21	0.43	0.50	0.00	0.01	0.01	0.04	0.01	0.00
	0.53	0.02	0.19	0.01	0.02	0.01	0.18	0.04	0.94	0.04	0.00	0.00	0.01	0.01	0.00	0.00

Subject 22	0.07	0.16	0.01	0.01	0.61	0.02	0.00	0.10	0.52	0.44	0.01	0.00	0.01	0.02	0.00	0.00
	0.13	0.03	0.06	0.01	0.03	0.67	0.05	0.01	0.09	0.87	0.00	0.00	0.00	0.02	0.01	0.00
	0.02	0.08	0.04	0.26	0.00	0.04	0.55	0.02	0.33	0.61	0.01	0.00	0.01	0.03	0.01	0.01
	0.10	0.02	0.05	0.04	0.01	0.01	0.06	0.71	0.60	0.31	0.00	0.01	0.04	0.02	0.00	0.01
	0.11	0.41	0.02	0.02	0.21	0.03	0.01	0.20	0.19	0.74	0.00	0.00	0.01	0.04	0.01	0.00
	0.56	0.02	0.11	0.02	0.12	0.04	0.11	0.02	0.31	0.61	0.00	0.00	0.01	0.06	0.00	0.00
	0.03	0.25	0.04	0.51	0.00	0.07	0.06	0.03	0.39	0.52	0.00	0.01	0.01	0.06	0.00	0.01
	0.09	0.03	0.65	0.01	0.01	0.01	0.16	0.05	0.49	0.43	0.01	0.00	0.03	0.03	0.01	0.00

Experiment 3 (Chapter 11)

Subject 11 (A)

0.07	0.58	0.02	0.33	0.00	0.00	0.02	0.97
0.48	0.22	0.26	0.05	0.00	0.01	0.94	0.05
0.01	0.40	0.03	0.57	0.01	0.00	0.72	0.26
0.41	0.05	0.48	0.06	0.01	0.00	0.04	0.95

Subject 11 (B)

0.07	0.61	0.00	0.32	0.92	0.07	0.01	0.00
0.54	0.11	0.33	0.03	0.02	0.97	0.01	0.00
0.01	0.33	0.09	0.57	0.06	0.92	0.00	0.02
0.36	0.05	0.48	0.11	0.97	0.02	0.00	0.01

Subject 11 (C)

0.16	0.42	0.28	0.14	0.27	0.48	0.24	0.01
0.13	0.08	0.04	0.75	0.02	0.95	0.01	0.01
0.17	0.48	0.21	0.13	0.15	0.60	0.23	0.02
0.41	0.18	0.14	0.27	0.04	0.89	0.05	0.01

Subject 11 (D)

0.08	0.54	0.01	0.37	0.97	0.01	0.00	0.01
0.43	0.10	0.46	0.02	0.01	0.97	0.02	0.00
0.01	0.23	0.15	0.61	0.03	0.95	0.01	0.01
0.40	0.01	0.48	0.10	0.98	0.01	0.00	0.01

Subject 11 (E)

0.05	0.11	0.84	0.00	0.98	0.01	0.01	0.00
0.17	0.03	0.01	0.79	0.02	0.75	0.00	0.23
0.15	0.79	0.06	0.01	0.96	0.01	0.02	0.01
0.86	0.03	0.09	0.02	0.03	0.42	0.01	0.54

Subject 11 (F)

0.23	0.59	0.01	0.18	0.99	0.01	0.00	0.00
0.42	0.19	0.38	0.01	0.00	0.99	0.00	0.00
0.02	0.12	0.21	0.65	0.03	0.96	0.01	0.00
0.55	0.02	0.37	0.06	0.97	0.02	0.00	0.00

Subject 12 (A)

0.09	0.64	0.07	0.20	0.01	0.01	0.40	0.58
0.44	0.07	0.40	0.09	0.01	0.01	0.24	0.74
0.02	0.41	0.03	0.53	0.00	0.02	0.34	0.63
0.50	0.03	0.40	0.07	0.01	0.01	0.41	0.58

Subject 12 (B)

0.07	0.65	0.04	0.24	0.52	0.46	0.01	0.01
0.46	0.03	0.49	0.02	0.53	0.45	0.01	0.01
0.02	0.21	0.02	0.75	0.53	0.45	0.01	0.01
0.53	0.08	0.30	0.10	0.54	0.45	0.00	0.01

Subject 12 (C)

0.13	0.45	0.39	0.03	0.86	0.09	0.04	0.01
0.19	0.03	0.03	0.75	0.23	0.75	0.01	0.01
0.12	0.69	0.06	0.13	0.79	0.18	0.02	0.01
0.77	0.08	0.07	0.09	0.15	0.83	0.01	0.02

Subject 12 (D)

0.01	0.56	0.06	0.36	0.60	0.38	0.00	0.02
0.49	0.02	0.47	0.02	0.25	0.72	0.02	0.00
0.07	0.14	0.02	0.76	0.62	0.36	0.01	0.01
0.46	0.07	0.44	0.03	0.57	0.42	0.01	0.01

Subject 12 (E)

0.05	0.18	0.75	0.02	0.94	0.04	0.01	0.01
0.66	0.02	0.01	0.31	0.26	0.72	0.01	0.01
0.05	0.68	0.16	0.11	0.75	0.21	0.02	0.01
0.87	0.04	0.07	0.02	0.62	0.34	0.01	0.04

Subject 12 (F)

0.08	0.62	0.02	0.28	0.70	0.30	0.00	0.00
0.70	0.01	0.27	0.01	0.21	0.78	0.01	0.00
0.07	0.05	0.13	0.76	0.29	0.68	0.01	0.01
0.45	0.08	0.40	0.07	0.69	0.29	0.00	0.01

Subject 13 (A)

0.05	0.57	0.07	0.31	0.03	0.02	0.69	0.25
0.45	0.26	0.26	0.02	0.03	0.01	0.71	0.24
0.01	0.42	0.08	0.50	0.01	0.00	0.64	0.35
0.45	0.01	0.40	0.14	0.01	0.00	0.22	0.77

Subject 13 (B)

0.24	0.46	0.01	0.29	0.90	0.09	0.00	0.01
0.57	0.15	0.28	0.01	0.22	0.77	0.01	0.00
0.04	0.40	0.10	0.45	0.69	0.29	0.00	0.02
0.29	0.01	0.64	0.06	0.94	0.05	0.01	0.00

Subject 13 (C)

0.11	0.31	0.55	0.03	0.62	0.35	0.01	0.01
0.11	0.07	0.00	0.82	0.35	0.64	0.01	0.00
0.08	0.84	0.04	0.04	0.43	0.52	0.03	0.01
0.90	0.04	0.02	0.04	0.16	0.81	0.01	0.01

Subject 13 (D)

0.06	0.58	0.01	0.35	0.93	0.06	0.00	0.01
0.50	0.21	0.27	0.02	0.07	0.93	0.00	0.00
0.04	0.24	0.04	0.68	0.59	0.40	0.01	0.01
0.22	0.05	0.51	0.22	0.77	0.21	0.01	0.01

Subject 13 (E)

0.06	0.23	0.70	0.00	0.92	0.07	0.01	0.00
0.31	0.17	0.00	0.52	0.13	0.86	0.00	0.00
0.03	0.78	0.19	0.00	0.84	0.12	0.04	0.00
0.88	0.08	0.02	0.03	0.15	0.83	0.01	0.01

Subject 13 (F)

0.23	0.58	0.01	0.18	0.89	0.10	0.00	0.00
0.55	0.27	0.17	0.01	0.14	0.85	0.00	0.00
0.04	0.37	0.03	0.55	0.36	0.61	0.01	0.03
0.28	0.02	0.60	0.10	0.89	0.09	0.01	0.01

Subject 14 (A)

0.05	0.44	0.00	0.50	0.01	0.02	0.03	0.95
0.58	0.23	0.02	0.18	0.01	0.04	0.14	0.81
0.02	0.66	0.00	0.31	0.02	0.15	0.31	0.52
0.40	0.34	0.11	0.16	0.01	0.15	0.12	0.72

Subject 14 (B)

0.14	0.39	0.01	0.46	0.59	0.36	0.01	0.04
0.40	0.24	0.32	0.04	0.38	0.61	0.00	0.01
0.01	0.48	0.02	0.48	0.08	0.92	0.00	0.00
0.22	0.15	0.61	0.03	0.64	0.33	0.01	0.02

Subject 14 (C)

0.08	0.24	0.65	0.03	0.71	0.19	0.05	0.05
0.22	0.06	0.02	0.69	0.05	0.83	0.05	0.08
0.07	0.76	0.04	0.14	0.47	0.27	0.18	0.07
0.80	0.10	0.07	0.03	0.14	0.66	0.07	0.12

Subject 14 (D)

0.03	0.53	0.04	0.41	0.51	0.46	0.01	0.03
0.63	0.13	0.22	0.02	0.19	0.79	0.01	0.01
0.02	0.45	0.03	0.50	0.02	0.94	0.01	0.02
0.12	0.13	0.66	0.10	0.52	0.46	0.02	0.00

Subject 14 (E)

0.03	0.17	0.78	0.02	0.78	0.19	0.01	0.02
0.45	0.08	0.01	0.46	0.11	0.85	0.01	0.03
0.04	0.87	0.03	0.06	0.40	0.55	0.04	0.01
0.91	0.05	0.02	0.02	0.58	0.37	0.00	0.04

Subject 14 (F)

0.09	0.58	0.02	0.31	0.76	0.21	0.01	0.02
0.65	0.11	0.23	0.01	0.20	0.77	0.01	0.02
0.03	0.37	0.05	0.55	0.20	0.78	0.02	0.00
0.29	0.04	0.62	0.04	0.79	0.19	0.01	0.01

Subject 15 (A)

0.10	0.55	0.20	0.15	0.01	0.01	0.85	0.12
0.22	0.18	0.47	0.14	0.01	0.00	0.70	0.29
0.00	0.37	0.03	0.59	0.01	0.00	0.46	0.53
0.61	0.03	0.28	0.09	0.00	0.00	0.65	0.35

Subject 15 (B)

0.02	0.72	0.04	0.21	0.24	0.74	0.01	0.01
0.30	0.06	0.62	0.02	0.61	0.38	0.00	0.00
0.02	0.03	0.22	0.73	0.83	0.15	0.01	0.01
0.75	0.02	0.18	0.05	0.87	0.12	0.01	0.00

Subject 15 (C)

0.07	0.15	0.48	0.31	0.13	0.63	0.21	0.02
0.09	0.28	0.26	0.38	0.04	0.91	0.04	0.01
0.01	0.15	0.29	0.56	0.13	0.73	0.11	0.03
0.13	0.52	0.24	0.11	0.03	0.91	0.04	0.01

Subject 15 (D)

0.12	0.61	0.01	0.26	0.85	0.15	0.00	0.00
0.25	0.33	0.41	0.02	0.47	0.52	0.01	0.00
0.18	0.13	0.18	0.51	0.90	0.09	0.01	0.01
0.36	0.03	0.36	0.25	0.84	0.15	0.00	0.01

Subject 15 (E)

0.12	0.22	0.65	0.01	0.96	0.04	0.00	0.00
0.28	0.10	0.01	0.61	0.86	0.14	0.00	0.00
0.37	0.54	0.08	0.01	0.85	0.15	0.00	0.00
0.81	0.14	0.02	0.03	0.96	0.03	0.01	0.01

Subject 15 (F)

0.26	0.54	0.01	0.18	0.98	0.02	0.00	0.00
0.55	0.21	0.23	0.00	0.91	0.07	0.01	0.00
0.02	0.07	0.15	0.76	0.98	0.02	0.00	0.00
0.52	0.01	0.42	0.05	0.97	0.02	0.00	0.01

Subject 16 (A)

0.01	0.40	0.35	0.24	0.02	0.00	0.79	0.19
0.37	0.11	0.37	0.16	0.01	0.01	0.89	0.09
0.01	0.40	0.01	0.58	0.01	0.01	0.61	0.37
0.50	0.04	0.37	0.09	0.02	0.01	0.38	0.59

Subject 16 (B)

0.02	0.65	0.02	0.31	0.23	0.75	0.01	0.00
0.56	0.07	0.33	0.05	0.36	0.62	0.00	0.01
0.04	0.08	0.00	0.88	0.66	0.33	0.01	0.00
0.54	0.22	0.18	0.06	0.94	0.06	0.00	0.00

Subject 16 (C)

0.09	0.23	0.48	0.20	0.55	0.42	0.01	0.02
0.34	0.05	0.05	0.56	0.27	0.70	0.02	0.02
0.13	0.46	0.03	0.37	0.25	0.69	0.03	0.02
0.64	0.08	0.09	0.19	0.25	0.70	0.02	0.03

Subject 16 (D)

0.03	0.52	0.01	0.44	0.56	0.42	0.01	0.01
0.35	0.05	0.58	0.02	0.10	0.88	0.02	0.01
0.06	0.26	0.12	0.55	0.57	0.40	0.01	0.01
0.23	0.09	0.61	0.07	0.79	0.18	0.02	0.02

Subject 16 (E)

0.10	0.16	0.73	0.01	0.94	0.05	0.01	0.00
0.21	0.02	0.04	0.73	0.23	0.77	0.00	0.00
0.17	0.71	0.09	0.02	0.84	0.16	0.00	0.00
0.76	0.03	0.13	0.07	0.26	0.73	0.01	0.00

Subject 16 (F)

0.13	0.58	0.01	0.28	0.95	0.05	0.00	0.00
0.66	0.02	0.32	0.01	0.33	0.65	0.01	0.01
0.04	0.24	0.05	0.67	0.68	0.29	0.01	0.02
0.33	0.30	0.33	0.05	0.86	0.14	0.00	0.00

Subject 17 (A)

0.03	0.05	0.04	0.88		0.00	0.01	0.15	0.83
0.36	0.36	0.04	0.24		0.00	0.01	0.24	0.74
0.03	0.85	0.01	0.11		0.00	0.00	0.01	0.99
0.01	0.27	0.56	0.16		0.00	0.01	0.04	0.94

Subject 17 (B)

0.02	0.31	0.03	0.63		0.04	0.96	0.00	0.00
0.59	0.09	0.19	0.14		0.10	0.90	0.00	0.00
0.04	0.62	0.02	0.33		0.02	0.98	0.00	0.00
0.32	0.09	0.56	0.03		0.56	0.43	0.00	0.01

Subject 17 (C)

0.11	0.31	0.51	0.07		0.12	0.58	0.27	0.03
0.23	0.17	0.04	0.56		0.11	0.72	0.13	0.03
0.10	0.44	0.27	0.18		0.02	0.76	0.14	0.08
0.54	0.25	0.09	0.13		0.18	0.60	0.16	0.07

Subject 17 (D)

0.07	0.35	0.05	0.54		0.04	0.95	0.00	0.01
0.53	0.12	0.21	0.14		0.27	0.72	0.01	0.00
0.03	0.33	0.04	0.60		0.40	0.60	0.00	0.00
0.23	0.06	0.35	0.37		0.92	0.07	0.00	0.01

Subject 17 (E)

0.11	0.17	0.46	0.26		0.14	0.82	0.01	0.02
0.65	0.15	0.07	0.14		0.74	0.24	0.01	0.00
0.21	0.18	0.14	0.47		0.40	0.53	0.04	0.04
0.41	0.46	0.05	0.08		0.05	0.92	0.00	0.02

Subject 17 (F)

0.08	0.55	0.03	0.35		0.29	0.70	0.00	0.00
0.54	0.26	0.18	0.01		0.45	0.54	0.00	0.00
0.05	0.37	0.09	0.48		0.08	0.92	0.00	0.00
0.24	0.06	0.60	0.10		0.71	0.28	0.01	0.00

Subject 18 (A)

0.14	0.57	0.05	0.23		0.01	0.07	0.15	0.76
0.35	0.26	0.14	0.26		0.01	0.02	0.02	0.95
0.03	0.54	0.02	0.40		0.00	0.02	0.31	0.67
0.44	0.16	0.25	0.15		0.00	0.01	0.05	0.93

Subject 18 (B)

0.07	0.49	0.01	0.42		0.38	0.61	0.00	0.00
0.60	0.12	0.25	0.03		0.17	0.82	0.00	0.01
0.03	0.26	0.06	0.65		0.31	0.68	0.01	0.01
0.38	0.03	0.39	0.21		0.18	0.81	0.00	0.00

Subject 18 (C)

0.09	0.22	0.59	0.10		0.62	0.36	0.00	0.02
0.41	0.14	0.03	0.43		0.42	0.53	0.03	0.02
0.10	0.35	0.02	0.53		0.72	0.24	0.03	0.02
0.51	0.27	0.05	0.18		0.35	0.62	0.00	0.03

Subject 18 (D)

0.15	0.52	0.02	0.31		0.18	0.81	0.00	0.01
0.58	0.12	0.29	0.01		0.09	0.91	0.00	0.00
0.04	0.25	0.14	0.57		0.19	0.76	0.02	0.03
0.26	0.03	0.46	0.25		0.19	0.77	0.02	0.02

Subject 18 (E)

0.05	0.10	0.75	0.10		0.36	0.61	0.02	0.01
0.78	0.07	0.01	0.15		0.22	0.78	0.01	0.00
0.10	0.42	0.05	0.43		0.12	0.87	0.00	0.01
0.71	0.22	0.01	0.06		0.14	0.83	0.01	0.01

Subject 18 (F)

0.17	0.61	0.02	0.21		0.36	0.63	0.00	0.00
0.65	0.11	0.22	0.02		0.27	0.71	0.01	0.01
0.03	0.15	0.05	0.77		0.70	0.28	0.01	0.01
0.39	0.02	0.34	0.25		0.59	0.38	0.01	0.01

Subject 19 (A)

0.05	0.64	0.10	0.21		0.01	0.00	0.76	0.23
0.40	0.06	0.31	0.23		0.01	0.01	0.70	0.29
0.03	0.20	0.15	0.62		0.00	0.01	0.82	0.17
0.43	0.05	0.48	0.04		0.01	0.01	0.47	0.50

Subject 19 (B)

0.27	0.58	0.04	0.11		0.97	0.02	0.00	0.00
0.66	0.02	0.27	0.04		0.80	0.19	0.00	0.00
0.24	0.04	0.10	0.62		0.92	0.06	0.01	0.01
0.51	0.08	0.39	0.02		0.95	0.04	0.00	0.01

Subject 19 (C)

0.18	0.27	0.33	0.23		0.47	0.24	0.25	0.04
0.05	0.12	0.56	0.27		0.05	0.88	0.07	0.00
0.02	0.28	0.21	0.50		0.15	0.36	0.45	0.03
0.07	0.28	0.49	0.16		0.03	0.88	0.07	0.02

Subject 19 (D)

0.07	0.67	0.02	0.24		0.86	0.13	0.00	0.01
0.64	0.07	0.27	0.02		0.14	0.84	0.00	0.01
0.17	0.16	0.09	0.58		0.84	0.15	0.00	0.01
0.31	0.10	0.46	0.13		0.84	0.15	0.01	0.00

Subject 19 (E)

0.06	0.25	0.65	0.05		0.88	0.06	0.05	0.00
0.56	0.04	0.10	0.30		0.29	0.68	0.02	0.01
0.11	0.57	0.10	0.21		0.81	0.12	0.07	0.00
0.70	0.12	0.17	0.01		0.19	0.76	0.04	0.01

Subject 19 (F)

0.16	0.72	0.02	0.11		0.93	0.07	0.01	0.00
0.75	0.07	0.16	0.02		0.17	0.83	0.00	0.00
0.12	0.17	0.10	0.62		0.94	0.04	0.02	0.00
0.38	0.12	0.47	0.03		0.94	0.05	0.01	0.01

Subject 20 (A)

0.08	0.46	0.04	0.43	0.00	0.00	0.74	0.25
0.52	0.09	0.07	0.32	0.01	0.01	0.51	0.47
0.01	0.28	0.07	0.65	0.01	0.00	0.77	0.22
0.34	0.16	0.47	0.03	0.00	0.00	0.65	0.34

Subject 20 (B)

0.09	0.67	0.00	0.23	0.94	0.05	0.00	0.01
0.64	0.02	0.28	0.06	0.83	0.17	0.00	0.00
0.02	0.43	0.12	0.43	0.93	0.06	0.00	0.00
0.18	0.19	0.55	0.08	0.87	0.10	0.00	0.02

Subject 20 (C)

0.09	0.31	0.39	0.20	0.58	0.39	0.01	0.02
0.46	0.06	0.15	0.33	0.65	0.33	0.00	0.01
0.20	0.14	0.08	0.58	0.29	0.68	0.01	0.02
0.44	0.17	0.27	0.12	0.22	0.75	0.01	0.02

Subject 20 (D)

0.10	0.51	0.01	0.38	0.97	0.02	0.01	0.01
0.76	0.03	0.18	0.03	0.75	0.23	0.01	0.01
0.02	0.62	0.08	0.27	0.97	0.02	0.01	0.01
0.01	0.16	0.72	0.11	0.98	0.02	0.01	0.00

Subject 20 (E)

0.03	0.26	0.67	0.03	0.97	0.01	0.01	0.00
0.44	0.04	0.16	0.35	0.60	0.39	0.01	0.00
0.20	0.52	0.10	0.19	0.95	0.03	0.01	0.00
0.79	0.05	0.15	0.02	0.92	0.06	0.01	0.01

Subject 20 (F)

0.10	0.61	0.01	0.27	0.98	0.01	0.01	0.00
0.75	0.02	0.21	0.02	0.66	0.33	0.01	0.00
0.04	0.43	0.10	0.43	0.98	0.02	0.00	0.00
0.18	0.11	0.66	0.05	0.97	0.03	0.00	0.01

Subject 21 (A)

0.01	0.37	0.13	0.49	0.00	0.01	0.54	0.45
0.49	0.06	0.34	0.11	0.00	0.00	0.96	0.03
0.05	0.48	0.06	0.42	0.00	0.00	0.80	0.20
0.46	0.01	0.41	0.12	0.00	0.01	0.37	0.62

Subject 21 (B)

0.10	0.46	0.02	0.42	0.85	0.15	0.00	0.00
0.53	0.07	0.39	0.01	0.42	0.57	0.00	0.00
0.06	0.44	0.07	0.44	0.74	0.24	0.02	0.00
0.37	0.01	0.54	0.07	0.97	0.02	0.00	0.00

Subject 21 (C)

0.10	0.25	0.52	0.14	0.90	0.07	0.03	0.01
0.28	0.08	0.04	0.60	0.03	0.92	0.01	0.04
0.10	0.55	0.11	0.24	0.72	0.22	0.05	0.01
0.67	0.14	0.04	0.14	0.11	0.85	0.01	0.03

Subject 21 (D)

0.04	0.44	0.05	0.47	0.47	0.50	0.02	0.01
0.42	0.05	0.50	0.03	0.35	0.61	0.04	0.01
0.13	0.34	0.08	0.45	0.88	0.09	0.02	0.01
0.27	0.02	0.61	0.10	0.95	0.03	0.00	0.02

Subject 21 (E)

0.07	0.16	0.67	0.11	0.86	0.12	0.01	0.01
0.47	0.03	0.04	0.46	0.31	0.67	0.00	0.02
0.15	0.56	0.08	0.21	0.38	0.57	0.01	0.03
0.87	0.06	0.05	0.02	0.57	0.41	0.01	0.02

Subject 21 (F)

0.24	0.44	0.02	0.30	0.79	0.20	0.00	0.00
0.57	0.10	0.32	0.01	0.32	0.67	0.00	0.00
0.09	0.34	0.09	0.47	0.85	0.13	0.01	0.01
0.27	0.01	0.63	0.09	0.97	0.01	0.00	0.01

Subject 22 (A)

0.24	0.31	0.01	0.44	0.02	0.02	0.01	0.94
0.65	0.21	0.02	0.12	0.01	0.04	0.01	0.94
0.05	0.66	0.10	0.19	0.01	0.02	0.04	0.93
0.19	0.30	0.27	0.25	0.01	0.01	0.04	0.93

Subject 22 (B)

0.41	0.25	0.02	0.32	0.32	0.67	0.01	0.01
0.71	0.19	0.07	0.03	0.15	0.83	0.01	0.00
0.01	0.67	0.10	0.21	0.04	0.95	0.01	0.00
0.06	0.07	0.79	0.08	0.09	0.87	0.02	0.02

Subject 22 (C)

0.10	0.14	0.75	0.01	0.64	0.34	0.01	0.01
0.09	0.11	0.02	0.78	0.26	0.70	0.01	0.03
0.12	0.71	0.14	0.02	0.65	0.32	0.01	0.02
0.52	0.11	0.29	0.08	0.40	0.56	0.01	0.04

Subject 22 (D)

0.14	0.54	0.01	0.31	0.73	0.26	0.00	0.01
0.48	0.27	0.25	0.01	0.26	0.74	0.01	0.00
0.03	0.26	0.14	0.57	0.43	0.56	0.01	0.00
0.22	0.01	0.49	0.28	0.75	0.24	0.00	0.01

Subject 22 (E)

0.08	0.24	0.67	0.02	0.75	0.24	0.01	0.00
0.21	0.07	0.08	0.63	0.30	0.69	0.01	0.01
0.41	0.44	0.10	0.06	0.62	0.38	0.00	0.00
0.55	0.06	0.35	0.04	0.57	0.41	0.01	0.00

Subject 22 (F)

0.39	0.40	0.00	0.21	0.74	0.25	0.00	0.00
0.55	0.18	0.26	0.02	0.26	0.72	0.01	0.01
0.03	0.29	0.08	0.60	0.47	0.50	0.02	0.01
0.18	0.04	0.59	0.18	0.71	0.27	0.01	0.01

Experiment 4 (Chapter 12)

Subject 23 (4P1)

0.14	0.44	0.31	0.12	0.04	0.04	0.29	0.63
0.24	0.13	0.49	0.15	0.02	0.03	0.27	0.68
0.06	0.09	0.32	0.53	0.00	0.00	0.43	0.56
0.34	0.19	0.24	0.24	0.01	0.02	0.17	0.79

Subject 23 (4R4)

0.05	0.81	0.11	0.03	0.00	0.00	0.98	0.01
0.09	0.07	0.80	0.03	0.00	0.00	1.00	0.00
0.12	0.01	0.19	0.67	0.00	0.00	0.98	0.01
0.68	0.01	0.18	0.13	0.00	0.00	0.89	0.10

Subject 23 (8P2)

0.03	0.32	0.02	0.03	0.45	0.14	0.01	0.00	0.68	0.18	0.01	0.02	0.05	0.03	0.01	0.01
0.13	0.01	0.30	0.02	0.09	0.43	0.02	0.00	0.58	0.28	0.00	0.01	0.06	0.07	0.00	0.00
0.06	0.03	0.02	0.40	0.01	0.01	0.22	0.26	0.44	0.41	0.01	0.00	0.07	0.05	0.02	0.01
0.21	0.05	0.09	0.03	0.03	0.01	0.13	0.45	0.28	0.57	0.01	0.01	0.06	0.05	0.01	0.01
0.03	0.38	0.02	0.09	0.28	0.19	0.01	0.01	0.64	0.22	0.01	0.01	0.06	0.04	0.01	0.01
0.09	0.03	0.48	0.06	0.19	0.08	0.06	0.02	0.62	0.12	0.01	0.01	0.11	0.11	0.01	0.01
0.07	0.14	0.03	0.19	0.01	0.02	0.31	0.22	0.28	0.50	0.00	0.00	0.11	0.07	0.01	0.01
0.44	0.08	0.03	0.04	0.05	0.02	0.19	0.16	0.43	0.48	0.01	0.01	0.04	0.02	0.01	0.01

Subject 23 (8R3)

0.01	0.33	0.01	0.01	0.58	0.05	0.01	0.01	0.94	0.01	0.01	0.00	0.02	0.00	0.01	0.00
0.03	0.01	0.36	0.01	0.06	0.45	0.08	0.01	0.82	0.03	0.00	0.00	0.12	0.02	0.01	0.01
0.01	0.01	0.00	0.73	0.00	0.01	0.14	0.10	0.94	0.00	0.00	0.00	0.02	0.02	0.00	0.00
0.15	0.02	0.02	0.01	0.02	0.01	0.48	0.29	0.96	0.00	0.00	0.00	0.01	0.01	0.01	0.00
0.01	0.69	0.01	0.02	0.12	0.09	0.04	0.01	0.90	0.02	0.00	0.00	0.05	0.01	0.00	0.01
0.05	0.01	0.48	0.02	0.15	0.03	0.22	0.03	0.87	0.00	0.00	0.00	0.06	0.03	0.02	0.01
0.03	0.02	0.01	0.05	0.02	0.03	0.37	0.47	0.87	0.02	0.01	0.00	0.06	0.04	0.01	0.00
0.54	0.03	0.02	0.00	0.08	0.02	0.21	0.10	0.92	0.01	0.00	0.01	0.04	0.01	0.00	0.00

Subject 24 (4P1)

0.04	0.24	0.17	0.55	0.05	0.02	0.44	0.49
0.32	0.04	0.30	0.35	0.09	0.12	0.58	0.22
0.13	0.36	0.05	0.46	0.01	0.02	0.72	0.25
0.13	0.06	0.70	0.11	0.01	0.00	0.14	0.85

Subject 24 (4R4)

0.09	0.31	0.15	0.44	0.01	0.00	0.02	0.97
0.05	0.13	0.63	0.19	0.00	0.00	0.98	0.02
0.00	0.33	0.25	0.42	0.00	0.00	0.98	0.01
0.02	0.11	0.81	0.06	0.00	0.00	0.02	0.98

Subject 24 (8P2)

0.06	0.17	0.02	0.24	0.47	0.02	0.00	0.02	0.87	0.04	0.01	0.00	0.04	0.02	0.00	0.01
0.08	0.04	0.22	0.01	0.11	0.53	0.01	0.01	0.01	0.91	0.01	0.00	0.01	0.05	0.01	0.01
0.06	0.31	0.05	0.08	0.01	0.02	0.22	0.25	0.04	0.80	0.02	0.00	0.03	0.09	0.01	0.01
0.20	0.14	0.31	0.02	0.03	0.01	0.14	0.15	0.53	0.10	0.01	0.04	0.30	0.01	0.00	0.01
0.24	0.36	0.01	0.24	0.08	0.05	0.00	0.02	0.62	0.03	0.01	0.01	0.30	0.03	0.00	0.01
0.21	0.02	0.25	0.03	0.43	0.03	0.01	0.01	0.03	0.67	0.00	0.00	0.04	0.25	0.00	0.01
0.05	0.24	0.09	0.11	0.03	0.09	0.08	0.31	0.05	0.77	0.01	0.01	0.06	0.08	0.01	0.01
0.41	0.11	0.13	0.05	0.06	0.02	0.13	0.09	0.49	0.07	0.03	0.03	0.26	0.06	0.03	0.03

Subject 24 (8R3)

0.02	0.15	0.00	0.28	0.50	0.02	0.00	0.03	0.96	0.00	0.00	0.00	0.02	0.00	0.00	0.00
0.14	0.02	0.03	0.01	0.02	0.74	0.04	0.00	0.00	0.97	0.01	0.00	0.00	0.01	0.01	0.00
0.02	0.32	0.02	0.22	0.01	0.01	0.18	0.23	0.03	0.94	0.00	0.00	0.00	0.01	0.00	0.00
0.11	0.07	0.14	0.01	0.03	0.00	0.41	0.23	0.96	0.01	0.00	0.01	0.02	0.00	0.01	0.00
0.08	0.50	0.00	0.32	0.03	0.03	0.00	0.04	0.72	0.04	0.00	0.01	0.22	0.00	0.01	0.00
0.61	0.01	0.11	0.02	0.11	0.01	0.12	0.01	0.10	0.66	0.01	0.01	0.03	0.17	0.01	0.01
0.02	0.47	0.03	0.20	0.01	0.02	0.09	0.17	0.03	0.92	0.00	0.01	0.01	0.03	0.00	0.00
0.25	0.10	0.11	0.02	0.05	0.01	0.35	0.11	0.92	0.02	0.00	0.00	0.04	0.00	0.01	0.00

Subject 25 (4P2)

0.04	0.52	0.42	0.02	0.01	0.02	0.80	0.17
0.46	0.10	0.02	0.42	0.01	0.04	0.56	0.39
0.43	0.04	0.06	0.46	0.01	0.01	0.89	0.10
0.02	0.42	0.50	0.06	0.00	0.00	0.09	0.90

Subject 25 (4R3)

0.01	0.46	0.02	0.52	0.01	0.01	0.11	0.87
0.58	0.05	0.34	0.02	0.02	0.08	0.28	0.62
0.02	0.51	0.04	0.43	0.00	0.00	0.92	0.08
0.29	0.06	0.62	0.04	0.00	0.00	0.04	0.96

Subject 25 (8P1)

0.08	0.19	0.01	0.06	0.05	0.52	0.00	0.10	0.65	0.19	0.01	0.01	0.04	0.09	0.01	0.01
0.13	0.09	0.06	0.04	0.61	0.02	0.05	0.00	0.16	0.67	0.00	0.01	0.12	0.04	0.01	0.00
0.02	0.22	0.06	0.28	0.03	0.01	0.35	0.03	0.19	0.61	0.01	0.02	0.10	0.04	0.01	0.01
0.15	0.03	0.18	0.11	0.00	0.06	0.01	0.46	0.29	0.52	0.01	0.06	0.02	0.06	0.01	0.03
0.45	0.08	0.14	0.06	0.09	0.07	0.10	0.01	0.22	0.42	0.01	0.01	0.29	0.04	0.01	0.01
0.07	0.46	0.03	0.05	0.10	0.10	0.01	0.19	0.29	0.33	0.00	0.00	0.12	0.25	0.00	0.00
0.03	0.15	0.04	0.67	0.04	0.01	0.03	0.04	0.21	0.26	0.08	0.08	0.11	0.09	0.08	0.08
0.17	0.04	0.59	0.03	0.01	0.06	0.03	0.07	0.38	0.40	0.01	0.03	0.05	0.10	0.00	0.02

Subject 25 (8R4)

0.04	0.21	0.00	0.16	0.57	0.00	0.00	0.01	0.88	0.02	0.00	0.00	0.08	0.00	0.00	0.01
0.16	0.03	0.18	0.01	0.00	0.62	0.01	0.00	0.03	0.93	0.01	0.00	0.01	0.03	0.00	0.00
0.00	0.11	0.02	0.39	0.00	0.04	0.44	0.00	0.49	0.45	0.01	0.01	0.02	0.01	0.02	0.01
0.12	0.01	0.12	0.03	0.01	0.00	0.01	0.69	0.63	0.27	0.01	0.01	0.05	0.01	0.01	0.01
0.01	0.57	0.01	0.38	0.01	0.01	0.00	0.01	0.43	0.14	0.00	0.02	0.39	0.01	0.01	0.00
0.48	0.02	0.43	0.01	0.01	0.02	0.02	0.01	0.04	0.69	0.02	0.00	0.00	0.23	0.02	0.00
0.02	0.16	0.01	0.66	0.00	0.06	0.07	0.01	0.28	0.67	0.00	0.01	0.01	0.01	0.01	0.00
0.32	0.01	0.56	0.02	0.03	0.00	0.04	0.02	0.55	0.33	0.01	0.01	0.05	0.01	0.01	0.01

Subject 26 (4P2)

0.02	0.42	0.07	0.50	0.01	0.01	0.10	0.88
0.51	0.11	0.31	0.07	0.01	0.00	0.12	0.87
0.02	0.36	0.06	0.56	0.00	0.00	0.12	0.87
0.22	0.05	0.56	0.17	0.00	0.00	0.09	0.91

Subject 26 (4R3)

0.03	0.43	0.05	0.49	0.01	0.01	0.16	0.83
0.65	0.02	0.24	0.09	0.01	0.01	0.49	0.50
0.05	0.44	0.04	0.48	0.00	0.00	0.52	0.48
0.22	0.03	0.55	0.20	0.00	0.00	0.21	0.79

Subject 26 (8P1)

0.09	0.18	0.07	0.02	0.49	0.04	0.01	0.10	0.58	0.24	0.02	0.01	0.06	0.03	0.03	0.01
0.20	0.08	0.03	0.02	0.04	0.56	0.06	0.01	0.19	0.65	0.01	0.01	0.07	0.04	0.01	0.01
0.06	0.16	0.06	0.15	0.01	0.07	0.42	0.07	0.18	0.67	0.02	0.02	0.05	0.02	0.02	0.01
0.04	0.13	0.13	0.14	0.09	0.05	0.04	0.39	0.24	0.50	0.05	0.03	0.04	0.08	0.03	0.03
0.06	0.52	0.05	0.02	0.10	0.15	0.02	0.09	0.47	0.24	0.00	0.00	0.22	0.06	0.00	0.00
0.41	0.05	0.10	0.04	0.07	0.15	0.15	0.03	0.36	0.51	0.01	0.01	0.06	0.03	0.01	0.01
0.04	0.08	0.04	0.49	0.03	0.06	0.05	0.21	0.20	0.58	0.02	0.02	0.04	0.09	0.04	0.02
0.04	0.11	0.38	0.07	0.10	0.05	0.09	0.16	0.26	0.47	0.02	0.03	0.10	0.06	0.03	0.04

Subject 26 (8R4)

0.02	0.23	0.00	0.08	0.59	0.05	0.00	0.03	0.63	0.27	0.01	0.00	0.04	0.04	0.00	0.01
0.28	0.02	0.15	0.01	0.06	0.46	0.00	0.01	0.26	0.64	0.00	0.00	0.02	0.06	0.00	0.01
0.01	0.03	0.01	0.23	0.01	0.08	0.60	0.03	0.26	0.60	0.00	0.01	0.01	0.10	0.02	0.01
0.16	0.02	0.35	0.01	0.01	0.02	0.02	0.41	0.80	0.11	0.01	0.01	0.02	0.02	0.00	0.03
0.01	0.47	0.01	0.29	0.01	0.11	0.00	0.10	0.28	0.66	0.00	0.01	0.02	0.02	0.00	0.01
0.48	0.03	0.31	0.01	0.06	0.09	0.01	0.01	0.40	0.52	0.00	0.00	0.02	0.04	0.00	0.01
0.02	0.06	0.01	0.58	0.01	0.20	0.02	0.11	0.25	0.60	0.00	0.01	0.02	0.09	0.02	0.02
0.27	0.03	0.54	0.02	0.02	0.04	0.04	0.03	0.75	0.13	0.00	0.00	0.03	0.08	0.00	0.01

Subject 27 (4P1)

0.19	0.33	0.12	0.35	0.01	0.00	0.39	0.59
0.60	0.08	0.30	0.02	0.00	0.00	0.43	0.56
0.01	0.56	0.09	0.34	0.00	0.01	0.58	0.41
0.39	0.01	0.58	0.02	0.01	0.00	0.30	0.69

Subject 27 (4R4)

0.19	0.33	0.12	0.35	0.01	0.00	0.39	0.59
0.60	0.08	0.30	0.02	0.00	0.00	0.43	0.56
0.01	0.56	0.09	0.34	0.00	0.01	0.58	0.41
0.39	0.01	0.58	0.02	0.01	0.00	0.30	0.69

Subject 27 (8P2)

0.13	0.08	0.01	0.29	0.46	0.03	0.00	0.00	0.37	0.55	0.01	0.01	0.01	0.04	0.01	0.01
0.08	0.03	0.39	0.02	0.04	0.43	0.01	0.00	0.01	0.94	0.00	0.01	0.01	0.03	0.00	0.00
0.01	0.22	0.11	0.20	0.06	0.07	0.20	0.12	0.16	0.75	0.00	0.00	0.05	0.02	0.01	0.01
0.03	0.08	0.39	0.03	0.29	0.01	0.06	0.11	0.04	0.89	0.01	0.00	0.04	0.01	0.01	0.00
0.07	0.40	0.01	0.37	0.07	0.06	0.00	0.00	0.29	0.62	0.01	0.01	0.02	0.04	0.01	0.01
0.38	0.02	0.19	0.05	0.25	0.08	0.02	0.01	0.08	0.80	0.00	0.00	0.01	0.10	0.00	0.00
0.01	0.29	0.06	0.16	0.02	0.01	0.12	0.33	0.09	0.81	0.01	0.00	0.05	0.03	0.00	0.00
0.03	0.12	0.39	0.02	0.13	0.04	0.13	0.15	0.12	0.79	0.01	0.00	0.06	0.00	0.00	0.00

Subject 27 (8R3)

0.07	0.13	0.00	0.22	0.54	0.02	0.00	0.00	0.66	0.30	0.01	0.00	0.01	0.02	0.00	0.00
0.07	0.02	0.20	0.01	0.01	0.66	0.02	0.00	0.14	0.73	0.00	0.00	0.01	0.10	0.00	0.00
0.02	0.12	0.05	0.04	0.00	0.01	0.49	0.27	0.41	0.55	0.00	0.00	0.01	0.01	0.00	0.00
0.18	0.02	0.34	0.07	0.01	0.00	0.15	0.23	0.17	0.79	0.00	0.01	0.00	0.01	0.00	0.00
0.06	0.20	0.01	0.61	0.09	0.03	0.00	0.01	0.61	0.32	0.01	0.01	0.01	0.03	0.01	0.01
0.56	0.02	0.21	0.04	0.10	0.03	0.03	0.01	0.42	0.28	0.00	0.00	0.02	0.26	0.01	0.01
0.02	0.31	0.05	0.03	0.00	0.02	0.11	0.46	0.15	0.79	0.01	0.01	0.02	0.03	0.01	0.00
0.19	0.06	0.28	0.09	0.01	0.00	0.24	0.13	0.36	0.57	0.00	0.01	0.02	0.03	0.00	0.01

Subject 28 (4P4)

0.15	0.42	0.02	0.41	0.00	0.00	0.01	0.99
0.01	0.14	0.63	0.22	0.00	0.00	0.61	0.39
0.00	0.22	0.34	0.44	0.00	0.00	0.90	0.09
0.01	0.09	0.62	0.28	0.00	0.00	0.02	0.98

Subject 28 (4R1)

0.07	0.19	0.24	0.50	0.01	0.01	0.13	0.85
0.38	0.07	0.18	0.36	0.02	0.01	0.33	0.64
0.15	0.43	0.09	0.33	0.01	0.01	0.61	0.36
0.14	0.04	0.72	0.10	0.00	0.00	0.18	0.82

Subject 28 (8P3)

0.04	0.07	0.00	0.11	0.77	0.01	0.00	0.00	0.80	0.09	0.01	0.01	0.07	0.01	0.01	0.01
0.13	0.04	0.11	0.01	0.01	0.67	0.01	0.00	0.02	0.94	0.00	0.00	0.01	0.02	0.01	0.01
0.01	0.17	0.06	0.17	0.00	0.01	0.43	0.13	0.01	0.97	0.00	0.00	0.01	0.01	0.00	0.00
0.09	0.03	0.23	0.06	0.01	0.00	0.18	0.40	0.16	0.72	0.01	0.01	0.07	0.02	0.00	0.01
0.03	0.41	0.00	0.28	0.22	0.05	0.00	0.01	0.33	0.19	0.01	0.01	0.42	0.03	0.01	0.01
0.61	0.04	0.22	0.02	0.04	0.03	0.02	0.01	0.06	0.87	0.00	0.00	0.00	0.06	0.01	0.00
0.01	0.27	0.10	0.21	0.00	0.01	0.21	0.19	0.00	0.96	0.01	0.00	0.01	0.01	0.00	0.00
0.10	0.06	0.26	0.15	0.01	0.01	0.30	0.11	0.32	0.58	0.00	0.01	0.05	0.03	0.00	0.01

Subject 28 (8R2)

0.15	0.12	0.03	0.29	0.28	0.11	0.01	0.01	0.27	0.54	0.01	0.01	0.01	0.16	0.01	0.01
0.27	0.04	0.14	0.04	0.05	0.44	0.01	0.01	0.01	0.93	0.00	0.01	0.01	0.03	0.01	0.00
0.07	0.26	0.06	0.15	0.01	0.01	0.27	0.17	0.00	0.96	0.00	0.01	0.01	0.01	0.00	0.00
0.06	0.16	0.22	0.07	0.01	0.01	0.20	0.26	0.07	0.88	0.01	0.01	0.01	0.02	0.01	0.01
0.06	0.18	0.05	0.41	0.12	0.15	0.01	0.03	0.20	0.60	0.01	0.01	0.02	0.14	0.01	0.01
0.49	0.03	0.20	0.10	0.12	0.02	0.02	0.02	0.16	0.66	0.00	0.00	0.03	0.13	0.00	0.00
0.04	0.24	0.07	0.19	0.01	0.01	0.17	0.28	0.01	0.93	0.00	0.02	0.00	0.02	0.01	0.01
0.10	0.17	0.28	0.06	0.01	0.02	0.26	0.11	0.26	0.68	0.01	0.01	0.01	0.03	0.01	0.01

Subject 29 (4P2)

0.03	0.75	0.20	0.02	0.01	0.00	0.44	0.55
0.26	0.19	0.46	0.09	0.01	0.01	0.19	0.79
0.27	0.00	0.15	0.57	0.01	0.00	0.85	0.14
0.33	0.01	0.61	0.06	0.01	0.00	0.13	0.86

Subject 29 (4R3)

0.03	0.76	0.03	0.18	0.02	0.01	0.69	0.28
0.47	0.08	0.18	0.27	0.01	0.01	0.93	0.04
0.08	0.12	0.03	0.77	0.01	0.01	0.91	0.08
0.29	0.03	0.43	0.25	0.01	0.00	0.16	0.83

Subject 29 (8P1)

0.03	0.18	0.08	0.09	0.04	0.54	0.02	0.02	0.87	0.08	0.00	0.00	0.01	0.03	0.01	0.01
0.18	0.05	0.05	0.03	0.65	0.02	0.01	0.00	0.14	0.82	0.01	0.01	0.01	0.01	0.01	0.01
0.05	0.27	0.03	0.30	0.01	0.00	0.28	0.06	0.61	0.36	0.00	0.01	0.00	0.00	0.00	0.01
0.08	0.15	0.15	0.06	0.01	0.01	0.03	0.51	0.76	0.14	0.02	0.02	0.01	0.03	0.01	0.01
0.53	0.07	0.12	0.08	0.08	0.06	0.04	0.01	0.27	0.65	0.00	0.00	0.05	0.01	0.00	0.00
0.04	0.55	0.10	0.11	0.15	0.03	0.02	0.01	0.75	0.17	0.01	0.01	0.02	0.04	0.01	0.01
0.05	0.23	0.03	0.50	0.01	0.01	0.06	0.11	0.57	0.40	0.01	0.00	0.01	0.00	0.00	0.00
0.11	0.24	0.44	0.02	0.02	0.01	0.09	0.07	0.84	0.06	0.01	0.02	0.02	0.03	0.01	0.01

Subject 29 (8R4)

0.02	0.22	0.02	0.11	0.50	0.12	0.00	0.01	0.94	0.02	0.01	0.00	0.03	0.00	0.00	0.00
0.11	0.06	0.20	0.01	0.03	0.59	0.00	0.00	0.13	0.83	0.00	0.00	0.01	0.03	0.00	0.00
0.08	0.03	0.01	0.71	0.00	0.00	0.10	0.06	0.93	0.04	0.01	0.00	0.01	0.00	0.01	0.00
0.07	0.02	0.09	0.03	0.00	0.00	0.08	0.69	0.96	0.02	0.00	0.01	0.01	0.00	0.00	0.01
0.01	0.47	0.03	0.12	0.04	0.33	0.00	0.00	0.65	0.29	0.00	0.00	0.04	0.01	0.00	0.00
0.18	0.12	0.34	0.02	0.22	0.12	0.00	0.00	0.27	0.60	0.00	0.00	0.03	0.08	0.00	0.00
0.27	0.07	0.03	0.18	0.02	0.02	0.11	0.30	0.91	0.02	0.01	0.01	0.01	0.01	0.02	0.01
0.15	0.03	0.24	0.03	0.01	0.01	0.45	0.08	0.93	0.04	0.01	0.00	0.00	0.01	0.01	0.00

Subject 30 (4P3)

0.02	0.43	0.01	0.54	0.00	0.01	0.15	0.84
0.57	0.08	0.32	0.03	0.01	0.00	0.86	0.12
0.00	0.39	0.06	0.54	0.00	0.00	0.83	0.16
0.27	0.01	0.69	0.02	0.00	0.00	0.82	0.17

Subject 30 (4R2)

0.02	0.19	0.01	0.77	0.01	0.00	0.17	0.82
0.67	0.05	0.19	0.09	0.00	0.00	0.98	0.02
0.01	0.43	0.04	0.52	0.00	0.00	0.80	0.20
0.12	0.04	0.81	0.04	0.00	0.00	0.76	0.24

Subject 30 (8P4)

0.04	0.28	0.01	0.43	0.15	0.04	0.01	0.05	0.22	0.18	0.01	0.01	0.51	0.06	0.01	0.01
0.09	0.04	0.13	0.01	0.29	0.41	0.03	0.00	0.02	0.86	0.00	0.00	0.06	0.03	0.01	0.01
0.01	0.34	0.04	0.44	0.01	0.03	0.10	0.02	0.03	0.83	0.01	0.01	0.08	0.03	0.01	0.01
0.03	0.05	0.54	0.01	0.06	0.02	0.04	0.26	0.25	0.38	0.00	0.01	0.34	0.01	0.00	0.00
0.02	0.53	0.01	0.34	0.01	0.07	0.00	0.04	0.14	0.23	0.01	0.01	0.54	0.05	0.01	0.01
0.18	0.01	0.17	0.01	0.57	0.01	0.04	0.00	0.09	0.60	0.01	0.02	0.22	0.02	0.01	0.03
0.02	0.29	0.01	0.54	0.03	0.03	0.02	0.06	0.02	0.69	0.02	0.02	0.13	0.07	0.03	0.02
0.06	0.06	0.61	0.01	0.15	0.02	0.07	0.02	0.25	0.45	0.01	0.00	0.25	0.03	0.00	0.00

Subject 30 (8R1)

0.11	0.25	0.02	0.11	0.29	0.02	0.02	0.18	0.62	0.04	0.00	0.00	0.31	0.01	0.00	0.01
0.15	0.16	0.01	0.01	0.13	0.41	0.13	0.00	0.08	0.80	0.00	0.00	0.07	0.03	0.01	0.00
0.03	0.52	0.07	0.18	0.03	0.02	0.09	0.04	0.16	0.64	0.01	0.01	0.13	0.02	0.01	0.01
0.08	0.04	0.20	0.05	0.22	0.01	0.13	0.27	0.64	0.06	0.01	0.02	0.23	0.02	0.02	0.02
0.08	0.50	0.02	0.10	0.09	0.03	0.02	0.16	0.41	0.12	0.00	0.00	0.45	0.00	0.01	0.01
0.32	0.03	0.02	0.02	0.22	0.03	0.35	0.01	0.49	0.25	0.01	0.01	0.21	0.03	0.01	0.01
0.02	0.23	0.03	0.61	0.03	0.02	0.03	0.04	0.26	0.52	0.01	0.02	0.12	0.04	0.01	0.01
0.04	0.04	0.38	0.02	0.20	0.01	0.28	0.03	0.44	0.15	0.01	0.01	0.34	0.01	0.01	0.03

Subject 31 (4P4)

0.13	0.40	0.31	0.15	0.00	0.01	0.24	0.75
0.49	0.07	0.40	0.04	0.01	0.01	0.43	0.55
0.02	0.30	0.19	0.48	0.01	0.00	0.30	0.69
0.35	0.06	0.55	0.04	0.00	0.00	0.12	0.87

Subject 31 (4R1)

0.06	0.35	0.37	0.22	0.03	0.04	0.54	0.39
0.16	0.08	0.50	0.26	0.00	0.02	0.77	0.20
0.04	0.18	0.09	0.69	0.00	0.03	0.77	0.21
0.49	0.19	0.22	0.10	0.01	0.01	0.20	0.78

Subject 31 (8P3)

0.02	0.11	0.01	0.09	0.75	0.02	0.00	0.00	0.62	0.24	0.00	0.00	0.11	0.02	0.00	0.00
0.21	0.02	0.18	0.02	0.06	0.51	0.01	0.00	0.08	0.86	0.01	0.00	0.01	0.03	0.01	0.00
0.04	0.12	0.04	0.35	0.01	0.01	0.35	0.08	0.18	0.80	0.00	0.00	0.00	0.02	0.00	0.00
0.19	0.03	0.04	0.02	0.06	0.00	0.16	0.49	0.76	0.21	0.00	0.00	0.02	0.01	0.00	0.00
0.03	0.55	0.02	0.12	0.16	0.12	0.00	0.00	0.35	0.35	0.00	0.00	0.26	0.02	0.00	0.00
0.14	0.03	0.62	0.09	0.05	0.02	0.05	0.01	0.16	0.72	0.00	0.00	0.01	0.09	0.00	0.00
0.02	0.11	0.08	0.54	0.00	0.00	0.17	0.07	0.18	0.76	0.01	0.00	0.00	0.03	0.01	0.00
0.47	0.05	0.05	0.02	0.15	0.01	0.13	0.10	0.76	0.17	0.00	0.00	0.05	0.01	0.00	0.01

Subject 31 (8R2)

0.02	0.21	0.02	0.15	0.52	0.06	0.00	0.01	0.50	0.03	0.01	0.01	0.41	0.02	0.01	0.01
0.21	0.04	0.27	0.05	0.05	0.33	0.04	0.01	0.08	0.85	0.01	0.00	0.02	0.03	0.00	0.00
0.18	0.12	0.05	0.30	0.02	0.01	0.20	0.12	0.56	0.28	0.01	0.01	0.08	0.04	0.01	0.01
0.41	0.08	0.07	0.09	0.06	0.01	0.11	0.17	0.66	0.20	0.01	0.01	0.06	0.04	0.01	0.01
0.01	0.57	0.03	0.14	0.02	0.22	0.00	0.01	0.15	0.04	0.01	0.00	0.78	0.01	0.00	0.00
0.10	0.05	0.55	0.13	0.04	0.05	0.07	0.02	0.10	0.52	0.01	0.01	0.13	0.22	0.01	0.01
0.07	0.08	0.03	0.40	0.00	0.01	0.23	0.17	0.63	0.20	0.01	0.02	0.04	0.06	0.02	0.02
0.51	0.06	0.05	0.10	0.02	0.01	0.14	0.11	0.70	0.18	0.01	0.01	0.09	0.01	0.01	0.01

Subject 32 (4P4)

0.09	0.14	0.29	0.48	0.00	0.01	0.08	0.91
0.02	0.02	0.66	0.30	0.00	0.00	0.90	0.10
0.00	0.39	0.06	0.54	0.00	0.00	0.92	0.08
0.02	0.24	0.68	0.06	0.00	0.00	0.07	0.93

Subject 32 (4R1)

0.04	0.10	0.38	0.48	0.01	0.01	0.07	0.91
0.17	0.10	0.16	0.58	0.06	0.05	0.11	0.78
0.21	0.55	0.08	0.16	0.01	0.01	0.82	0.17
0.25	0.05	0.58	0.12	0.02	0.01	0.09	0.88

Subject 32 (8P3)

0.02	0.28	0.04	0.12	0.40	0.12	0.01	0.01	0.77	0.19	0.00	0.00	0.00	0.00	0.01	0.00
0.05	0.03	0.11	0.01	0.26	0.50	0.02	0.01	0.07	0.90	0.00	0.00	0.00	0.01	0.01	0.00
0.10	0.34	0.03	0.01	0.01	0.09	0.20	0.21	0.10	0.89	0.00	0.00	0.00	0.01	0.00	0.00
0.03	0.24	0.04	0.01	0.00	0.01	0.56	0.11	0.83	0.15	0.01	0.00	0.00	0.01	0.00	0.00
0.02	0.06	0.01	0.61	0.05	0.25	0.01	0.01	0.90	0.04	0.01	0.01	0.04	0.00	0.00	0.01
0.05	0.06	0.22	0.05	0.50	0.03	0.05	0.04	0.12	0.74	0.01	0.01	0.01	0.11	0.01	0.01
0.01	0.45	0.01	0.00	0.00	0.02	0.08	0.42	0.11	0.86	0.01	0.00	0.00	0.01	0.01	0.00
0.60	0.21	0.03	0.01	0.01	0.02	0.09	0.04	0.92	0.05	0.00	0.00	0.01	0.01	0.00	0.00

Subject 32 (8R2)

0.04	0.15	0.09	0.45	0.08	0.13	0.04	0.02	0.53	0.40	0.00	0.00	0.02	0.05	0.00	0.00
0.36	0.04	0.11	0.05	0.17	0.23	0.02	0.01	0.06	0.89	0.01	0.00	0.00	0.03	0.00	0.00
0.10	0.31	0.03	0.05	0.02	0.19	0.13	0.16	0.08	0.85	0.00	0.00	0.00	0.04	0.01	0.01
0.04	0.44	0.16	0.01	0.02	0.08	0.18	0.07	0.15	0.83	0.01	0.00	0.00	0.00	0.01	0.00
0.05	0.19	0.06	0.44	0.03	0.16	0.05	0.03	0.46	0.47	0.00	0.02	0.02	0.02	0.00	0.00
0.41	0.03	0.21	0.12	0.08	0.05	0.05	0.05	0.10	0.67	0.01	0.00	0.01	0.20	0.00	0.00
0.03	0.42	0.05	0.10	0.01	0.03	0.16	0.21	0.06	0.87	0.01	0.03	0.01	0.02	0.01	0.00
0.12	0.44	0.04	0.04	0.04	0.05	0.14	0.13	0.21	0.72	0.01	0.02	0.01	0.01	0.01	0.01

Subject 33 (4P3)

0.03	0.58	0.07	0.32	0.01	0.01	0.57	0.41
0.41	0.27	0.29	0.04	0.01	0.01	0.44	0.54
0.01	0.37	0.02	0.60	0.00	0.00	0.81	0.18
0.47	0.01	0.46	0.06	0.00	0.01	0.66	0.33

Subject 33 (4R2)

0.03	0.47	0.40	0.11	0.01	0.01	0.55	0.42
0.46	0.23	0.12	0.19	0.01	0.01	0.52	0.46
0.37	0.05	0.16	0.43	0.00	0.01	0.46	0.53
0.01	0.36	0.59	0.04	0.00	0.00	0.27	0.73

Subject 33 (8P4)

0.05	0.20	0.05	0.23	0.39	0.06	0.01	0.01	0.67	0.25	0.01	0.01	0.02	0.04	0.01	0.01
0.10	0.06	0.37	0.01	0.17	0.27	0.01	0.00	0.39	0.49	0.01	0.01	0.04	0.06	0.01	0.01
0.06	0.06	0.02	0.33	0.01	0.10	0.14	0.28	0.50	0.39	0.02	0.01	0.01	0.04	0.03	0.01
0.23	0.02	0.15	0.11	0.02	0.03	0.11	0.34	0.31	0.61	0.00	0.01	0.03	0.02	0.00	0.01
0.10	0.39	0.10	0.23	0.05	0.12	0.01	0.01	0.60	0.32	0.00	0.00	0.01	0.04	0.01	0.00
0.11	0.06	0.39	0.01	0.38	0.05	0.01	0.01	0.61	0.22	0.01	0.01	0.10	0.05	0.01	0.01
0.05	0.14	0.09	0.11	0.04	0.30	0.13	0.14	0.25	0.59	0.01	0.01	0.01	0.07	0.04	0.02
0.36	0.01	0.09	0.17	0.02	0.01	0.24	0.10	0.52	0.36	0.00	0.01	0.03	0.07	0.00	0.00

Subject 33 (8R1)

0.07	0.35	0.01	0.11	0.04	0.32	0.02	0.08	0.54	0.36	0.01	0.00	0.02	0.04	0.01	0.02
0.23	0.08	0.02	0.10	0.45	0.04	0.05	0.03	0.27	0.65	0.00	0.00	0.02	0.05	0.00	0.01
0.04	0.50	0.05	0.13	0.03	0.02	0.18	0.05	0.22	0.70	0.01	0.01	0.02	0.01	0.01	0.01
0.07	0.16	0.26	0.04	0.01	0.02	0.09	0.37	0.24	0.64	0.01	0.03	0.01	0.04	0.01	0.03
0.44	0.09	0.02	0.17	0.07	0.07	0.09	0.04	0.32	0.53	0.01	0.01	0.02	0.09	0.01	0.01
0.05	0.51	0.02	0.14	0.04	0.08	0.02	0.15	0.37	0.44	0.01	0.01	0.03	0.13	0.01	0.01
0.08	0.20	0.01	0.46	0.03	0.03	0.07	0.12	0.20	0.73	0.00	0.00	0.03	0.03	0.00	0.00
0.07	0.17	0.49	0.02	0.02	0.02	0.14	0.06	0.12	0.75	0.02	0.01	0.01	0.05	0.01	0.03

Subject 34 (4P3)

0.11	0.52	0.10	0.27	0.01	0.02	0.07	0.89
0.59	0.07	0.31	0.03	0.00	0.02	0.49	0.48
0.00	0.42	0.02	0.55	0.00	0.00	0.80	0.19
0.20	0.02	0.71	0.07	0.00	0.00	0.02	0.98

Subject 34 (4R2)

0.25	0.33	0.09	0.33	0.03	0.13	0.66	0.19
0.59	0.13	0.23	0.06	0.01	0.03	0.82	0.14
0.04	0.18	0.10	0.68	0.00	0.01	0.84	0.15
0.09	0.14	0.66	0.11	0.00	0.00	0.03	0.97

Subject 34 (8P4)

0.04	0.09	0.01	0.24	0.61	0.02	0.00	0.00	0.64	0.23	0.01	0.01	0.07	0.02	0.01	0.01
0.14	0.06	0.22	0.01	0.06	0.50	0.01	0.00	0.04	0.92	0.01	0.00	0.01	0.02	0.00	0.00
0.01	0.25	0.09	0.13	0.01	0.01	0.48	0.02	0.05	0.86	0.02	0.00	0.01	0.03	0.02	0.01
0.08	0.01	0.10	0.02	0.07	0.00	0.00	0.70	0.22	0.65	0.02	0.02	0.05	0.01	0.01	0.02
0.02	0.33	0.01	0.50	0.02	0.11	0.00	0.01	0.70	0.05	0.00	0.00	0.23	0.01	0.01	0.00
0.42	0.03	0.21	0.01	0.30	0.02	0.01	0.01	0.01	0.88	0.00	0.00	0.03	0.06	0.00	0.00
0.02	0.24	0.01	0.62	0.03	0.01	0.03	0.05	0.01	0.88	0.03	0.01	0.01	0.05	0.01	0.01
0.15	0.04	0.61	0.02	0.13	0.01	0.02	0.03	0.20	0.67	0.01	0.00	0.05	0.02	0.00	0.03

Subject 34 (8R1)

0.13	0.16	0.02	0.03	0.44	0.08	0.02	0.11	0.53	0.09	0.00	0.00	0.31	0.07	0.01	0.00
0.20	0.16	0.03	0.06	0.08	0.36	0.09	0.02	0.07	0.65	0.01	0.01	0.07	0.18	0.01	0.01
0.11	0.12	0.12	0.17	0.04	0.06	0.31	0.05	0.48	0.18	0.01	0.02	0.22	0.06	0.01	0.01
0.20	0.06	0.14	0.14	0.11	0.03	0.10	0.22	0.42	0.19	0.01	0.03	0.29	0.03	0.02	0.02
0.06	0.42	0.03	0.05	0.09	0.20	0.03	0.12	0.19	0.38	0.01	0.01	0.32	0.09	0.01	0.01
0.38	0.10	0.04	0.07	0.13	0.10	0.13	0.05	0.09	0.62	0.01	0.00	0.11	0.16	0.00	0.01
0.08	0.08	0.03	0.37	0.05	0.02	0.15	0.20	0.39	0.10	0.01	0.12	0.27	0.08	0.01	0.01
0.13	0.11	0.26	0.09	0.09	0.05	0.19	0.09	0.30	0.20	0.02	0.02	0.28	0.16	0.01	0.02