Psychology Dissertations   Department of Psychology

8-11-2020

Modifiability Of Strategy Use In Probabilistic Categorization By Rhesus Macaques (Macaca Mulatta) And Capuchin Monkeys (Cebus [Sapajus] Apella)

William Whitham

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

MODIFIABILITY OF STRATEGY USE IN PROBABILISTIC CATEGORIZATION

BY RHESUS MACAQUES (*MACACA MULATTA*)

AND CAPUCHIN MONKEYS (*CEBUS [SAPAJUS] APELLA*)

by

WILL WHITHAM

Under the Direction of David Washburn, PhD

ABSTRACT

Humans and nonhuman animals categorize the natural world, and their behaviors can reveal how they use the stimulus information they encounter in service of these categorizations. Rigorous psychological study of categorization has offered many insights into the processes of categorization and their relative strengths and weaknesses across species. Probabilistic categorization, in which the relationships among stimulus information and category membership that are observed by an individual are fundamentally probabilistic, presents unique challenges both to the categorizer and to the psychologist attempting to model their behavior. Challenges notwithstanding, probabilistic categorization is an exceptionally ecologically relevant problem to human and nonhuman animal cognition alike. This dissertation reports the effects of many manipulations of theoretical interest on computer-trained rhesus macaques' and

capuchin monkeys' inferred cognitive strategy use in a computerized version of a classic probabilistic categorization task. Experiment 1 probed cognitive strategy use across five variants of the same task in which the probability structure was constant, but the appearances and onscreen locations of cues and responses changed. Experiment 2 presented a series of manipulations of theoretical interest to the animals by changing the probability and reward structures of the task. Experiment 3 manipulated the stimuli of the task in ways motivated by findings across perceptual psychology literature. Experiment 4 extended the reward rate manipulations of Experiment 2 even further. Across four experiments, inferred strategy use was remarkably stable. Those animals that used cue-based strategies often returned to the same specific strategy experiment after experiment, as the cues, responses, probabilities, and contingencies changed around them. This finding is discussed in relation to questions of a real or functional ceiling on sophistication of strategy use, the robustness of cognitive individual differences in nonhuman primates, and future directions for comparative study of cognitive strategy use in probabilistic categorization.

MODIFIABILITY OF STRATEGY USE IN PROBABILISTIC CATEGORIZATION

BY RHESUS MACAQUES (*MACACA MULATTA*)

AND CAPUCHIN MONKEYS (*CEBUS [SAPAJUS] APELLA*)

by

WILL WHITHAM

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

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2020

MODIFIABILITY OF STRATEGY USE IN PROBABILISTIC CATEGORIZATION

BY RHESUS MACAQUES (*MACACA MULATTA*)

AND CAPUCHIN MONKEYS (*CEBUS [SAPAJUS] APELLA*)

by

WILL WHITHAM

Committee Chair: David Washburn

Committee: Sarah Barber

Michael Beran

Sarah Brosnan

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

August 2020

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

The relationships among minds, behaviors, and environments are fundamentally uncertain, and in many ways the cognitive competencies of mind can be understood as specific adaptations that serve to mediate this uncertainty (Glimcher, 2005). For example, the structure of many behavioral economic games is such that one or more players attempts to maximize the unpredictability of their responses as observed by other players (e.g., Rapoport & Budescu, 1992). Likewise, the structure of complex nervous systems are such that neuronal activity and neurotransmitter release are probabilistic events (Stevens, 2003). Cognitive scientists need not appeal to such evolutionarily ultimate or neuroscientifically molecular processes to find unpredictability in animal brains and behaviors, as the stimulus environments we interact with daily do not present us with transparent, deterministic, causal relationships.

A mind's transformation of the uncertainty of the environment into discrete representations, decisions, and behaviors is termed inference. Inference from probabilistic evidence is a key feature of human and nonhuman animal (hereafter: animal) life. A jury can only be uncertain as to a defendant's guilt or innocence, but they can make an informed inference via physical evidence, witness testimony, and reasonable doubt. The new restaurant on the corner may be delicious or nauseating, but inference from online reviews, a health inspector's report, and word of mouth can rapidly clarify the relative likelihood of food poisoning. A hungry monkey faces a similar dilemma, when a nearby fungus might be a poisonous toadstool or an edible mushroom. The relative probabilities of fitness advantage and disadvantage that might be won by consuming the fungus must be evaluated amidst considerable uncertainty and nontrivial stakes.

One way to describe this practice of inference is by using principles of probability theory, especially the rearrangement of conditional probabilities described by Bayes' rule (Etz & Vandekerckhove, 2018). Bayes' rule offers a straightforward way to synthesize probabilistic information into a form that is useful to a decision-maker: it is the probability that one event will occur over another, or that one hypothesis will be supported versus another. Interestingly, to cognitive psychologists, normative human behavior does not typically reflect these calculations in any straightforward way (Tversky & Kahneman, 1974). Moreover, there are other ways, beyond differing from Bayesian computation, in which inferences of humans and other species are not so sterile as probability theory might suggest. Adult humans (Washburn et al., 2005), human children (Beran et al., 2012), chimpanzees (*Pan troglodytes*; Beran et al., 2015), orangutans (genus *Pongo*; Marsh & MacDonald, 2012), rhesus macaques (*Macaca mulatta*; Tu et al., 2015), Guinea baboons (*Papio papio*; Malassis et al., 2015), a dolphin (*Tursiops truncatus*; Smith, 2010), and rats (*Rattus norvegicus*; Foote & Crystal, 2007) are among the populations that have demonstrated the ability to behave in accordance with recognition of their own uncertainty.

At least for humans, this relationship with uncertainty is fraught. Objectively, there are few situations in which it is adaptive to be underinformed about the contingencies of one's environment. But subjectively, the story is not so clear. On one hand, live events like sports are uniquely profitable entertainment (with 89 of the 100 most watched broadcasts in the United States in 2018 related to live sports), in no small part because their contents are fundamentally uncertain (Crupi, 2019). On the other hand, and less trivially, individuals at risk of high trait anxiety or clinically significant anxiety disorders are intolerant of the subjective feeling of not knowing that comes with such uncertainty

(Carleton et al., 2007). These examples are expository only, and scarcely scratch the surface of how probability and psychological processes intersect. Uncertainty is, to a large or small degree, explicitly or implicitly, a component of any psychological process concerned with information (Garner, 1962).

This dissertation will describe the strategies that species use when identifying objects (categorizing a stimulus into one or many unique psychological groups) and coordinating behavior on the basis of such judgments (engaging in decision-making by using the external and internal stimulus information at its disposal to guide behavior) in service of probabilistic inference. I will use this introductory chapter to catalog previous research on the cognitions and behaviors that are characteristic of categorization and decision-making in conditions of uncertainty. Next, I will introduce one specific methodology for bringing into the psychology laboratory the types of probabilistic categorization and decision-making that organisms face in their environments, discuss what this methodology has already contributed to our understanding of such constructs, and identify its limitations. In the second chapter, I will present a role for comparative psychology in answering a number of outstanding questions about probabilistic categorization, alongside discussion of species-specific considerations that suggest a match between mechanisms of probabilistic categorization and the cognitive competencies of New World and Old World monkeys. Further chapters will describe a set of experiments that probed the degree to which these characteristic cognitions and behaviors are specific to the stimulus properties and/or embedded probabilities of previously published experiments, the results of these experiments, and implications for the comparative study of how cognition intersects with chance.

1.1 Probabilistic Categorization

Very high categorization performance must be required of humans and nonhuman animals in their natural habitats, where dangerous predators cannot be mistaken as benign and new chances to elect political officials in the United States come only every two to six years (Ashby & Ennis, 2006; Kruschke, 2005; Smith, 2014). Psychologists often present relationships between psychological inputs and their observable outputs as if they were deterministic: an organism that perfectly used the information at its disposal would unerringly predict the toxicity of a mushroom or the danger posed by a large animal's silhouette (Ashby & Maddox, 2005). Indeed, three of the four major classes of categorization experiments identified by Ashby and Ell (2001) model relationships of this kind. In the first, rule-based tasks, including the influential Wisconsin Card Sort Task, an explicit, verbally declarable stimulus dimensions or some conjunction of such stimulus dimensions are mapped to categories (Eling et al., 2008). In the second, information-integration tasks, like many variants of the sinewave grating task of Ashby and colleagues (1998), stimulus dimensions are mapped to categories in a way that cannot be explicitly declared (e.g., Qadri et al., 2019; Smith et al., 2012). In the third, prototype-distortion tasks, like the "A, not A" tasks of Posner & Keele (1968, 1970), stimulus dimensions are mapped to categories based on the degree of similarity a prospective category member shares with an unseen, prototypical category member. It is only the fourth major class of categorization experiment, unstructured tasks, that maps items to categories arbitrarily without regard to any stimulus dimension (Little et al., 2006).

In each of these designs, with few exceptions (the subset of information-integration tasks that does not provide deterministic mappings between stimulus properties and category membership will be discussed later in this chapter), any object

with a given set of values on stimulus dimensions of interest is mapped, with absolute certainty, to a specific category member or response. From these experiments, a psychologist can evaluate the degree to which an individual or population evidenced learning of these deterministic mappings of stimulus information to a category. Where perfect utilization of the available information is not achieved, a mélange of exploratory behaviors, novelty biases, and motoric considerations is often assumed to be to blame.

Any deterministic conception of categorization understates the degree to which real relationships of interest between psychological inputs and observable outputs are probabilistic (Kruschke & Johansen, 1999). Most charitably, deterministic categorization tasks are a simplification of how a human or animal's categorization mechanisms work in a single, specific context. For example, a child is either tall enough to ride a roller coaster, or not tall enough to ride it. A single, deterministic dimension of height guides categorization, and many other dimensions (e.g., clothing, hairstyle, freckle count) are irrelevant. Decades of categorization research might inform whether such a categorization is made by explicit, rule-based reasoning, or implicit, procedural learning, or other orthodoxy. Both this depiction of real-life categorization and its experimental models face limits to their validity and implementation. Such a perspective can only be agnostic about how learners manage less deterministic cue information across contexts. This is a serious shirking of duty, because demand for the ability to combine many probabilistically noisy, conflicting cues is high. Many other categorizations about our prospective coaster rider would require an evaluation of many dimensions that combine far more fitfully. A school administrator would, one hopes, categorize the child into an advanced, remedial, or on-level classroom based on an exceedingly complex mixture of teacher impressions, classroom behavior, and scores from math, reading, and science

assessments. None of these cues are perfect, and the administrator can only hope to make an informed probabilistic categorization about the student's true category membership.

In addition to these questions of the ecological validity of classic categorization tasks, other domains of the social sciences also caution against simplifications of natural category structures. To economists, there are insidious consequences to positing any stable, deterministic relationship between an observable stimulus dimension and category membership (Goodhart, 1984). To adopt a popular example: in academic settings there is probably some relationship between a citation metric and the quality of scholarship of an academic (i.e., a citation metric dimension can be used to categorize academics as high quality or low quality scholars). However, as soon as the citation metric becomes conventionally used as a deterministic predictor of scholarly quality, it ceases to be one, because selection pressure for both academic and categorizer is to maximize on that metric instead of maximizing on scholarly quality. In education policy a similar caution is made regarding standardized achievement testing. Standardized tests may deterministically predict scholastic success in normal teaching and testing conditions, but not when school and student resources are subverted to maximize standardized test scores (Sidorkin, 2016). Even in a relatively sterile social science like applied statistics, the concept of overfitting codifies this idea that a well-validated relationship observed in one context cannot be expected to be perfectly maintained in a new context (Babyak, 2004).

Psychologists use weather prediction tasks to model naturalistic, probabilistic category structures by structuring the cue information available to an organism to be probabilistically related to the value of the responses an organism can make (Gluck & Bower, 1988). In this way, the uncertainty that permeates real life categorization contexts

can be accounted for with exacting specificity, and titrated continuously. In most classic weather prediction designs, each trial presents a participant with one to three imperfect cues (often stylized as medical symptoms or tarot cards; Knowlton et al., 1994). The participant is asked to choose which of two outcomes will be scored as correct in that trial. Perfect performance is impossible in such a task, because even the greatest weight of cue information a participant might receive will still underspecify the correct outcome. Instead, the participant, like the school administrator, must work within their uncertainty to make the best decision possible with the information that they have.

1.2   Probabilistic Decision-Making

The inferences made during these probabilistic categorizations are observable in how individuals behave in naturally-occurring and experimentally-derived decision-making contexts. Humans and animals are often found to deviate from probability theories' recommendations given a set of probabilistic information, even when the task is well-learned (Nickerson, 2004). To chronicle these deviations, cognitive psychologists have cultivated an impressive bestiary of the many heuristics and biases that blight civil, rational minds (e.g., Kahneman & Tversky, 1984, Tversky & Kahneman, 1974; Tversky & Kahneman, 2014). For example, humans and pigeons neglect the prior probability of an estimable occurrence and conversely assign undue decisional weight to any predictive information, termed base-rate neglect (Bar-Hillel, 1980; Fantino et al., 2005; Tversky & Kahneman, 1982; Zentall & Clement, 2002). A positive test for a rare disease might be understood as highly likely to reflect that the disease is present, even though the low prior probability of the presence of the disease would instead suggest that a false alarm by the test is much more likely. In a related example, this cue information is weighted so greatly

that human evaluations of the probability space are not internally, logically consistent (Tversky & Kahneman, 1983). This conjunction fallacy is frequently illustrated via a short vignette about a young woman with a history of social activism (sometimes called the Linda problem). Human participants often evaluate that she is more likely to be a bank teller who is active in feminist causes than that she is "merely" a bank teller. The young woman's activism is so strongly weighted by human observers that they propose the logical impossibility that the conjunction of two states could be greater than one of the states in isolation. Even at the earliest stages of probabilistic decision-making, humans are not particularly successful at identifying where probabilistic information is and where it is not. Humans routinely report that probabilistic relationships exist among sequences or stimuli when the process that generated them was actually paradigmatically random (Zhao et al., 2014).

There is a sense in which the size of this apparent gulf between recommendations derived from Bayesian computation of probabilistic information and observed human behavior has been exaggerated. The way in which such probabilistic decision-making problems are typically framed, with percentile probabilities disseminated in words, is ideal for Bayesian computation. This is highly different from how humans and animals would sample probabilistic information from their natural environments (Gigerenzer & Hoffrage, 1995). When experimenters present human participants with mechanically identical probabilistic decision-making problems in a representational format more akin to how evolution shaped minds to intuit probabilistic information, many of the irrational biases can be substantially or entirely muted (Gigerenzer, 1991; Gigerenzer et al., 1991).

Accounts of human probabilistic decision-making behavior that seek to coax biases from participants and those that forgive these biases as part of an entropy or effort

reduction shorthand compete to provide superior accounts for probabilistic decision-making behavior (Gigerenzer & Goldstein, 1996; Kahneman & Tversky, 1996; Shah & Oppenheimer, 2008). Similarly, conceptions of probability theory that draw from perfect knowledge of the problem space compete with subjective, observed event frequency to describe human interactions with probability theory best (Gigerenzer, 1991; Gigerenzer, 1994). Post-hoc labeling of characteristic responses of humans in highly specific, highly verbal probabilistic decision-making contexts is common. In contrast, process-based or mechanistic accounts of cognition and behavior are relatively uncommon (Gigerenzer, 1996). Stanovich and West (2000), and more popularly Kahneman (Kahneman, 2011; Kahneman & Frederick, 2002), understand the cognitive processes that produce these characteristic responses as derived from two systems: one quick, effortless, intuitive, and automatic and the other deliberate, effortful, and systematic. This clustering represents an improvement on merely labeling biases, and is intellectually fruitful in that the individual and correlational psychologies of how individuals may deploy these systems can be explored (Cronbach, 1957). But even this approach reduces to a fairly trivial labeling of inferred cognitive processing: "fast" or "slow", with some inconsistently correlated features (Melnikoff & Bargh, 2018).

The restricted nature of these dual-process accounts distracts from the substantial challenge that cognitive psychologists face in seeking to describe decision-making processes molecularly, via their inputs, outputs, processes, neuronal substrates, and proximate and ultimate causes. Kruschke and Johansen (1999) took up this challenge with their Rapid Attention SHifts N Learning (RASHNL) model for probabilistic cognition in which mathematical relationships among inputs from the environment, their representation by an individual, and behavioral outputs can be specified exactly. More

specifically, the RASHNL model built on previous connectionist models of human cognition (e.g., Gluck & Bower, 1988) to describe a probabilistic cognition that changes continuously and dynamically across experience. More recently Trueblood, Busemeyer, Pothos and their collaborators have advanced models of human probabilistic decision-making that trade connectionist principles for those borrowed from particle physics and, in doing so, make explicit the characteristic indeterminacy of human brains and behaviors (Pothos & Busemeyer, 2009; Trueblood & Busemeyer, 2011; Trueblood et al., 2017). This computationally intensive account, termed quantum cognition, has furnished formal mathematical descriptions of how many of the characteristic responses humans make in probabilistic situations may arise. These formal models, and others like them, provide accounts of human probabilistic decision-making at unprecedented levels of detail. Interestingly, the degree to which these models impact scientific discourse would seem to be less a question of their accuracy (they provide a good fit to data, both real and simulated) and more of the degree to which their core ideas can be disseminated to and internalized by cognitive and applied psychologists who have already adopted the fundamentally intuitive, but scientifically fraught dual-process typology.

Probabilistic decision-making theories thus appear to fall along a continuum. At one extreme, researchers figuratively plant a tiny flag on the response that they will consider rational and hundreds or thousands of additional flags on every other point (each with its own unique identifier and publication). At the other extreme, researchers seek to recreate, via pure mathematics, artificial neural networks, and/or principles of particle physics, the computations undertaken by the brain-shaped black box that transforms psychological inputs into behavioral outputs. An approach that appreciates each of these perspectives may be uniquely valuable. Rigorously and mathematically

defining experimentally derived inputs and predicted outputs is a requirement of scientific progress, and a prerequisite of most statistical analyses of merit. That said, in many contexts people and their cognitive processes really are clustered, such that labels can properly represent psychological relationships (Bartlema et al., 2014). An epistemological midpoint might borrow features from each extreme to make predictions that are not only accurate, but are also transparently interpretable, accessible to other scientists, and portable to new areas of cognitive inquiry.

Lee and colleagues have done much to advance such a perspective by using principles of statistical model selection to identify which account, among competing theories of involved cognitive processes, best accords with behavioral data of probabilistic decision-making tasks. For example, Lee and Newell (2011) applied this strategy to a classic decision-making problem in which the size of German cities is estimated. They used behavioral and simulated data with competing formal models arranged using the principles of the "Take-the-Best" heuristic and the influential weighted-additive rule. In this way classic ideas about normative human decision-making behavior, complete with heuristics and biases, are ported intact for principled, statistical evaluation alongside competing accounts. Critically, these evaluations may inform what information was utilized by participants and how participants' cognitive competencies utilized this information, alongside estimation of which rule or process provided the best fit. In the case of Lee and Newell (2011), this approach facilitated novel inferences about participant search strategies and stopping rules. In other cases, this approach has served to discriminate among categorical and continuous differences in human cognitive processes in a Multidimensional Scaling design (Bartlema et al., 2014) and demonstrate participant sensitivity to probabilistically weighted evidence in a decision-making task (Lee, 2016).

1.3 Strategy Analyses in Weather Prediction

A key feature of weather prediction tasks is that they can be completed at high rates of success by any one of many different response strategies (Gluck et al., 2002; Meeter et al., 2006). For example, in the set of relationships between cues and response probabilities that is often used by weather prediction researchers, a simple Take-the-Best-style heuristic is nearly optimal (Ashby & Maddox, 2005; Todd & Gigerenzer, 1999). For this reason, the more meaningful inference about participant performance in a weather prediction task is not merely the proportion at which participants chose the correct outcome, but rather the degree to which participants' responses are consistent with use of one strategy or another. Because equivalent task performance by individual participants does not require equivalent response mechanisms, response strategy use in the task can vary wildly within the realm of roughly-equivalent rates of correct responding.

These inferences are arrived at via strategy analyses techniques that compare many theoretically-interesting models that make different, explicit predictions about participant behavior in the task. Gluck, Shohamy, and Myers (2002) inaugurated the use of such techniques in weather prediction designs. In their task, participants were asked to predict whether 'sun' or 'rain' was a likelier outcome on the basis of a set of 1-3 arbitrary, probabilistic cues. To infer participant strategy use, they used a simple least mean squared error measure in which a normalized count of the expected number of sun responses predicted by each strategy minus the observed number of sun responses was computed for each candidate model. Candidate models fell into three major designations. Multi-Cue strategies assumed the optimal response (i.e., response that would be correct

most often given the available cue information) to every cue combination, as if the participant perfectly integrated the cue information they received. Single-cue strategies assumed that responses were made based on the presence or absence of one cue. This type of strategy assumed fundamentally rule-based responses (e.g., if cue1 is present, choose response 1. if cue1 is not present, choose response 2). Finally, Singleton strategies assumed that the participant made the optimal response any time only one cue was presented, but failed to successfully integrate the information yielded by multiple cues. The initial strategy analysis of Gluck and colleagues described most participants as best fit by Singleton strategies, with few participants best fit by Single-Cue strategies and even fewer best fit by optimal, Multi-Cue strategies.

Substantial elaborations on the strategy analysis of Gluck and colleagues were made by Meeter, Myers, Shohamy, Hopkins, and Gluck (2006). Meeter and colleagues translated the analysis to one that could be guided by standardized principles of model selection and fit measures like Akaike's Information Criterion (Akaike, 1974), rather than the opaquer inference yielded by the least mean squared error measure. Model parameters representing participants' response rates to a predicted outcome were fixed at a proportion less than one, modeling the assumption that even at mature responding participants don't respond in perfect accordance with their strategy, and then simulated response data were used to guide parameter assignment in the model. With their refined modeling procedures, Meeter and colleagues tested a greater variety of plausible response strategies. A class of Intermediate strategies was considered, including strategies that integrate information across multiple cues so long as the cues provide converging evidence. Perhaps most importantly, Meeter and colleagues included a Random strategy that acted as a buffer against false alarms of strategy use when none was really present.

These strategy analyses are part of a larger movement in the psychological sciences towards categorical inferences about participant brains and behaviors. As previously discussed, Lee (2016) expanded upon an earlier analysis of Hilbig and Moshagen (2014) to make inferences about probable models for participant decision-making. In cognitive neuroscience, mode networks like the default mode network (Greicius et al., 2003) or, more recently, quasi-periodic patterns (Abbas et al., 2019), are similar conceits for describing consistent patterns among isolable observations. In each of these analyses, as in that of Meeter and colleagues, many correlated observations, be they consistent patterns of activation and inhibition of independent neural networks or consistently strategic responses made to specific sets of cue information, were synthesized using modern model selection tools to offer novel insights into the phenomena of interest. In each, the phenomena of interest were better understood by describing where they were consistent with an unobservable, common cause. Indeed, this exact manner of inference is precisely the one sought by the scientific method: which of the available candidate sets of predictions (e.g., models, hypotheses, theories) best predicts observed data?

In weather prediction applications, these analyses reveal a characteristic pattern of responses made by neurotypical adult humans. Responses are initially unstructured and best fit by the Random strategy, as in the absence of any experience with the task any relationship between cues and outcomes cannot be known (there is no logical relationship between cues and outcomes to which a participant might reason). Random responses typically give way to Single-Cue or Singleton responses based on single elements of the available cue information. This trend mirrors inferences from absolute categorization tasks in which rule-based responding is dominant over other types of responding such that if an explicit rule is made available to participants, they will invariably use it (Church

et al., 2018, but see Duncan et al., 2018 for demonstration of a preference for multicue strategies over single-cue responding in an atypical weather prediction task). In many weather prediction designs, the majority of participants do not use more complex strategies than these rule-based, Take-the-Best strategies. These results are not surprising, with analogs in the human decision-making literature (Lee & Newell, 2011; Todd & Gigerenzer, 2000) and deterministic categorization literature across species (Smith et al., 2012). A minority of adult, neurotypical humans have sometimes used more optimal strategies that include Multi-Cue integration of the cue information at late stages of the task (Gluck et al., 2002; Kemény & Lukács, 2010; Meeter et al., 2006; Newell et al., 2007).

1.4   Cognitive Neuroscience of Weather Prediction

Of greater theoretical significance are the inferences made about participant populations other than neurotypical adults. Theories of language impairment, like the procedural deficit hypothesis, posit that cognitive impairments to processing of probabilistic information are the root cause of diagnoses like specific language impairment or developmental dyslexia (Ullman & Pierpont, 2005; West et al., 2018). Theorists have tested such theories by comparing weather prediction performances of groups with and without these diagnoses. Finn and colleagues (2016) reported that 10-year-old children made optimal responses in a weather prediction task at the same rate as adults, but did not report the results of a strategy analysis comparing the two populations. This result contrasts sharply with measures of declarative cognition of the two groups, with adults demonstrably superior in working memory and long-term memory measures. Gabay and colleagues (2015) reported that adults with developmental

dyslexia demonstrated impaired learning on weather prediction tasks relative to matched controls but were not significantly different from controls in their frequency of use of Single-cue, Singleton, or Multicue strategies. Kemény and Lukács (2010) compared language-impaired children (age M = 11.3 years, SD = 1.3), with age-matched typically-developing children and adults. Language-impaired children performed at chance levels across increasing experience in the task, and 11 out of 13 participating language-impaired children were not well fit by any cue-based strategy. In contrast, both typically-developing children and adults performed at above-chance levels within the first 50 trials and adopted increasingly complex strategies over the course of testing.

Use of weather prediction tasks across participant populations have also been important to the formulation and development of dual-process theories of learning and memory that posit dissociable memory systems that vary in, among other things, the degree to which the participant has conscious access to the inputs, outputs, and processes of the system (Ashby et al., 1998). Knowlton et al. (1994) innovated this application of weather prediction tasks in their exploration of the learning and memory abilities of eight amnesic patients (two Korsakoff's diagnoses, one bilateral medial thalamic infarction diagnosis, four with known or strongly evidenced hippocampal damage diagnoses, and one idiopathic case). The amnesic participants performed as well as matched, neurotypical control participants early in learning, but were markedly worse than controls later in learning. The authors argued that because early weather prediction responding is better accomplished without declarative memory (the probabilistic nature of the task means that relying on discrete memory of what happened in a previous trial may be misleading), amnesic participants' performance was not impaired. It was only later in learning, after many opportunities to cultivate declarative knowledge about the

probabilistic structure of the task, that control participants performed better than participants with amnesia.

Knowlton et al. (1996) offered additional support to this argument. They presented a weather prediction task to 7 participants with medial-temporal lobe amnesia and 5 participants with diencephalic amnesia, to 20 participants with Parkinson's disorder (damage to striatal regions and others), and to age-matched control participants. As an additional measure, they also asked participants verbally to report what they had learned from the task. As in Knowlton et al. (1994), participants with amnesia were as effective at the weather prediction task as controls during early task performance. However, these participants could not answer basic multiple-choice questions about their testing experience. In contrast, participants with Parkinson's disorder were not effective at the core weather prediction task, but they were as effective as controls in reporting basic facts about the testing session. This double-dissociation has been considered prime evidence for the existence of two separable neurocognitive systems: a conscious, explicit, declarative system underpinned by the hippocampus and mediotemporal regions (and expressed via frontal regions) and an unconscious, implicit, procedural system underpinned by the basal ganglia (and expressed via motor regions).

Neuronal encoding of the probabilisitic information associated with visual cues in weather prediction tasks is thought to occur via the plasticity of synapses in the lateral intraparietal cortex (Soltani & Wang, 2010; Yang & Shadlen, 2007). This region is implicated in a variety of primate psychological processes, including memory for and attention to visual stimuli (Sereno & Maunsell, 1998) and reinforcement learning in the more probabilistic games of behavioral economics (Seo, Barraclough, & Lee, 2009). Unlike the structured deviations from normative probability theory that humans

frequently evidence in their behavior in the task, extracellular recordings of the synaptic activity of the lateral intraparietal cortex have been found to very faithfully represent the probability structure of a weather prediction task (Yang & Shadlen, 2007). The nature of this synaptic activity can even explain certain systematic biases that are typically exhibited in specific probabilistic decision-making contexts, like base rate neglect (Soltani & Wang, 2010). Given this relatively precise encoding of probabilistic information at the neuronal level, it is interesting that humans who complete weather prediction tasks are reliably highly variable in both their task behavior and in their explicit reports about their cognitive strategy use in the task.

These data are from a small number of neurons, from a small number of nonhuman primates. As such, the generalizability of these neuronal readings is extremely restricted. These reports perhaps suggest that not only is the probability structure of the task identical across variably behaving animals, but also the encoding of this probability structure is equivalent across variably behaving animals. In that case the variability must arise somewhat further down the [neuro]cognitive stream, perhaps at one or both of the regions identified by the separable systems account. The more rule-like processes described by the separable systems account (and modeled by weather prediction strategy analyses) take on a special role in culling the variability in psychology of probabilistic inference that remains unexplained.

1.5   Interim Summary of Weather Prediction Research

Critically, the reviewed weather prediction studies (1) report behavioral dissociations that reveal differential engagement of multiple memory systems over the course of the task, and (2) demonstrate a marked link between the requirements of the

task and clinical language and memory deficits. Such a link would have substantial neurocognitive and clinical applications (Krishnan et al., 2016). However, other lines of research make such a conclusion more fraught. Newell, Lagnado, and Shanks (2007) presented multiple experiments that tested the predictions of dual-process theories in weather prediction tasks. If classic weather prediction performance is the product of implicit, procedural learning mechanisms, not explicit, declarative mechanisms, then the inclusion of a concurrent task that taps into more declarative mechanisms should not meaningfully disrupt weather prediction performance. Their Experiment 1 demonstrated that inclusion of a verbal Stroop-like concurrent task disrupted learning of the weather prediction task and changed the distribution of strategies used by participants. Dual-process descriptions of weather prediction tasks also suggest that the task is completed by procedural learning mechanisms because reward feedback is administered to participant responses, and that more declarative mechanisms would be involved when reward feedback is not administered in this way. Newell and colleagues' Experiment 2 tested groups of participants on weather prediction tasks either in which feedback was administered after responses or in which participants passively viewed cues and their associations to outcomes and reported that participant performance and strategy use did not significantly differ between groups. Taken together, the invariability of classic weather prediction tasks to changes in feedback contingencies and primary reinforcers suggests no special role for uniquely procedural learning mechanisms in the task. In contrast, declarative insight into what was learned in the task was strong throughout. No manipulation in either experiment affected the accuracy of participants' verbal estimations of the relationships between cues and outcomes.

Classic conceptions of weather prediction are also critiqued on grounds of philosophy of science (e.g., Keren & Schul, 2009). Newell, Dunn, and Kalish (2011) argued that the dual-systems conception of weather prediction performance is presented in such a way that it cannot be readily falsified. That the performance of amnesic participants is equivalent to that of control participants early in learning but not as good late in learning is regarded as key evidence for both the mostly-procedural nature of weather prediction tasks and the existence of dissociable memory systems (Knowlton et al., 1994). However, precise behavioral indicators for when a participant might be using procedural versus declarative mechanisms to complete the weather prediction task are conspicuously absent. Instead, the structure of the argument from dual-systems theorists seems to be one that fallaciously affirms the consequent: a theory that permits fluid switching between procedural and declarative systems could accommodate the behavioral data from amnesic and control participants, so the theory is supported. In such a case the fluid switching offers its own inferential minefield, since the dual-systems are malleable enough to accommodate nearly any behavioral data (Newell et al., 2011; Palmeri & Flanery, 1999). Ashby and Maddox (2005) offered the more moderate criticism of classic weather prediction designs in which what they call "explicit memorization" might be so dominant a response strategy that much of the nuance of weather prediction tasks and their analyses is ultimately wasted, and the centrality of weather prediction to dual-process theory-crafting is overstated.

An alternative conception of weather prediction tasks and data (implied by the earlier claim of affirming the consequent) is a family of theories and analyses that have described weather prediction learning as happening gradually (Speekenbrink et al., 2008). In these conceptions, any observed differences in weather prediction

performances by different participant populations are described as the product of differences in learning rates, not discrete, rule-like shifts. In a way that makes explicit the distinction between processes of categorization and decision-making response behavior outlined at the outset of this manuscript, Speekenbrink et al. (2008) utilized an analysis strategy that divorces modeling of learning from modeling of responding (Cooksey, 1996). This lens model yielded similar inferences to those of previous weather prediction analyses; that is, that amnesic participants are very nearly as good as controls at the task, they learn much about specifics of the task, and suffer an episodic memory deficit about other aspects of the experiment. Their analyses differ in that they describe nuances of participant performance via quantitative differences in learning rate between the groups rather than the types of modular, systematic, qualitative differences described by dual-process theorists.

Notably, at least in the case of weather prediction tasks, there is not disagreement about whether the competing models fit to data (they do). Indeed, the opposed theorists and their opposed models make largely the same predictions about each dataset (Meeter et al., 2008). As in other realms of cognitive psychology in which complementary models are wielded by theoretically opposed parties, relatively semantic explanatory virtues are leveraged against one another in relatively subjective ways (Barsalou, 1992; Fidler, 2018; Smith et al., 2016; Wills & Pothos, 2012). A strategy analysis perspective assumes discrete shifts in behavior and describes them using language that implies dual influences of procedural learning and cognitive control (i.e., participants [choose to] adopt one [cognitive] strategy or another to complete the task). The lens model perspective assumes gradual updating of learned associations between cues and outcomes, and gradual

updating of the utilization of those learned associations. Each approach has its merits, but a strategy analysis perspective boasts the following explanatory virtues (Fidler, 2018):

- Analogy between the inferences of strategy analyses and those of other disciplines are many, and were reviewed above (e.g., comparable models from the judgment and decision-making psychological literatures, neuroscience mode networks, etc.).

- The assumptions of strategy analyses are testable in a more robust manner - strategies posit specific behaviors to specific sets of cues and test whether responses conform to that specification at an arbitrarily frequent rate (e.g., thousands of times in Whitham & Washburn, 2020).

- Ultimately a strategy analysis perspective is more intellectually fruitful. The perspective yields a dissertation of theoretically interesting confirmatory and exploratory hypotheses that can enrich our understanding of an important set of cognitions, whether or not the strategy analysis perspective itself withstands the tests of time.

[For reasons of aesthetic sensibility, the remainder of this manuscript will use strategy use to connote response strategy use inferred by strategy analysis. I will imply discrete shifts from one strategy to another, pursuant to published research that makes a well-evidenced case that weather predictions tasks are better understood as mostly declarative and rule-based in nature (Newell et al., 2007). Note that despite this choice, these conclusions are analytically-derived inferences about discrete behaviors (not the direct observation of mind the concise language may imply), and but one among many interesting ways of exploring mind and behavior via weather prediction tasks.]

Divorced from any role as standard-bearer in dual-process theorizing, weather prediction tasks are in many ways an ideal model of human and animal cognition and

behavior in response to complex, probabilistic information. Unlike nearly every experimental task of probabilistic decision-making research programs, the task need not be a verbal problem in which each word is subject to etymological scrutiny (Hertwig & Gigerenzer, 1999) and changing just a few words changes the nature of the problem (Dulany & Hilton, 1991). Instead, probabilistic information can be disseminated across experience in a way more akin to how natural frequencies are sampled by many species (Gallistel & Gibbon, 2000). Indeed, how this probabilistic information is disseminated is in many ways as flexible as the experimenter is: cues need not be restricted to one modality or another, the number of cues can be as high or as low as needed (e.g., Soltani & Wang, 2010; Yang & Shadlen, 2007), cue information can be weighted continuously, and so forth. Common analyses of weather prediction, too, facilitate flexible matching of statistics to experimental design. So long as a stable probability structure for the task can be crafted, and theoretically interesting hypotheses for how individuals solve the task can be specified, weather prediction tasks represent a well of potential for the experimental study of probabilistic cognition in humans and animals.

2 CORRELATIONAL, EXPERIMENTAL, AND COMPARATIVE PSYCHOLOGY OF WEATHER PREDICTION

Previous research programs have provided rich descriptions of different individual humans' brains and behaviors in a weather prediction task. What these studies cannot answer is the degree to which the strategy use is a trait of an individual that is stable across time, context, task manipulation, or other intraindividual factors. In the language of Cronbach (1957), much research has been done on the correlational psychology of weather prediction tasks and relatively little on their experimental psychology (Conway & Kovacs, 2013). Indeed, many designs are fundamentally correlational - how do different people with different neuronal structures, or different ages, or different diagnoses, solve the weather prediction task (see Table 2.1) - with titles that follow a discrete Mad Lib format: "How does [population] solve the weather prediction task" (Gluck et al., 2002; Meeter, et al., 2008; Whitham & Washburn, 2020).

The strategy analysis perspective often adopted by weather prediction researchers offers promise for an experimental psychology component in the form of the longitudinal tracking of participant strategy use. Whereas comparisons of the strategy-use trajectories of different participants, from different participant populations, is fundamentally a correlational psychology pursuit, that a trajectory of individual inferences - that is, a set of inferences made about cognition as it is manifest across experience - was charted in the first place is a nontrivial step toward understanding the intraindividual factors that guide probabilistic categorization.

Comparative psychology has much to offer to the study of outstanding questions about probabilistic cognition. Many of the characteristic behaviors that humans evidence in probabilistic decision-making tasks are present in animal species as well (Blanchard et

al., 2014; Klein et al., 2013; Mazur & Kahlbaugh, 2012; Watzek et al., 2018). Most species have not been demonstrated to monitor their own uncertainty, and are likely to engage processes of categorization and decision-making without actually evaluating the amount of uncertainty in their environment. That these species solve probabilistic categorization and decision-making tasks is tautology – they make decisions in their native environments, and their native environments are rife with probabilistic cues. But despite some animals' demonstration of flexible, strategic leveraging of rule-based responses in other categorization tasks, they have rarely been previously studied in such terms (Smith et al., 2012). Furthermore, comparative psychology can offer a greater wealth of data from individuals within and across task manipulations, great potential for correlation of strategic behavior to other cognitive measures, and a window into the evolutionary trajectory of this fundamental cognitive competency.

Whitham and Washburn (2020) explored what patterns of responses nonhuman primates would spontaneously use to solve a classic weather prediction task. These monkeys were not constrained to ~200 trials of the task in an hour or less, like most weather prediction designs with humans, and therefore had their strategy use assessed at unprecedented levels of experience in the task. The authors adapted the weather prediction task of Gluck and colleagues (2002) for use with a computerized test system for nonhuman primates (the Language Research Center - Computerized Test System; Evans et al., 2008; Richardson et al., 1990). Cue information was disseminated via clipart images on a nearby computer screen, instead of via manipulable tarot card cues. Monkeys used a joystick apparatus to make a response towards one of two additional clipart images that acted as the binary response options 'Sunny' and 'Rainy'. The authors maintained all probability and relative frequency information from Gluck, Shohamy, and Myers, with

the key difference that the overall exposure to the task was far greater (with many monkeys completing thousands of trials).

These monkeys followed a distinctly different trajectory of strategy use in probabilistic categorization tasks than do typical adult human participant populations (Whitham & Washburn, 2020). Whereas humans often transition smoothly from using responses that are not strategic to responses that depend on only a single item of cue information to (in some cases) responses that flexibly integrate multiple sources of cue information, many monkeys used no cue information at all in favor of other vaguely strategic responding (e.g., use of only one response per session). Many other monkeys were eventually best fit by a Single-Cue strategy that depended on the presence or absence of the same, single, highly-predictive cue, with some monkeys shifting to this strategy only after many hundreds of trials. No monkeys were consistently best fit by Singleton strategies (that respond optimally to any trial that has only one cue onscreen) or to strategies requiring integration of probabilistic information yielded by multiple cues. Most divergent from studies with healthy human adults, the monkeys adopted a strategy very quickly and adhered to it with incredible fidelity. No monkey smoothly transitioned from less sophisticated strategy use to more sophisticated strategy use (beyond the initial transition away from random responding). In this sense the monkeys' strategy use was more like that of language-impaired children than typically-developing adults (although whether this apparent similarity is due to task motivation, cognitive similarities, both, or neither is an outstanding question; Kemény & Lukács, 2010),

Whitham and Washburn (2020) were the first to apply the strategy analyses of Meeter et al. (2006) to weather prediction task performance of animals. However, using relative probabilities of participant responses to specifically structured stimuli to yield

unique information about how participants learn and utilize the information with which they are presented is a classic technique in cognitive psychology. Indeed, weather prediction designs and analyses have much in common with identification and [deterministic] categorization designs that have played a crucial role in cognitive psychology's understanding of object identification and categorization processes (with weather prediction tasks taxonomized as "probabilistic categorization" here and elsewhere). These literatures even share the useful metaphor of stimulus representations and response probabilities as distributed in geometric (typically 2-dimensional, Euclidean) space (e.g., Ashby & Maddox, 2005). In each case, structure in the experimentally observed behaviors of an individual is used to reveal how an individual psychologically clustered the information they learned from the presented stimuli.

For example, multidimensional scaling has proven an exceptionally powerful tool for identifying these clusters and how they are distributed in different kinds of stimuli and by different individuals or groups. Shepard (1957, 1958) inaugurated the study of how minds represent physically different stimuli as conceptually so by innovating multidimensional scaling analyses for the metaphorical mapping of psychology to geometric space. This core conceit, initially relatively exploratory in scope, is among the most influential confluences of mathematics and psychology (Jaworska & Chupetlovska-Anastasova, 2009). Multidimensional scaling underpins many influential models of human [deterministic] categorization (Nosofsky, 1992; Smith & Minda, 1998), and has been extended to considerations of decision-making utility and evolutionary optimality (Herbranson et al., 1999). In this way it can inform our understanding of the cognitive evolution of human and animal minds by quantifying the degree of similarity with which psychologically relevant clusters are distributed across taxa (Smith et al., 2012).

Furthermore, the challenge faced by Whitham and Washburn (2020) of using nonverbal, behavioral measures to describe how animals complete tasks is endemic to comparative psychology and vital to its progress as a discipline. Rumbaugh (1971) described the ways in which the distribution of cognitive-behavioral data from disparate primate taxa were poorly accounted for by associative principles alone, and how percentage differences in percent correct in a simple learning task were better understood as the product of qualitatively different cognitive processes rather than of differences in parameters of associative value like learning rate. Cook, Brown, and Riley (1985) demonstrated differential leveraging of retrospective and prospective memory processes by rats as they completed radial arm mazes with interspersed interruptions. More generally, many of the scholars in comparative psychology's long tradition of maze-running designs have registered dissatisfaction with the limitations of using response latency or maze errors alone to describe the complex cognition evidenced by an animal in a maze, and sought alternative measures to offer a more complete account of how the animal completed the task (Fragaszy et al., 2003; Washburn, 1992). In these examples and many others, a premium is placed on accounts that appreciate psychology as a science of process-based accounts for psychological data and not 'merely' the rate at which animals can earn rewards in different kinds of tasks (Dutilh et al., 2019).

This dissertation will borrow liberally from these traditions in order to describe more fully the conditions in which rhesus macaques and capuchin monkeys adopt more optimal use of the provided probabilistic cues. This will allow us to better understand the processes that underpin a complex, ecologically-valid cognitive-behavioral task. These species have been previously used as an appropriate model for many high-level cognitive-behavioral phenomena of interest, including analogical reasoning (Basile et al., 2015;

Flemming et al., 2011), self-control (Beran et al., 2016; Evans et al., 2014), artificial grammar abstraction (Wilson et al., 2013; Wilson et al., 2015), and many others. Each of these species contributes uniquely to this dissertation beyond their ability to complete many thousands of trials. Both species are hardy, long-lived, omnivorous primates with relatively large brains and highly complex social dynamics, with each of these features among the strongest predictors of sophisticated cognition in animals (Deaner et al., 2007; Dunbar & Schultz, 2007; Fragaszy et al., 2004; Lindburg, 1971). Comparison of Old World (i.e., macaque) and New World (i.e., capuchin) monkey species that diverged some 40 million years ago may enrich our understanding of the evolution of the cognitive competencies that facilitate probabilistic cognition. Critically, the two species seem to vary in the degree to which they mobilize what would be called the more explicit, declarative aspects of cognition in dual-process typology. Hampton, Engelberg, and Brady (2020) reviewed a wealth of cognitive-behavioral data from rhesus macaques to conclude that these animals indeed have explicit dimensions to their cognition that are inconsistently engaged by experimental tasks, especially a restricted subset of memory, selective attention, and metacognition paradigms. Capuchin monkeys are often found to engage less successfully with this same set of memory, selective attention, and metacognition tasks (Basile et al., 2009; Beran et al., 2014; Beran et al., 2015; Beran & Smith, 2011). Newell, Lagnado, and Shanks (2007, reviewed above) convincingly demonstrated that, in humans, weather prediction tasks are principally rule-based and explicit. The apparent distinction between explicit cognition utilization of the two species, along with the greater trial experience the monkeys will receive, thus affords a unique description of the trajectory of rule-based responding as it evolves across low versus high levels of task experience for species with relatively lower and relatively higher degrees of

explicit cognition. These pre-existing group differences set up interesting comparisons which are complementary to more focal hypotheses about the underpinnings of probabilistic cognition.

By better understanding the network of individual predictors and task conditions associated with different patterns of performance on the weather prediction task, we may better understand the processes involved in the real-life situation for which the weather prediction task is a sincere model. By understanding predictors of more optimal responding specifically, we may more fully understand how to elicit more sophisticated response patterns to the situations modeled by weather prediction tasks. Finally, the inferences about probabilistic categorization gleaned by the exceedingly large trial counts and statistical precision afforded by comparative psychology may be applied to renewed explorations of how optimal responding may be encouraged in humans. In particular, it is unclear what has limited the ability of humans and other animals to make decisions that make use of more of the available cue evidence in previous weather prediction tasks. One possibility is that the 30 to 60 minute experimental sessions that human participants complete in typical weather prediction designs yields too few exposures to the embedded probabilistic relationships of the task, and that this restricts the range and complexity of strategy use. Although studies that offer humans more comprehensive experience with the task have not been published, the comparative data of Whitham and Washburn (2020) demonstrated very little continued elaboration of strategy use after initial strategy adoption. For monkeys, at least, the more fundamental determinant of behavior must be more cognitive, perceptual, motivational, or procedural. A comprehensive evaluation of the evidence for each of these possibilities may guide both the basic science of probabilistic cognition and application of this science wherever inferences are made.

2.1   Experiment 1

Experiment 1 was designed as a set of conceptual replications of Whitham and Washburn (2020) to explore the extent to which strategy use by an animal was consistent across mechanically identical weather prediction tasks. Rhesus macaques (*Macaca mulatta*) and capuchin monkeys (*Cebus [Sapajus] apella*) completed five iterations of a computerized adaptation of the cognitive-behavioral weather prediction task of Gluck, Shohamy, and Myers (2002), with each iteration using a different set of onscreen stimuli for each cue and response. If cognitive strategy use during probabilistic categorization was indeed a reliable trait of an individual, strategy analysis should reveal a similar set of strategies for the same animal across each iteration of the task. If responses were qualitatively changed by the practice that five iterations of the design might afford, disproportionately guided by early experience in the task, or demonstrably unstructured across the animals' experience with the task, strategy analysis should reveal no such consistency.

H0: The null hypothesis for each of the succeeding experiments suggests that strategy use is the product of random sampling from a population-level distribution of strategies and their frequencies, and that the strategy adopted in each new task by each monkey is a random draw from that distribution.

Hypothesis I: Strategy use during probabilistic categorization is a stable trait of an individual animal, and the same strategies used by an animal in one weather prediction task will be those used in another task.

2.2 Experiment 2

In contrast, Experiment 2 was a set of directly theory-motivated attempts to upset any consistent strategy use observed in Experiment 1 and to encourage use of multicue strategies. Experiments 2a, 2b, and 2c assessed the degree to which strategy use by the animals was dictated by parameters of rote associative evaluation like reward rate. Optimal foraging theory, monkeys' behavior in other cognitive-behavioral computer games, and motivational/associative theory converge to predict that when additional cognitive or motoric effort is required to acquire primary reinforcement, organisms are likely to adjust their behavior to meet that challenge (in a way consistent with their species and individual powers; Blanchard & Hayden, 2015; Kangas et al., 2016; Ross & Winterhalder, 2015; Washburn & Rumbaugh, 1992). In probabilistic decision-making designs with humans, the nature of strategy use has been demonstrated to change with endogenous motivation (Giner-Sorolila & Chaiken, 1997). As the reward rate for chance performance is made less and less lucrative, any animals that had not previously demonstrated above-chance performance on the weather prediction task should adopt a strategy that will yield a success rate that maintains the animal at the individual reward rate it expects of its tasks. It is possible that some animals may use more complex, multicue strategies to maximize a decremented reward rate, but this is unlikely in weather prediction designs in which single-cue and multicue strategies are nearly equivalently successful.

Hypothesis II: Behavior in weather prediction tasks is guided by factors other than an individual's preferred strategy, and titration of overall reward rate in the task will alter

strategy use accordingly. Specifically, the animals that used Random or Side-Bias strategies in Experiment 1 will use cue information to guide responding in Experiments 2a, 2b, and 2c.

Experiments 2d and 2e similarly probed the degree to which another relationship of fundamentally associative value – that of the reward frequency of responses in the presence of the different cues – guided strategy use. Nearly every previous weather prediction design has used the table of cue validities and probabilities introduced by Experiment 2 of Gluck, Shohamy, and Myers (2002) (e.g., Meeter et al., 2006; Newell et al., 2007; Shohamy et al., 2008). Although this makes the differential psychology of weather prediction easier (see Table 2.1) by facilitating comparisons across groups on the same task, other probability tables may be crafted that might yield different strategy use trajectories. As the amount of information yielded by each cue becomes more similar, multicue strategies (especially those that optimize according cue information) may become relatively more attractive whereas single-cue strategies become relatively less so. As above, we would expect the animals' behavior in the task to change accordingly.

Hypothesis III: Behavior in weather prediction tasks is guided by factors other than an individual's preferred strategy, and a set of more egalitarian cue probabilities will beget greater multicue responding in the animals that readily use strategies. Animals that do not readily use strategies will revert to Random or Side-biased strategies.

2.3   Experiment 3

In Experiment 2, each manipulation of psychological significance was titrated exactly by the experimenter to test theory-motivated hypotheses. In contrast, the manipulations included in Experiment 3 were exploratory and designed to probe theoretically interesting questions about the nature of probabilistic categorization in animals. Note that throughout Experiment 3, manipulation checks of whether monkeys perceived the stimuli in the intended manner were not readily available.

Monkeys process complex visual stimuli locally, via their more minute features, rather than globally via the whole stimulus array in its entirety (Hopkins & Washburn, 2002; Spinozzi et al., 2003). In the task of Whitham and Washburn (2020), the cue information was distributed across the computer screen such that animals might reasonably have only processed one cue or region of the screen each trial. To test whether animals might more optimally integrate cue information when local processing of this kind is discouraged, Experiment 3a explored strategy use by monkeys to a weather prediction task in which the cue information was presented in such a way that this local processing bias could be minimized. In Experiment 3a, simple cue images appeared in close spatial proximity. Another way to manipulate local processing biases of the monkeys may be to present the cue information in a context in which monkeys have been demonstrated to process information more globally, like illusory motion or faces (Leopold & Rhodes, 2010; Taubert et al., 2017). Experiment 3b oriented the probabilistic cue information as a face that the animals might be more likely to process holistically.

Hypothesis IV: Utilization of cues is, in part, a perceptual process, and design considerations that have been demonstrated to engender multicue perception will engender multicue strategy use.

Similarly, many cognitive processes are seemingly deployed differently in social contexts than they are in nonsocial contexts (Jack et al., 2013). Moreover, weather prediction tasks can be understood to mimic problems presented in many social decision-making situations. For example, a socially subordinate monkey assessing whether or not an attempt to exploit food resources would be punished by dominant conspecifics would do well to assess probabilistic information yielded by the number and identities of nearby conspecifics, the direction of conspecifics' gazes, the severity of recent aggression events, and so forth. Although a computerized task struggles to model such a context exactly, an exploration of whether strategy use is sharper (or duller) when cues resemble highly salient, socially interesting information is a natural extension of the present research. Experiments 3c and 3d provided animals with weather prediction tasks in which the predictive cues were photographs of conspecific faces and bodies, respectively.

Hypothesis V: Utilization of cues is, in part, a perceptual process, and cues that are qualitatively different than those of other experiments (socially interesting rather than asocial; discrete rather than abstract; inherently salient rather than sterilely controlled) should engender qualitatively different strategy use.

2.4 Experiment 4

Because many monkeys did not use cue information in Experiments 1, 2, or 3, the reward rate manipulations of Experiments 2a, 2b, and 2c were extended even further. The hypotheses and rationale for this manipulation were unchanged from Experiment 2.

Table 2.1 Correlational Psychology of Weather Prediction

Diagnosis	Citation	Clinical N	Performance deficit?
Alzheimer's disorder	Eldridge, Masterman, & Knowlton, 2002	8	No performance deficit, declarative knowledge of task environment impaired.
Anterograde amnesia	Knowlton, Squire, & Gluck, 1994	8	Performance deficit appears only late in testing, declarative knowledge of task environment impaired.
	Knowlton, Squire, & Mangels, 1996	12	Performance deficit appears only late in testing, declarative knowledge of task environment impaired.
	Speekenbrink, Channon, & Shanks, 2008	9	No performance deficit, declarative knowledge of task environment intact.
	Meeter et al., 2006	7	Deficit to ability to utilize and elaborate response strategy.
Developmental dyslexia	Gabay, Vakil, Schiff, & Holt, 2015	15	Performance deficits throughout testing on weather prediction tasks both with and without trial-by-trial feedback.
Parkinson's disorder	Knowlton, Squire, & Mangels, 1996	20	Performance deficits throughout testing, declarative knowledge of task environment intact.
	Moody, Bookheimer, Vanek, & Knowlton, 2004	8	No performance deficit, declarative knowledge of task environment intact.
Specific language impairment	Kemény & Lukács, 2010	16	Performance deficits throughout testing.

3   METHODS

3.1   Participants

Seven rhesus macaques (*Macaca mulatta*) and twenty capuchin monkeys (*Cebus [Sapajus] apella*) housed at the Language Research Center of Georgia State University took part in the series of experiments. Due to time and testing constraints, and differences in monkeys' enthusiasm for the task, every animal did not complete every experiment (specific participant numbers will be reported in the Results chapter). Each animal had an extensive history participating in computerized cognitive-behavioral research on tasks that probed the cognitive and comparative psychology of phenomena as diverse as perception of visual illusions (e.g., Parrish et al., 2015), self control (e.g., Evans et al., 2014), metacognition (e.g., Smith et al., 2010), and behavioral economics (Parrish et al., 2014). Many of these animals (~41%) also participated in the weather prediction task of Whitham and Washburn (2020).

Capuchin monkeys were housed in small social groups of 2 to 9 animals in large indoor enclosures with attached outdoor enclosures. Each morning, animals were given the opportunity to voluntarily separate from their main enclosure and social group into a smaller, testing enclosure. Any animal that voluntarily separated was given access to a testing apparatus for four to six hours before being returned to their main enclosure and social group. Rhesus macaques were single-housed in indoor enclosures with daily access to an outdoor play area. Testing apparatus was available to each animal 23-24 hours per day, six days per week. Animals with a socially-compatible conspecific often spent one hour each day in their outside play area with that conspecific and without access to testing apparatus.

Animals were not food- or water-deprived at any time for testing purposes, and elected to participate in the task or pursue other activities at their individual discretion. All animals were maintained on a nutritionally-complete diet of nuts, fruits, vegetables, and commercial primate food independently of their participation in any research activities. All research protocols were approved by an Institutional Animal Care and Use Committee.

3.2   Apparatus and General Procedure

Animals were tested using an adapted version of the Language Research Center - Computerized Test System (Evans et al., 2008; Richardson et al., 1990). Computer-generated stimuli were presented to the animal on a computer monitor positioned just outside the animal's testing enclosure. Animals responded to these stimuli using a joystick mounted inside their testing enclosure (capuchin monkeys) or within an arm's reach outside of their enclosure (rhesus macaques). Each correct joystick response to onscreen stimuli was rewarded with a 45 mg banana-flavored primate pellet and an ascending tone played from a speaker adjacent to the computer monitor. Each incorrect response caused a buzzing tone to be played from the same speaker, and no food reward to be dispensed. For each new manipulation in each experiment, new images for cues and outcomes were used and new spatial locations for cues and outcomes were assigned.

3.3 Experimental Procedures

3.3.1 Experiment 1 Procedures

Monkeys completed a computerized version of the classic weather prediction task of Gluck and colleagues (2002). On each trial, one, two, or three of four total cues simultaneously appeared onscreen as computer-generated visual images. The animal responded to these cues via a joystick deflection to one of two responses (R or S), also denoted onscreen as computer-generated images (Figure 3.1).

Critically, the cues combined to offer incomplete, probabilistic evidence about which of the two responses would be rewarded (see Table 3.1). These probabilities were structured such that each cue appeared equally often in each 200 trial block and each response was rewarded equally often. Cues 1, 2, 3, and 4 were associated with Response R being rewarded at validities .2, .4, .6, and .8, respectively.

A 1-second intertrial interval elapsed after correct responses, and a 4-second intertrial interval elapsed after incorrect responses. After each 1,000 completed trials, the images associated with cues and outcomes were replaced with new images, and the spatial orientations of the cues were reassigned (i.e., if the leftmost cue was cue 3 for the first 1,000 trials, it might cease to be so in subsequent experiments). Experiment 1 concluded after 5,000 trials and five sets of images for cues and outcomes.

3.3.2 Experiment 2 Procedures

3.3.2.1 Experiment 2abc Procedures.

Monkeys completed 1,000 trials of the design described in Experiment 1 for each of three new sets of response contingencies. In Experiment 2a, 4 seconds elapsed after correct

responses and 6 seconds elapsed after incorrect responses. In Experiment 2b, 5 seconds elapsed after correct responses and 8 seconds elapsed after incorrect responses. In Experiment 2c, 6 seconds elapsed after correct responses and 10 seconds elapsed after incorrect responses. These manipulations changed the overall reward rate of the task, in attempts to change the strategy use of the animals accordingly.

3.3.2.2 *Experiment 2d Procedures.*

Monkeys completed 1,000 trials of the design described in Experiment 1 with a modified probability table (Table 3.2). Cues 1, 2, 3, and 4 were associated with Response R at validities .28, .41, .59, and .72, respectively. This manipulation changed the relative information yielded by each cue, in an attempt to change the strategy use of the animals accordingly.

3.3.2.3 *Experiment 2e Procedures.*

Monkeys completed 1,000 trials of the design described in Experiment 1 with a modified probability table (Table 3.3). Cues 1, 2, 3, and 4 were associated with Response R at validities .33, .45, .55, and .67, respectively.

3.3.3 Experiment 3 Procedures

3.3.3.1 *Experiment 3a Procedures.*

Monkeys completed 1,000 trials of the design described in Experiment 1 but with cues oriented such that they are spatially contiguous (Figure 3.2).

3.3.3.2 Experiment 3b Procedures.

Monkeys completed 1,000 trials of the design described in Experiment 1 but with cues that were oriented as features on the face of a computerized version of a popular vegetable face man toy (Figure 3.3). Other weather prediction designs have used these toys as stimuli with human participants (e.g., Meeter et al., 2006; Shohamy et al., 2008).

3.3.3.3 Experiment 3c Procedures.

Monkeys completed 1,000 trials of the design described in Experiment 1 but public domain photographs of conspecific faces as cues. The animals depicted in the photographs were not familiar to any of the tested animals.

3.3.3.4 Experiment 3d Procedures.

Monkeys completed 1,000 trials of the design described in Experiment 1 but with public domain photographs of conspecific bodies as cues. The animals depicted in the photographs were not familiar to any of the tested animals.

3.3.4 Experiment 4 Procedures

Monkeys completed 1,000 trials of the design described in Experiment 1 for each of four new sets of response contingencies. In Experiment 4a, 2 seconds elapsed after correct responses and 6 seconds elapsed after incorrect responses. In Experiment 4b, 3 seconds elapsed after correct responses and 8 seconds elapsed after incorrect responses. In Experiment 4c, 4 seconds elapsed after correct responses and 10 seconds elapsed after

incorrect responses. In Experiment 4d, 4 seconds elapsed after correct responses and 14 seconds elapsed after incorrect responses.

3.3.5 *Methods Summary*

The proposed sequence of experiments promised greater depth of experimental exploration (via the wealth of data comparative psychology can offer) and greater breadth of experimental manipulation (via the diversity of intraindividual designs that experimental psychology can offer) than any previous experimentation with the uniquely ecologically valid weather prediction task (see Table 3.4). How individuals independently come to conclude that available evidence is relatively strong or weak and how it may be associated with the true state of the universe is a fundamental, pressing issue for cognitive psychology.

Table 3.1 Probability Structure of Experiment 1

Cue Pattern	Cue 1 Present?	Cue 2 Present?	Cue 3 Present?	Cue 4 Present?	Times Response R Rewarded	Times Response S Rewarded	Pattern appearances per 200 trials
A	0	0	0	1	17	2	19
B	0	0	1	0	7	2	9
C	0	0	1	1	24	2	26
D	0	1	0	0	2	7	9
E	0	1	0	1	10	2	12
F	0	1	1	0	3	3	6
G	0	1	1	1	17	2	19
H	1	0	0	0	2	17	19
I	1	0	0	1	3	3	6
J	1	0	1	0	2	10	12
K	1	0	1	1	5	4	9
L	1	1	0	0	2	24	26
M	1	1	0	1	4	5	9
N	1	1	1	0	2	17	19

Table 3.2 Probability Structure of Experiment 2d

Cue Pattern	Cue 1 Present?	Cue 2 Present?	Cue 3 Present?	Cue 4 Present?	Times Response R Rewarded	Times Response S Rewarded	Pattern appearances per 200 trials
A	0	0	0	1	16	3	19
B	0	0	1	0	6	3	9
C	0	0	1	1	21	5	26
D	0	1	0	0	3	6	9
E	0	1	0	1	7	5	12
F	0	1	1	0	3	3	6
G	0	1	1	1	16	3	19
H	1	0	0	0	3	16	19
I	1	0	0	1	3	3	6
J	1	0	1	0	5	7	12
K	1	0	1	1	5	4	9
L	1	1	0	0	5	21	26
M	1	1	0	1	4	5	9
N	1	1	1	0	3	16	19

Table 3.3 Probability Structure of Experiment 2e

Cue Pattern	Cue 1 Present?	Cue 2 Present?	Cue 3 Present?	Cue 4 Present?	Times Response R Rewarded	Times Response S Rewarded	Pattern appearances per 200 trials
A	0	0	0	1	14	5	19
B	0	0	1	0	5	4	9
C	0	0	1	1	19	7	26
D	0	1	0	0	4	5	9
E	0	1	0	1	8	4	12
F	0	1	1	0	3	3	6
G	0	1	1	1	14	5	19
H	1	0	0	0	5	14	19
I	1	0	0	1	3	3	6
J	1	0	1	0	4	8	12
K	1	0	1	1	5	4	9
L	1	1	0	0	7	19	26
M	1	1	0	1	4	5	9
N	1	1	1	0	5	14	19

Table 3.4 Experimental Psychology of Weather Prediction

Confirmatory / Exploratory	Experiment	Manipulation	Hypothesis
Confirmatory	Experiment 1abcde	Within-experiment replication, practice with task	Hypothesis I: Strategy use during probabilistic categorization is a trait of an individual animal, and the same strategies used by an animal in one weather prediction task will be those used in another task.
	Experiment 2abc, 4abcd	Reward rate for chance responding decremented	Hypothesis II: Behavior in weather prediction tasks is guided by factors other than an individual's preferred strategy, and titration of overall reward rate in the task will titrate strategy use accordingly. Specifically, the animals that use Random or Side-Bias strategies in Experiment 1 will use cue information to guide responding in Experiments 2a, 2b, and 2c.
	Experiment 2de	Cue validities decremented	Hypothesis III: Behavior in weather prediction tasks is guided by factors other than an individual's preferred strategy, and a set of task probabilities that demands multicue responding to maintain high reward rates will beget such responding in the animals that readily use strategies. Animals that do not readily use strategies will revert to Random or Side-biased strategies.
Exploratory	Experiment 3a	Cues spatially contiguous	Hypothesis IV: Utilization of cues is, in part, a perceptual process, and design considerations that have been demonstrated to engender multicue perception will engender multicue strategy use.
	Experiment 3b	Cues as face	
	Experiment 3cd	Socially-interesting cues	Hypothesis V: Utilization of cues is, in part, a perceptual process, and cues that are qualitatively different than those of other experiments (socially-interesting rather than asocial; discrete rather than abstract; inherently salient rather than sterilely controlled) should engender qualitatively different strategy use.
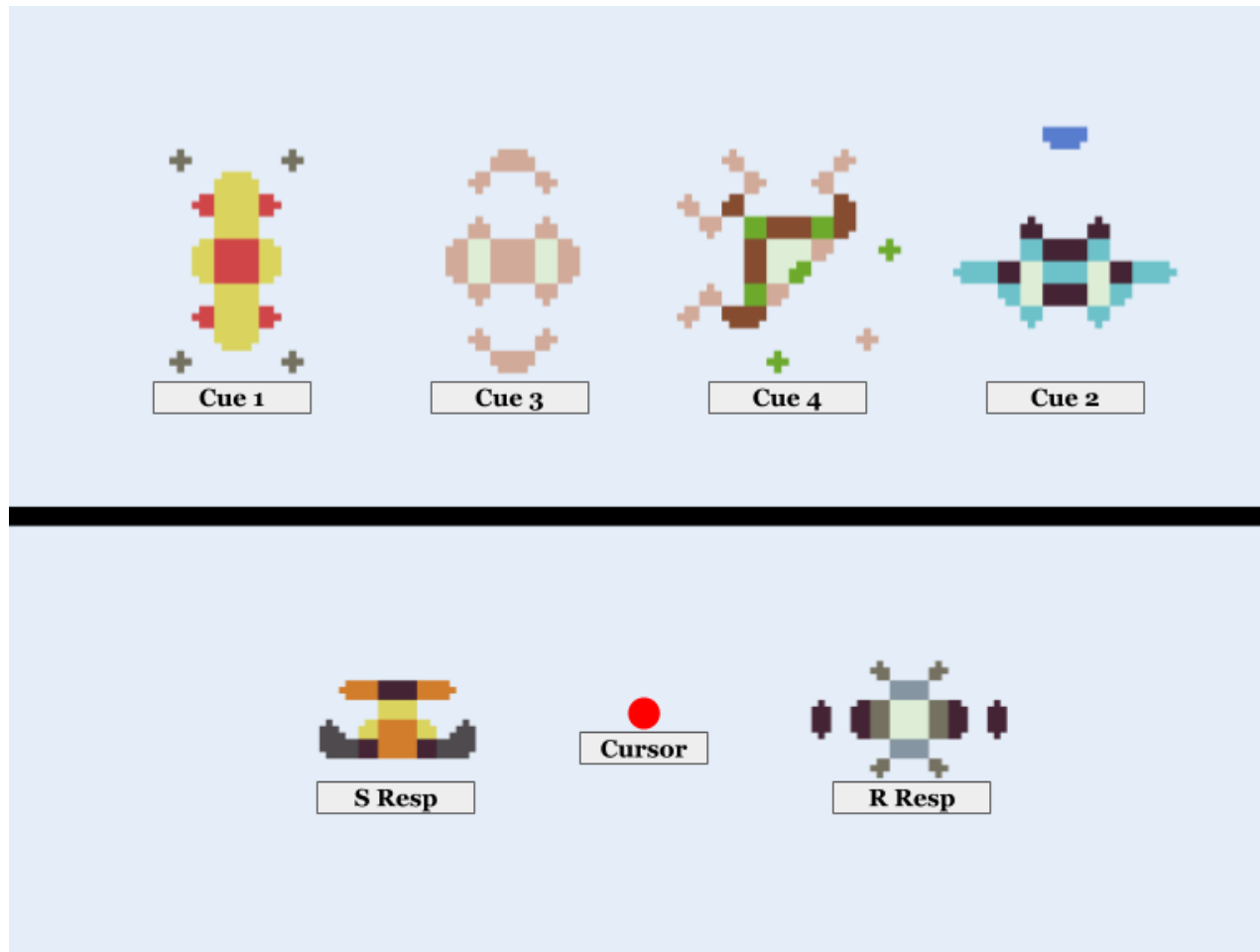
Figure 3.1 Demonstration of task screen for Experiments 1, 2, and 4. Cues, responses, and joystick-controlled cursor are labeled for clarity. All four cues are presented together only for demonstration purposes, as this pattern will not appear in any experiment.
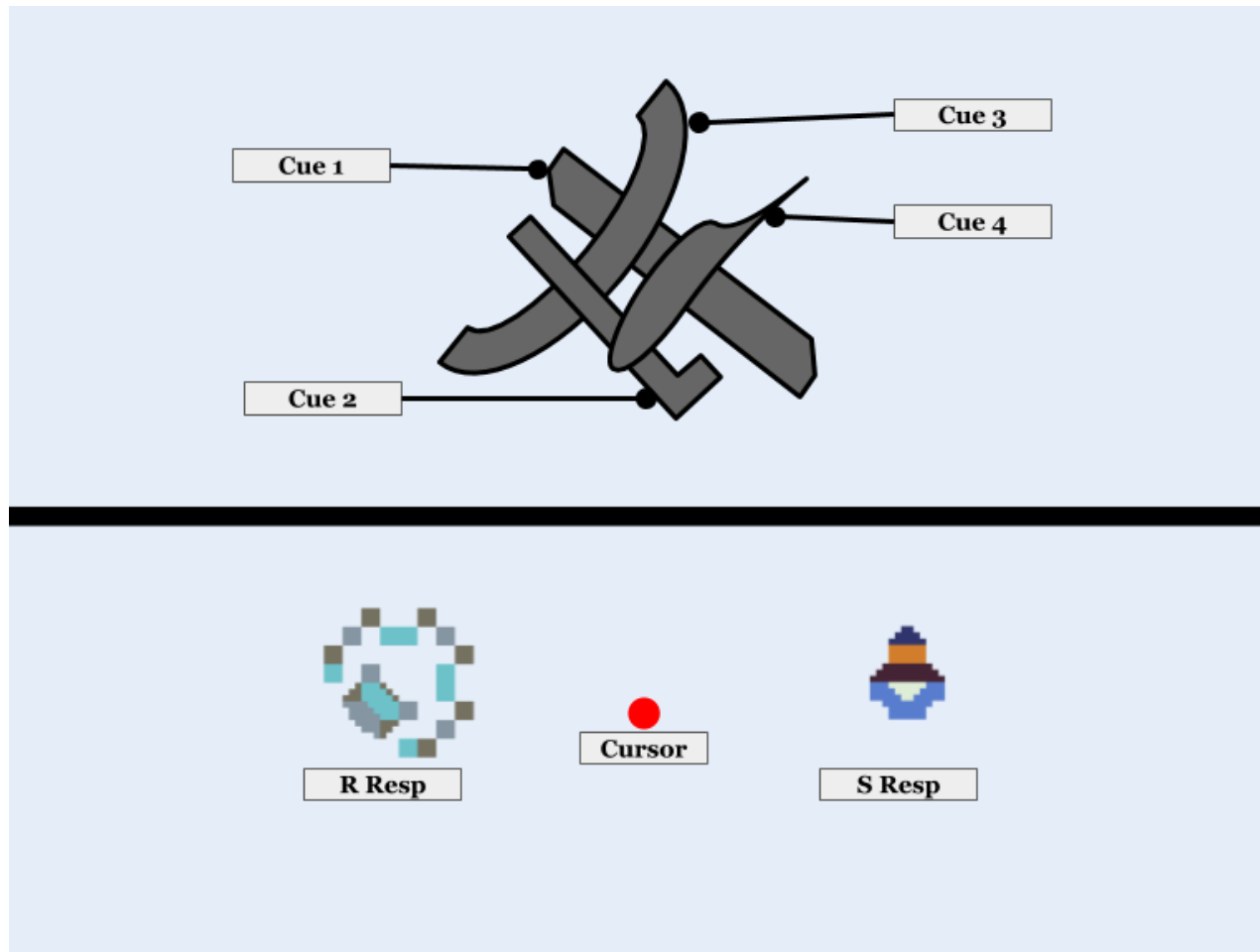
Figure 3.2 Demonstration of task screen for Experiment 3a
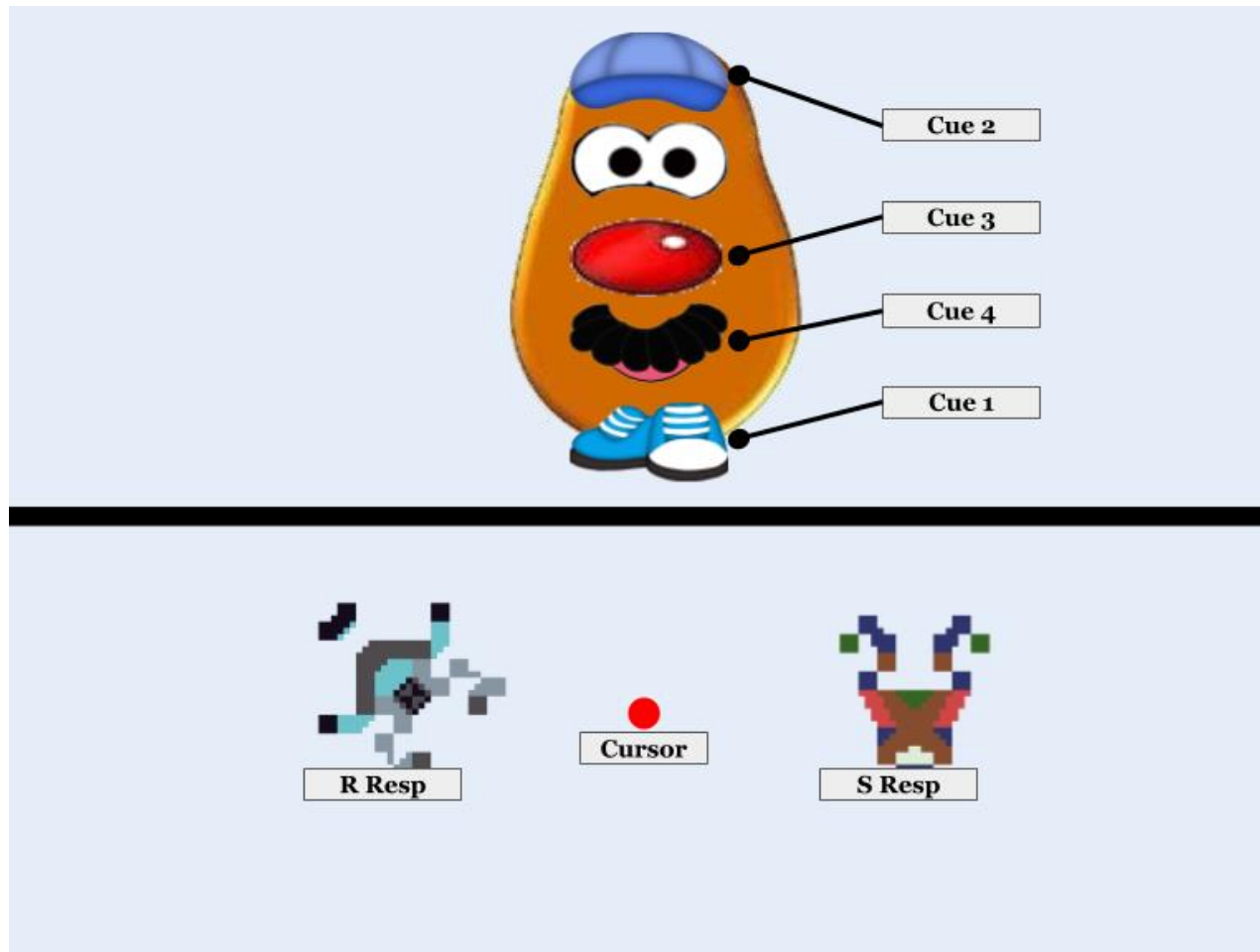
Figure 3.3 Demonstration of task screen for Experiment 3b

4 RESULTS

4.1 Analysis Plan

Whitham and Washburn's (2020) elaboration of the strategy-analysis technique of Meeter and colleagues (2006) was used to analyze data from each experiment. In such an analysis, each of an arbitrary number of possible response strategies that an animal might employ to complete the task is expressed formally as a model for the observed data, then model selection procedures are used to identify which of the strategies/models best accounts for the data. Any hypothetical strategy that can predict consistent responses to each cue pattern can be specified and tested. I tested the wide range of easily verbalizable strategies that was previously probed by Whitham and Washburn (2020).

Table 4.1 describes many of the strategies that humans and monkeys might use to solve the task. These strategies vary greatly in their cognitive and computational complexity. Some strategies predicted use of no cue information at all (e.g., Random, S or R Biases). Other strategies predicted responses on the basis of only a single cue (e.g., Singleton, Single-cue). Still other strategies predicted that the probabilistic information yielded by multiple cues would be synthesized suboptimally (e.g., Singleton-Prototype, Two v One, All but Two Strong Cards). Finally, a Perfect response strategy predicted that the optimal response will be made to each of the cue patterns. Because we expected that monkeys might adopt side biases when the strategy that they adopted does not offer a specific prediction, rather than choosing among the responses randomly, side-biased variants of many of the strategies were included. For example, a Singleton strategy predicted that the animal would make optimal responses any time a single cue was onscreen, and choose at chance otherwise; the side biased variants instead predicted that

the animal would make optimal responses any time a single cue is onscreen and make side-biased R or S responses otherwise.

The specific predictions made by each strategy, to each cue pattern, are listed in Table 4.2. Even an individual that adopted a strategy exactly as it is described in Table 4.2, however, would be very unlikely to make every predicted response to each iteration of the various cue patterns. Perseverative or motoric errors, probability matching behaviors, and exploration of the problem space are but a few of the reasons why an individual might have deviated from the predicted response (Shanks et al., 2002; Speekenbrink et al., 2008). To account for this, the rate at which forecasted responses were made to cue patterns was instead described via an adherence rate parameter. Meeter and colleagues (2006), who innovated this approach with human participants, fixed the adherence rate parameter at .95. This indicates that individuals made responses consistent with the strategy they were understood to have adopted 95% of the time. Despite the apparent incongruity between inadherent monkeys and very high predicted adherence, the .95 adherence rate remained a sensible choice. Meeter and colleagues reported simulation data that suggested that even when simulated participants adhered to a strategy on fewer than 95% of trials, a .95 adherence rate did a superior job of correctly identifying the strategy that was used to simulate the data. Lee (2016) reported that in similar strategy analyses of human behavioral data, adherence is reliably very high. Although the monkeys' decremented attention and propensity for perseverative errors with the joystick, to name two likely species differences, may have made them unlikely to adhere to a strategy so closely as do humans, assuming high adherence is the recommendation from both simulation and comparative cognition perspectives.

Each 40 trials completed by an animal was analyzed as a separate block. The fit of a strategy to each of these 40 trials blocks was estimated in the following manner. First, the binomial likelihood of seeing x S responses in n exposures to a cue pattern when the probability of an S response is .95, .05, or .5 (see Table 4.2) was computed. Such a likelihood was computed for each cue pattern that an animal was exposed to in those 40 trials (typically all cue patterns, for 14 likelihoods). Next, the product of these likelihoods was used to describe the extent to which that strategy was a good fit to the 40 trials. These two steps were repeated for each of the 25 probable strategies that were specified. The greatest value among these products (corresponding to the lowest Akaike's Information Criteria value; Akaike, 1974) was taken as the best fitting strategy for that set of trials. These steps were repeated for each block of 40 trials in an experiment, for every other experiment, and for every monkey. After the likeliest strategies were identified, inferences about strategy use trends were then assessed via analyses of strategy frequencies within and across experiments and animals, estimation of stability of strategy from block to block, visualization of strategy adoption/maintenance, and other analyses.

The optimal number of trials used to identify strategy use via this analysis, with these species, is unclear. Use of many trials would facilitate stronger fits of strategies to data when strategy use is consistent, because the influence of rare strategy inadherence on strategy fit is diluted. However, taking many trials as model input when strategy use is more inconsistent risks the model attempting to label use of multiple strategies as a single strategy. On the other end of the continuum, use of too few trials prevents the analysis from making accurate identifications in the first place. The sample size of 40 trials that this analysis plan takes as model input is informed by simulated data analyses reported by Meeter and colleagues (2006) and Whitham and Washburn (2020). This represents a

relative "Goldilocks number" at which the model input is neither too large (and risks misidentifying heterogeneous use of multiple strategies as a single strategy) nor too small (thwarting accurate strategy identification). Figure 4.1 shows the ability of the analysis perfectly or near-perfectly to recover the strategy used to simulate the data when high adherence is assumed. Accurate strategy recovery is reliably very high, with only the Perfect strategies (that make the most specific predictions) not always easily recovered from samples of 40 trials.

4.2 Descriptive Analyses

The probabilistic nature of weather prediction tasks divorces an animal's responding from the feedback contingencies for completing a trial correctly. An animal can make the optimal response that is most likely to dispense a reward, but receive no reward. The proportion of trials in which a reward was received can be useful in some modeling contexts in which rewards are the molecular unit of learning. More useful for this study was a measure of optimal responding that described the rate at which the animal made the response that was most likely to cause a pellet to be dispensed. Such a measure describes, post-hoc, mature animal learning in the task. Greater-than-chance use of the optimal response suggests structured learning of the relationships among cues, responses, and outcomes.

Figure 4.2 presents a visualization of the proportion of trials in which an animal made the optimal response in each block, of each experiment, by each monkey. Each cell describes an experiment (i.e., Experiment 1a, Experiment 1b, etc.), with the proportion of trials in which the optimal response was made to all cue patterns on the y-axis of the cell (ranging from no use of the optimal response to 100% use of the optimal response),

successive blocks presented along the x-axis of the cell, and a dashed line denoting chance use of the optimal response. Heavy black lines denote borders between Experiments 1, 2, 3, and 4. Different line colors were used based on the number of experiments completed by a monkey. Monkey names are listed on the y-axis of the figure.

Tables 4.3-6 report additional descriptive statistics for Experiments 1, 2, 3, and 4, respectively. The grand mean and standard deviation of the proportion of trials the optimal response was made were computed using the observed proportions of each of the 25 blocks of 40 trials (25 x 40 = 1,000 blocks per experiment). The grand mean and standard deviation of response latency (time between stimulus onset and response, with outliers >7 sec removed) were computed in the same manner.

Many animals (e.g., Han, Hank) never deviated from chance use of the optimal response. Many other animals (e.g., Albert throughout Experiment 2, Ingrid and Griffin throughout Experiments 1 and 2) demonstrated use of optimal responses at chance rates in early blocks of an experiment but at much higher rates as the experiment continued. The plotted data often form a characteristic, "lowercase m" shape. These individuals seemingly learned something of the relationships among cues, responses, and outcomes, with optimal response rates increasing over time as a result. When these relationships changed at the start of a new experiment, optimal response rates returned to chance and the animal began the learning process anew. The descriptive data suggest that many animals learned to use the optimal response at above-chance levels within the various experiments.

Visual inspection and conventional statistical analyses of these optimal response proportions in comparison to chance only inconsistently describes how these animals solved the task. For example, a multilevel logistic regression model could estimate the

effects of individual differences, cue information, task experience, and any number of other variables on likelihood of selecting one response or another. This analysis would output correlation coefficients for individual parameters and individual animals, and faithfully identify experiments in which observed use of the optimal response differed from chance. But such an analysis would strain to effectively describe the root cause of a network of correlation coefficients. That a monkey learned to use the optimal response in Experiment 2a, then again in Experiment 2b, is largely tangential to the question of how, qualitatively, the animal completed the task. The strategy analyses, however, can identify these more qualitative differences.

4.3 Experiment 1 Strategy Analysis

Twenty capuchin monkeys and seven rhesus macaques finished Experiment 1, combining to produce 594 blocks of responses that used cue information in an identifiably strategic way. A visualization of the best fit strategies for each block of responses by each monkey is presented in Figure 4.3. Cue-based strategies were overwhelmingly those that used one cue to guide all responses. Many monkeys used a Single-cue Strong strategy in which one response was made when a specific high validity cue was present, and the opposite response was made otherwise. At other times the Strong Singleton strategy was adopted, in which optimal responses were made only when one of the cues with highest predictive validity was the only cue presented.

Although this experiment functioned as a series of replications of Whitham and Washburn (2020), some patterns of responding were markedly different. Use of [Strong] Singleton strategies was not reported in that design, but was evidenced by several monkeys that used cue-based strategies in Experiment 1 (none of these monkeys was

included in the previous weather prediction design). The strategies used by those monkeys that used cue-based strategies did not change across the five iterations of the task.

4.4   Experiment 2 Strategy Analysis

Ten capuchin monkeys and seven rhesus macaques finished Experiment 2, in which ITIs, timeouts, and cue validities were adjusted. These animals combined to produce 274 blocks of responses that used cue information in an identifiably strategic way. A visualization of the best fit strategies for each block of responses by each monkey is presented in Figure 4.4. These [null] effects are noteworthy because, unlike in the mechanically identical manipulations of Experiment 1, Experiments 2a, 2b, 2c, 2d, and 2e each manipulated either the reward rate or the validities of the cues. Importantly, there was no evidence for a change in strategy distribution towards new or more sophisticated strategy use was observed with these changes, and the rate of non-Single Cue strategy use remained very low. Changes to reward rate and cue validity also did not cause higher rates of cue-based strategy use. One capuchin monkey, Lily, that used cues strategically in Experiment 1 ceased to do so in Experiment 2, while a second capuchin monkey, Albert, started using cue information strategically in Experiment 2 after not doing so in Experiment 1.

Curiously, both Griffin (capuchin) and Ingrid's (capuchin) Single Cue Strong strategy adoption was strongest when the validity of the strong cues was weakest, in Experiment 2e. The reasons for this are unclear, and may represent artifacts of fitful data collection, but this result minimally demonstrates that the more equivalent validities of the four cues in Experiments 2d and 2e were tractable to the animals. That is, even though

the animals overwhelming used single, strongly predictive cues to guide cue-based strategy use, and even though these experiments presented strongly predictive cues that were not as strong as in other experiments, these animals still made reliable use of single-cue strategies. This is a novel demonstration, given that manipulation of cue validities is very rare in probabilistic categorization designs, and that the cue validities and probability table of Gluck, Shohamy, and Myers (2002) (see Table 3.1) has been used in nearly every published probabilistic categorization task.

4.5   Experiment 3 Strategy Analysis

Nine capuchin monkeys and seven rhesus macaques finished Experiment 3, in which the cues were spatially contiguous or were depicted as photographs of conspecific faces or bodies. Together, these animals combined to produce 197 blocks of responses that used cue information in an identifiably strategic way. A visualization of the best fit strategies for each block of responses by each monkey is presented in Figure 4.5. The responses of each of Griffin (capuchin), Ingrid (capuchin), and Obi (rhesus) were consistently best fit in this experiment by strategies that use multiple cues, with the three animals best fit by these strategies ~93%, ~92%, and ~85% of blocks in the latter half of Experiment 3a, respectively.

Obi was identified as using a Singleton-Prototype strategy that, despite inclusion in analyses of Meeter and colleagues (2006, 2008) and Whitham and Washburn (2020), had never been previously reported to be a good fit to human or animal probabilistic categorization data. The Singleton-Prototype strategy predicted optimal responses to cue combinations in which only one cue was onscreen, or those in which two cues were onscreen that predicted the same response. A restricted subset of these rhesus macaques

had taken part in many thousands of discrete categorization tasks with spatially localized, polygonal stimuli of similar appearance to those of Experiment 3a, and Obi's data are often strongly represented in these designs (e.g., Smith et al., 2012). Obi may have been uniquely sensitive to this manipulation, and non-representative of strategy use in weather prediction tasks.

But this fact of the animals' experimental history was not true of Ingrid and Griffin. Instead, presenting the cues more locally may have caused a greater proportion of the cue information to be attended to among these three animals that, across all experiments, were generally more attentive to the task overall. The spatially localized, polygonal cues of Experiment 3a may have facilitated slightly more global processing of the cues and their associated validities, as intended. Besides the aforementioned cue-based strategy use in Experiment 3a, the experimental manipulations of Experiment 3 had no effect on the distribution or speed of cue-based strategy use. Instead, use of the more unfamiliar stimulus images in the rest of Experiment 3 seemed to be less interesting or tractable to the animals, and strategic use of cues was nearly entirely absent.

4.6 Experiment 4 Strategy Analysis

Five capuchin monkeys and seven rhesus macaques finished Experiment 4, which extended the ITI and timeous manipulations of Experiment 2. These animals combined to produce 246 blocks of responses that used cue information in an identifiably strategic way. A visualization of the best fit strategies for each block of responses by each monkey is presented in Figure 4.6. Experiment 4's extension of Experiment 2's reward rate manipulations resulted in cue-based strategy use by five monkeys – Applesauce (capuchin), Irene (capuchin), Murph (rhesus), Liam (capuchin), and Lou (rhesus) - that

had not demonstrated any appreciable degree of strategy use in Experiments 1, 2, or 3 (Liam, capuchin, used the Single-Cue strategy in unpublished data from the experiment of Whitham & Washburn, 2020). The changes to these animals' performance was limited to their adoption of strategies that other animals used in previous experiments. The manipulations did not speed up the rate of strategy adoption, nor did they cause adoption of cue-based strategies that were new to this series of experiments.

The strategy analysis (and descriptive data, see Figure 4.2 and Table 4.6) suggests strongly that Obi (rhesus) used multicue strategies in Experiment 4. In the latter halves of Experiments 4a, 4b, 4c, and 4d, Obi was best fit by multicue strategies ~71% of the time overall. However, successive blocks were rarely identified as the product of the same strategy (hence the paucity of plotted points in Figure 4.6, which only depicts multiple identical strategy identifications in a row). Without this control against strategy misidentification, the specific implications of this result can only be evaluated with caution. There are a few reasons why the strategy analysis may have reported such a diversity of multicue strategy use. Obi might have indeed changed strategies very rapidly among the multicue strategies, and the strategy analysis faithfully captured this suite of strategy use. Such a demonstration would be novel but intuitive: maintaining complex rules across hundreds or thousands of trials is cognitively effortful for humans, and would likely be more so for a rhesus macaque. Alternatively, Obi's multicue strategy use may have been consistent throughout, but the strategy analysis failed to identify it reliably. Analyses of simulated data (see Figure 4.1) discount this possibility somewhat, but a highly inconsistent adherence rate to a given strategy, for example, could spike the false identification rates much higher than those depicted in the figure. Finally, it is possible

that Obi consistently adopted the use of a strategy that was not specified in the strategy analysis.

4.7 Strategy Analysis Summary

Monkeys that reliably adopted cue-based strategies retained them within an experiment and often returned to them in each new experiment. Figure 4.7 presents a simplified Markov State diagram depicting the rates of strategy retention and switching among three groups of strategies: one group of side-biased blocks (R bias and S bias), one group of random blocks, and a third group of all blocks in which cue-based strategies were used. The state transitions depict rates of shifts from use of one of the strategy groups in a block to a different one of the strategy groups, as well as rates at which that strategy group was retained. Retention was very high overall, with the likeliest strategy for a block of responses by a monkey predicting the likeliest strategy for the next block a majority of the time for a majority of monkeys. Figure 4.8 depicts the proportion of blocks in which the highest likelihood strategy was the same across two blocks for each monkey, and overall.

Monkeys that used cue-based strategies did not demonstrate strategy use at earlier blocks as experience with weather prediction tasks increased. Figures 4.9-10 depict the block at which a cue-based strategy was acquired (defined as the same strategy having the highest likelihood for a majority of five successive blocks) for each monkey and overall, respectively. No obvious relationship existed between cue-based strategy acquisition and experience in the task. This result, along with the strategy analyses depicted in Figures 4.3-6, suggests no role of practice effects in the task and, instead, repeated measurement of identical series of behaviors to a series of mechanically identical tasks (with the

exceptions of Experiments 2d and 2e, which varied the probability table for the task yet still maintained the general results).

4.8 Other Analyses

Response latencies, both the time from trial onset to response and the total time in which the joystick was depressed, were also recorded for each trial. Although there were no specific hypotheses about response latencies, analyses of this additional measure might capture aspects of the monkeys' performance in the task to which the strategy analysis was blind. For example, it may be the case that each strategy entails a certain amount of processing time, and that response time may be used as a behavioral marker of strategy use alongside rates of responding to the different cue patterns. Differences in processing time would perhaps be evidenced in response latencies, and not captured by the strategy analysis alone (see Tables 4.3-6 for descriptive statistics of response latencies).

Bayesian modeling was used to detect any differences in response latencies. Central tendency and scale parameters were estimated to describe the mean and standard deviation, respectively, of a latent distribution that modeled response latency. These parameters were estimated once for each of three groups of strategies. One central tendency and scale pair was estimated for trials in which the strategy analysis suggested R or S biased responding. A second pair was estimated for trials in which the strategy analysis suggested random responding. A third pair was estimated for trials in which the strategy analysis suggested cue-based responding. Because response latencies are heavily skewed (response latencies can extend to the end of the test session, but never be faster than around 1 second), two transformations of response latencies were made to facilitate

modeling (for strengths and weaknesses of this approach, see Grayson, 2004; Lo & Andrews, 2015). First, only response latencies faster than 7 sec were included in the analysis. This arbitrary cutoff removes the slowest 7% of trials, that likely represented monkey behaviors that are not of interest to the present investigation (e.g., drinking water, taking a nap, interactions with conspecifics, etc.), while preserving response latencies that could reasonably represent time taken to process the cue information and make a response. Second, response latencies were natural log transformed to remove remaining skew from the distribution and sharpen parameter estimates. The response latencies that capture time from trial onset to response were used, as processing time could plausibly include times in which the animal was not deflecting the joystick.

Bayesian modeling techniques were used because they offer flexible matching of analysis to design, emphasize estimation of parameters of interest instead of binary decision criteria, facilitate identification of where differences demonstrably do not exist (akin to accepting the null hypothesis in frequentist significance testing), and cope well with unbalanced designs (Kruschke, 2018; Kruschke & Liddell, 2018a; McElreath, 2020). Although the central tendency and scale of response latency distributions would be expected to vary across individual monkeys as well, and thus could have been included in a Bayesian multilevel model, the severely unbalanced distribution of inferred strategy use prevented effective modeling of this potential predictor. Some monkeys only very rarely contributed response latencies due to cue-based strategy use, and the high retention rates described above made the individual monkeys highly correlated with their strategy of choice. Adding monkey identity to the model would only convolute measurement of the strategy group parameters of interest. Analyses were completed using the pandas and pymc3 packages of Python 3.7 (McKinney, 2011; Salvatier et al., 2016). Parameter

estimates were made using 20,000 Markov Chain Monte Carlo steps (10,000 tuning steps).

Results of this response latency analysis, with distribution mean estimates transformed back to the original units (msec), are plotted in Figure 4.11. Trials from blocks in which side biases were inferred were demonstrably faster than trials from blocks in which random or cue-based strategies were inferred. Because processing of cues was likely very low for monkeys employing these side-biased strategies, akin to lever-pressing on a variable ratio reward schedule, this effect was to be expected. It is more interesting that no difference was observed between blocks in which random strategies were inferred and those in which cue-based strategies were inferred. If greater response latency suggests greater time processing the stimulus information, the equivalent response latencies of random strategy blocks and cue-based strategy blocks suggests sincere processing of cue information in random strategy blocks.

This result may suggest use of a strategy that was not specified in the strategy analysis in these blocks. This failure of model specification and selection would be analogous to a Type II error – there is structure to the data, but the model is not sensitive enough to capture it and the analyst erroneously reports no effect. Many animals use win-stay and lose-shift strategies in highly serialized tasks with binary response options (e.g., Levine, 1959). Figure 4.12 depicts the rate of win-stay/lose-shift strategy use by monkeys in blocks of random strategy use. As in Whitham and Washburn (2020), no evidence for use of this particular serial dependency in the monkeys' responses was observed. Further descriptive analyses did not reveal any obvious cue dependency in the random blocks as might suggest specification of an additional, heretofore unidentified strategy. Taken together these analyses suggest that random strategy use reflected failed attempts to make

use of the available cue information, or otherwise misinformed guesses, rather than structured responding using an approach not included in the strategy analysis.

Table 4.1 Strategies

Description
Random	Each response is made randomly.
R bias	R responses are predicted to every cue combination.
S bias	S responses are predicted to every cue combination.
Singleton	Optimal responses are predicted when only one cue is onscreen.
	Random or side-biased responses are predicted when multiple cues are onscreen.
Single-cue [cue]	The presence or absence of one cue guides all responses - to the response predicted by that cue if that cue if it is present, or to the opposite response if it is not present.
Singleton-Prototype	Optimal responses are predicted when one cue is onscreen or two cues that predict the same response are onscreen.
	Random or side-biased responses are predicted when cues that predict opposite responses are onscreen.
Two v One	Optimal responses are predicted when one cue is onscreen, when two cues that predict the same response are onscreen, or when two cues that predict the same response and one cue that predicts the opposite response are onscreen.
	Random or side-biased responses are predicted when cues that predict opposite responses are onscreen and majority rule cannot guide responses.
All but Two Strong Cues	The optimal response is predicted for every cue combination except those in which both of the strongly predictive cues are onscreen together.
Perfect	The optimal response is predicted to every cue combination.

Table 4.2 Cue Patterns, Frequencies, and Strategies. Each row describes the set of responses a strategy predicts to the different cue patterns. A strategy predicted a deflection to the R response, the S response, or make no specific prediction (chance rates of responding among the two two responses, Rand). Cues can be strong (R or S) or weak (r or s) evidence that a response will be rewarded. The rate at which an animal will be assumed to adhere to a strategy is not 100%, to allow for anticipated deviations due to exploratory behavior, probability matching, and other factors. Each strategy predicted a unique set of R responses, S responses, and random responses to the 14 unique cue patterns.

	Cue Patterns													
	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**
	Cues Present													
	R	**r**	**Rr**	**s**	**Rs**	**rs**	**Rrs**	**S**	**RS**	**rS**	**RrS**	**Ss**	**RSs**	**rSs**
Strategies not based on cue information														
Random	Rand	Rand	Rand	Rand	Rand	Rand	Rand	Rand	Rand	Rand	Rand	Rand	Rand	Rand
R bias	R	R	R	R	R	R	R	R	R	R	R	R	R	R
S bias	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Strategies based on a single cue														
Strong Singleton	R	Rand	Rand	Rand	Rand	Rand	Rand	S	Rand	Rand	Rand	Rand	Rand	Rand
Strong Singleton w/ R bias	R	R	R	R	R	R	R	S	R	R	R	R	R	R
Strong Singleton w/ S bias	R	S	S	S	S	S	S	S	S	S	S	S	S	S
Singleton	R	R	Rand	S	Rand	Rand	Rand	S	Rand	Rand	Rand	Rand	Rand	Rand
Singleton w/ R bias	R	R	R	S	R	R	R	S	R	R	R	R	R	R
Singleton w/ S bias	R	R	S	S	S	S	S	S	S	S	S	S	S	S
Single-Cue Strong R	R	S	R	S	R	S	R	S	R	S	R	S	R	S
Single-Cue Weak R	S	R	R	S	S	R	R	S	S	R	R	S	S	R
Single-Cue Weak S	R	R	R	S	S	S	S	R	R	R	R	S	S	S
Single-Cue Strong S	R	R	R	R	R	R	R	S	S	S	S	S	S	S
Strategies based on multiple cues														
Singleton Prototype	R	R	R	S	Rand	Rand	Rand	S	Rand	Rand	Rand	S	Rand	Rand
Singleton Prototype w/ R bias	R	R	R	S	R	R	R	S	R	R	R	S	R	R
Singleton Prototype w/ S bias	R	R	R	S	S	S	S	S	S	S	S	S	S	S
Two v One	R	R	R	S	Rand	Rand	R	S	Rand	Rand	R	S	S	S
Two v One w/ R bias	R	R	R	S	R	R	R	S	R	R	R	S	S	S
Two v One w/ S bias	R	R	R	S	S	S	R	S	S	S	R	S	S	S
All but Two Strong Cards	R	R	R	S	R	Rand	R	S	Rand	S	Rand	S	Rand	S
All but Two Strong Cards w/ R bias	R	R	R	S	R	R	R	S	R	S	R	S	R	S
All but Two Strong Cards w/ S bias	R	R	R	S	R	S	R	S	S	S	S	S	S	S
Perfect	R	R	R	S	R	Rand	R	S	Rand	S	R	S	S	S
Perfect w/ R bias	R	R	R	S	R	R	R	S	R	S	R	S	S	S
Perfect w/ S bias	R	R	R	S	R	S	R	S	S	S	R	S	S	S

Table 4.3 Descriptive statistics for Experiment 1

	Experiment 1a				Experiment 1b				Experiment 1c				Experiment 1d				Experiment 1e			
	p(Optimal)		Latency		p(Optimal)		Latency		p(Optimal)		Latency		p(Optimal)		Latency		p(Optimal)		Latency	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Albert	0.50	0.06	1924	291	0.51	0.07	1847	270	0.51	0.09	1818	255	0.62	0.12	1857	228	0.65	0.11	1735	213
Applesauce	0.50	0.07	2253	396	0.50	0.07	2357	327	0.50	0.08	2212	258	0.50	0.07	2294	220	0.50	0.08	2346	504
Bailey	0.51	0.09	2357	281	0.64	0.16	2194	236	0.50	0.08	1688	613	0.50	0.07	1374	419	0.52	0.09	2026	230
Benny	0.50	0.08	1301	255	0.50	0.08	1175	177	0.50	0.06	1211	158	0.50	0.07	1166	157	0.50	0.09	1075	87
Chewie	0.50	0.08	1108	150	0.50	0.06	1134	151	0.50	0.08	1185	110	0.49	0.08	1147	119	0.50	0.07	1136	195
Gabe	0.50	0.06	2160	252	0.50	0.07	2273	187	0.50	0.08	2089	217	0.50	0.09	2158	189	0.50	0.07	2065	178
Gambit	0.55	0.08	3665	316	0.56	0.09	3531	366	0.52	0.09	3386	294	0.76	0.16	2983	234	0.79	0.17	3424	280
Gonzo	0.53	0.08	2436	302	0.60	0.15	2513	249	0.58	0.12	2515	201	0.82	0.14	2509	269	0.60	0.23	2449	272
Gretel	0.54	0.08	2543	279	0.69	0.17	2738	325	0.49	0.08	2633	274	0.63	0.09	2523	256	0.78	0.12	2265	231
Griffin	0.51	0.09	2526	259	0.64	0.12	2243	253	0.76	0.16	2357	216	0.75	0.15	2150	283	0.66	0.15	2090	346
Han	0.52	0.09	2094	402	0.49	0.08	2168	200	0.49	0.05	2232	265	0.49	0.07	1892	273	0.50	0.09	1631	288
Hank	0.50	0.08	3222	267	0.50	0.07	2959	401	0.50	0.05	2799	228	0.50	0.07	2790	176	0.50	0.06	3015	330
Ingrid	0.55	0.09	2531	311	0.77	0.16	2458	366	0.77	0.15	2351	216	0.75	0.19	2408	240	0.79	0.17	2312	264
Ira	0.51	0.06	2609	438	0.50	0.07	2073	281	0.57	0.13	2083	258	0.80	0.12	2261	229	0.60	0.1	2252	206
Irene	0.50	0.08	2875	307	0.50	0.05	2305	175	0.48	0.08	2540	219	0.54	0.11	2720	353	0.61	0.13	2775	225
Ivory	0.50	0.10	2610	327	0.50	0.08	2285	211	0.50	0.07	2141	198	0.53	0.09	2061	328	0.51	0.1	2166	329
Liam	0.57	0.07	2818	291	0.52	0.07	2325	265	0.55	0.07	2261	371	0.50	0.06	1883	224	0.47	0.08	2316	474
Lily	0.49	0.07	2441	186	0.61	0.16	2176	213	0.59	0.13	2077	208	0.78	0.13	2257	252	0.60	0.12	2085	211
Logan	0.51	0.09	2355	214	0.52	0.08	2517	241	0.49	0.1	2412	326	0.51	0.06	2555	211	0.58	0.1	2388	215
Lou	0.48	0.07	2211	287	0.51	0.07	2263	286	0.54	0.06	2713	248	0.53	0.08	2368	265	0.49	0.07	2119	251
Luke	0.50	0.05	1813	307	0.50	0.06	1805	238	0.50	0.08	1602	241	0.50	0.08	1297	179	0.50	0.08	1450	280
Lychee	0.48	0.09	2698	290	0.48	0.08	2237	214	0.52	0.07	2257	305	0.50	0.08	2277	240	0.49	0.11	2004	219
Mason	0.51	0.09	2758	411	0.50	0.07	2681	169	0.52	0.09	2583	194	0.48	0.07	2590	131	0.53	0.08	2809	297
Murph	0.51	0.07	1475	235	0.50	0.08	1636	150	0.49	0.08	1646	135	0.51	0.08	1867	308	0.50	0.08	1510	180
Nkima	0.53	0.08	2425	417	0.50	0.09	1634	423	0.50	0.07	1362	242	0.48	0.1	2626	218	0.48	0.06	2583	255
Obi	0.53	0.08	2105	326	0.51	0.06	1945	176	0.51	0.08	1809	224	0.49	0.09	2055	317	0.50	0.07	1412	214
Paddy	0.48	0.07	2879	209	0.52	0.08	2985	376	0.58	0.08	2802	448	0.70	0.13	2569	467	0.50	0.08	1070	40
Widget	0.50	0.09	2047	119	0.50	0.06	1813	232	0.66	0.17	1750	218	0.73	0.13	1779	128	0.80	0.11	1871	213
Wren	0.52	0.06	2280	310	0.57	0.09	2146	236	0.52	0.11	2024	178	0.71	0.13	2084	147	0.69	0.10	2261	223

Table 4.4 Descriptive statistics for Experiment 2

| | Experiment 2a | | | | Experiment 2b | | | | Experiment 2c | | | | Experiment 2d | | | | Experiment 2e | | | |
| | p(Optimal) | | Latency | | p(Optimal) | | Latency | | p(Optimal) | | Latency | | p(Optimal) | | Latency | | p(Optimal) | | Latency | |
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Albert	0.75	0.15	2307	601	0.63	0.12	2353	265	0.69	0.19	2285	793	0.6	0.14	2704	310	0.68	0.15	2533	428
Applesauce	0.51	0.07	2258	298	0.50	0.08	2317	261	0.56	0.14	2511	348	0.53	0.10	2999	312	0.50	0.08	2354	242
Bailey	0.51	0.07	1644	314	0.51	0.07	1956	281	0.49	0.09	1835	210	0.49	0.09	1973	217	0.49	0.09	2282	363
Chewie	0.50	0.06	1145	97	0.50	0.07	1158	87	0.50	0.08	1088	79	0.51	0.09	1135	98	0.50	0.07	1145	162
Gabe	0.51	0.08	1474	239	0.50	0.06	1689	263	0.50	0.07	1919	258	0.50	0.07	2289	233	0.50	0.07	2261	210
Griffin	0.62	0.11	2251	502	0.55	0.11	2732	337	0.57	0.11	2912	284	0.63	0.12	2718	271	0.76	0.15	2473	236
Han	0.51	0.08	2110	349	0.52	0.08	2007	629	0.50	0.06	1043	0	0.50	0.07	1043	0	0.50	0.07	1043	0
Hank	0.50	0.07	2388	250	0.50	0.07	2160	312	0.48	0.07	2602	435	0.50	0.07	3058	274	0.49	0.07	2998	352
Ingrid	0.77	0.16	2955	445	0.61	0.09	2889	433	0.47	0.08	3021	231	0.64	0.22	2319	214	0.86	0.10	2586	211
Irene	0.56	0.10	2302	316	0.54	0.10	2281	174	0.64	0.12	2535	390	0.50	0.09	2840	331	0.51	0.08	2301	339
Ivory	0.50	0.09	1982	243	0.54	0.10	2233	313	0.50	0.07	2008	225	0.50	0.10	2189	381	0.50	0.06	1892	301
Liam	0.53	0.08	1756	288	0.50	0.08	1471	235	0.51	0.06	1491	230	0.50	0.08	1415	324	0.50	0.09	1181	119
Lily	0.49	0.08	2164	279	0.50	0.07	2428	626	0.48	0.07	2538	432	0.50	0.08	1999	238	0.50	0.08	2252	316
Lou	0.48	0.09	2245	230	0.49	0.07	2069	212	0.56	0.09	2468	228	0.50	0.07	2073	263	0.49	0.07	2167	301
Luke	0.50	0.07	1796	299	0.50	0.07	1711	202	0.50	0.08	1418	291	0.50	0.08	2002	280	0.50	0.08	1575	195
Murph	0.52	0.09	1609	171	0.58	0.07	2101	359	0.52	0.09	1903	219	0.49	0.08	1685	281	0.52	0.09	1307	145
Obi	0.54	0.08	2521	448	0.58	0.08	2545	370	0.55	0.08	2424	287	0.52	0.07	1761	172	0.50	0.07	1780	249
Wren	0.56	0.09	2272	278	0.53	0.09	2427	784	0.42	0.10	2636	356	0.50	0.08	2524	268	0.55	0.07	2184	222

Table 4.5 Descriptive statistics for Experiment 3

	Experiment 3a				Experiment 3b				Experiment 3c				Experiment 3d			
	p(Optimal)		Latency		p(Optimal)		Latency		p(Optimal)		Latency		p(Optimal)		Latency	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Albert	0.69	0.14	2379	316	0.55	0.07	2668	467	0.65	0.11	2683	484	0.60	0.13	2157	451
Applesauce	0.51	0.06	2416	369	0.49	0.07	2540	408	0.50	0.09	2652	378	0.51	0.09	2272	290
Bailey	0.50	0.08	2188	268	0.51	0.07	2030	168	0.49	0.07	1842	272	0.49	0.10	1778	255
Chewie	0.50	0.09	1142	149	0.50	0.08	1225	253	0.50	0.10	1079	80	0.50	0.08	1031	13
Gabe	0.50	0.05	2177	296	0.50	0.06	2155	230	0.49	0.08	2141	246	0.50	0.09	2313	432
Griffin	0.82	0.11	2132	146	0.53	0.08	2409	323	0.56	0.12	2413	335	0.70	0.17	2104	188
Han	0.50	0.07	1042	1	0.50	0.08	1042	0	0.50	0.07	1043	0	0.51	0.09	1769	402
Hank	0.51	0.05	3061	234	0.50	0.09	2924	240	0.50	0.08	2903	244	0.50	0.07	3058	316
Ingrid	0.73	0.14	2502	177	0.52	0.09	2486	186	0.49	0.07	2653	228	0.47	0.08	2466	291
Irene	0.49	0.07	2341	181	0.50	0.07	2505	302	0.60	0.07	2403	308	0.57	0.06	2529	331
Liam	0.51	0.07	1933	522	0.51	0.06	1992	238	0.51	0.09	1747	130	0.50	0.10	1715	187
Lily	0.50	0.08	2201	180	0.50	0.07	2385	530	0.50	0.09	2037	171	0.50	0.06	2150	364
Lou	0.51	0.07	2119	321	0.50	0.06	1992	219	0.48	0.08	1902	247	0.51	0.08	2130	191
Luke	0.50	0.07	1498	146	0.50	0.06	1711	271	0.50	0.08	1678	207	0.50	0.08	1611	184
Murph	0.50	0.08	1242	94	0.51	0.09	1551	216	0.53	0.08	1282	136	0.52	0.08	1621	154
Obi	0.71	0.13	2103	292	0.52	0.09	2116	252	0.55	0.09	1519	327	0.60	0.13	1578	177

Table 4.6 Descriptive statistics for Experiment 4

	Experiment 4a				Experiment 4b				Experiment 4c				Experiment 4d			
	p(Optimal)		Latency		p(Optimal)		Latency		p(Optimal)		Latency		p(Optimal)		Latency	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Applesauce	0.60	0.10	1988	289	0.51	0.07	2434	334	0.50	0.08	2610	525	0.77	0.08	2263	294
Chewie	0.50	0.08	1087	73	0.50	0.06	1055	30	0.50	0.07	1232	200	0.50	0.07	1257	212
Han	0.52	0.08	2257	207	0.54	0.08	2283	233	0.52	0.09	1388	402	0.51	0.08	1820	688
Hank	0.51	0.07	2865	207	0.51	0.09	2414	254	0.50	0.06	2235	171	0.50	0.09	2458	334
Ingrid	0.46	0.10	2499	204	0.48	0.06	2723	157	0.47	0.10	3039	415	0.53	0.11	3177	417
Irene	missing data				0.57	0.09	1967	195	0.53	0.12	2291	244	0.68	0.20	2394	242
Liam	0.57	0.12	2203	660	0.61	0.10	2457	382	0.68	0.11	2381	237	0.76	0.11	2390	296
Lily	missing data				0.50	0.08	1832	212	0.52	0.07	2080	404	0.50	0.07	2315	570
Lou	0.51	0.09	2280	271	0.46	0.08	2172	239	0.51	0.08	2237	246	0.78	0.15	2261	183
Luke	0.50	0.07	2111	445	0.50	0.09	1572	245	0.50	0.09	1956	476	0.50	0.07	1767	397
Murph	0.60	0.14	1515	372	0.64	0.11	1972	390	0.74	0.11	2199	284	0.83	0.07	2227	231
Obi	0.71	0.10	2114	311	0.73	0.09	1880	200	0.74	0.15	2176	305	0.76	0.11	2141	303

Figure 4.1 Simulated strategy recovery. The strategy used to simulate response data is printed in the top-right corner of each plot. 'Exact' recoveries are the rate at which the strategy analysis recovered the exact strategy that was used to simulate response data. 'Strategy' recoveries are the rate at which the strategy analysis identified the correct strategy except for misidentifying the presence or absence of any S or R biases. Both of these types of recoveries are accurate. In 'Other' recoveries the strategy that was used to generate the data was misidentified. Figure adapted from Whitham & Washburn (2020).
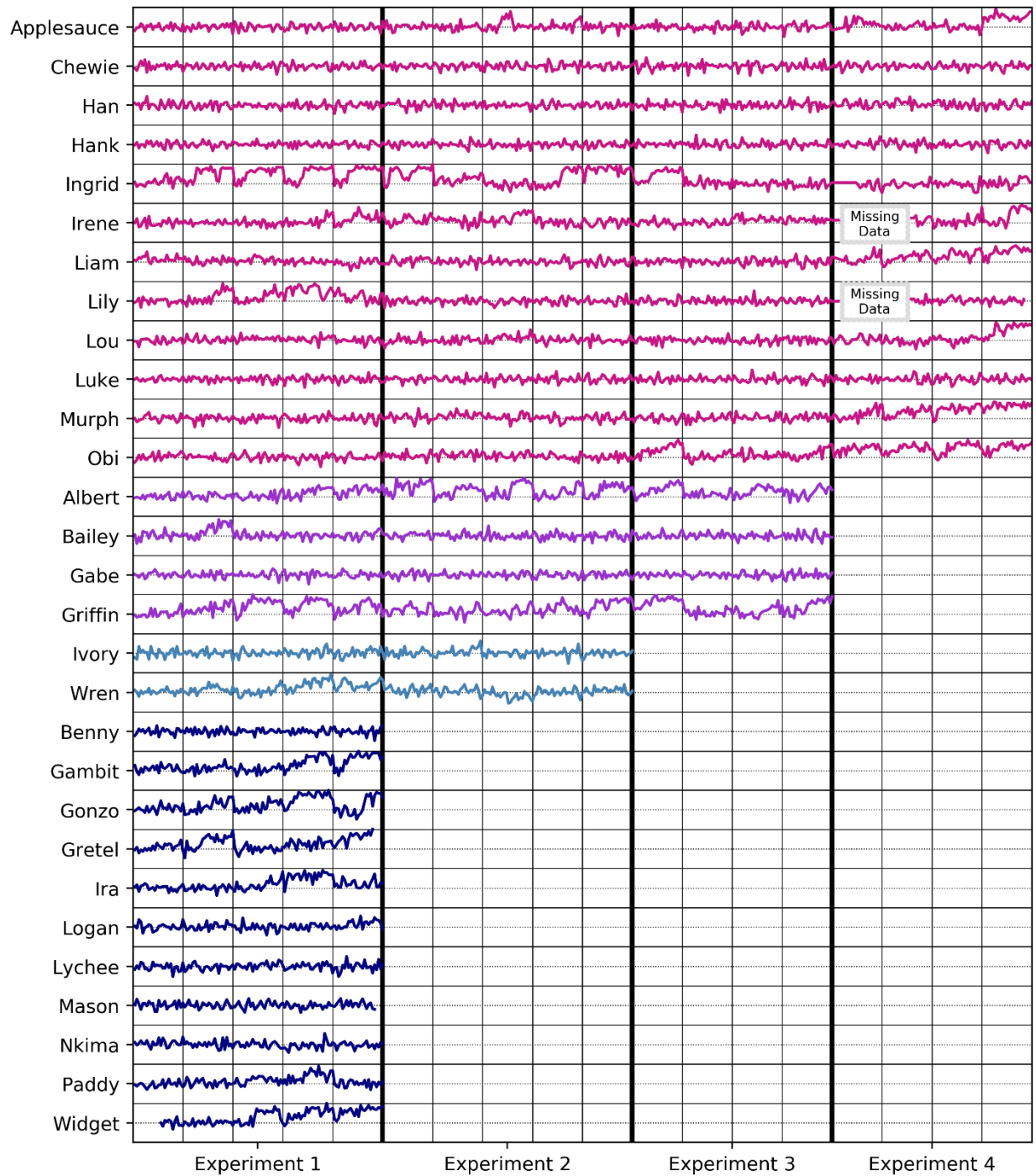
Figure 4.2. Response optimality. The y-axis of each plot depicts the proportion of trials in which the optimal response was used in each experiment by each monkey.
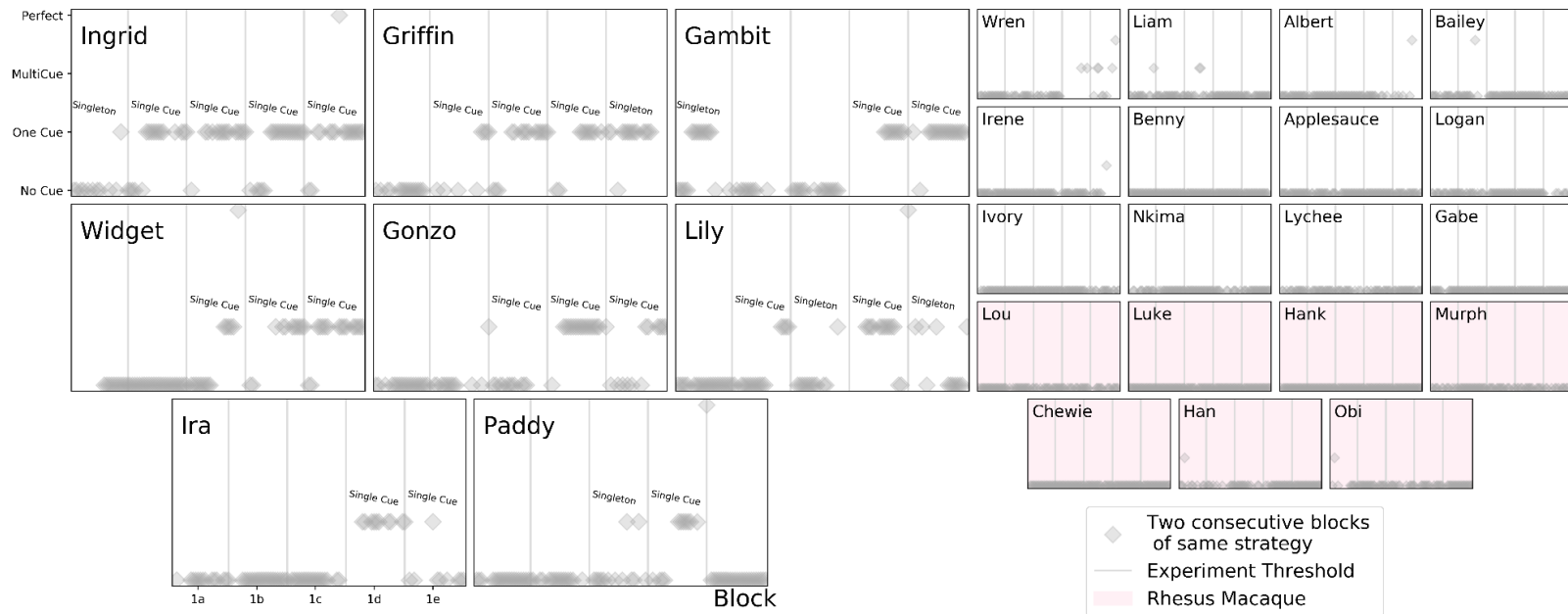
Figure 4.3 Strategy use in Experiment 1. The many possible strategies of the strategy analysis are grouped into four bins for plotting on the y-axis: no cue use (random or side-biased strategies), use of one cue (single-cue or singleton strategies), use of multiple cues (singleton-prototype, two vs one), or [nearly] perfect strategies (all but two strong cards and perfect strategies). The eight capuchin monkeys on the left side of the plot used cues in their responding substantially more than the other animals, and the strategies they used are annotated within each experiment accordingly. Note that these bins are only for visualization purposes, and were not incorporated into the modeling.

Figure 4.4 Strategy use in Experiment 2. The many possible strategies of the strategy analysis are grouped in the same way as in Figure 4.3. The three capuchin monkeys on the left side of the figure used cues in their responding substantially more than the other animals, and the strategies they used are annotated within each experiment accordingly.

Figure 4.5 Strategy use in Experiment 3. The many possible strategies of the strategy analysis are grouped in the same way as Figure 4.3. The four monkeys on the left side of the figure used cues in their responding substantially more than the other animals, and the strategies they used are annotated within each experiment accordingly.
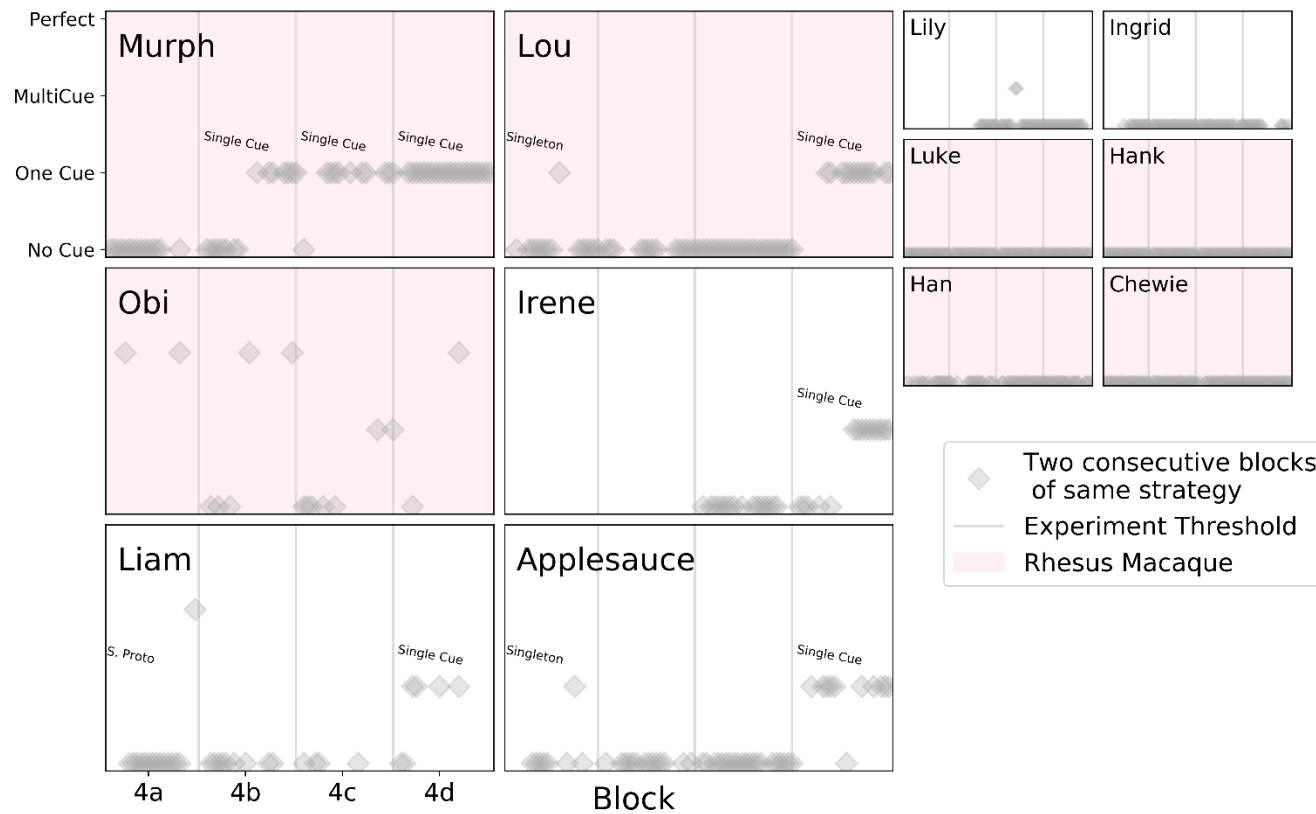
Figure 4.6 Strategy use in Experiment 4. The many possible strategies of the strategy analysis are grouped in the same way as Figure 4.3. The five monkeys on the left side of the figure used cues in their responding substantially more than the other animals, and the strategies they used are annotated within each experiment accordingly.
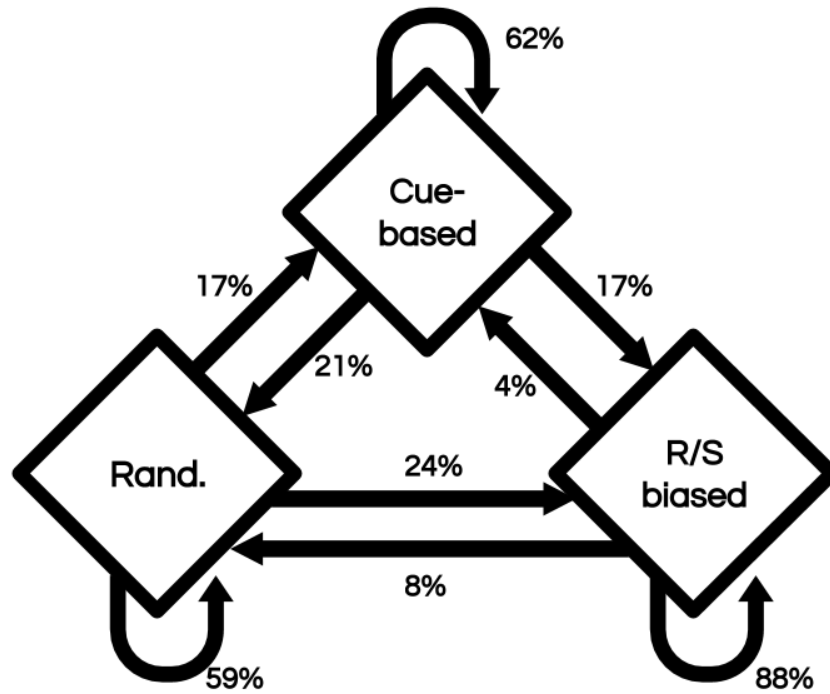
Figure 4.7 Simplified Markov state diagram. Arrows denote block-over-block rates of switching and retention among R or S biased strategies, the random strategy, and cue-based strategies.
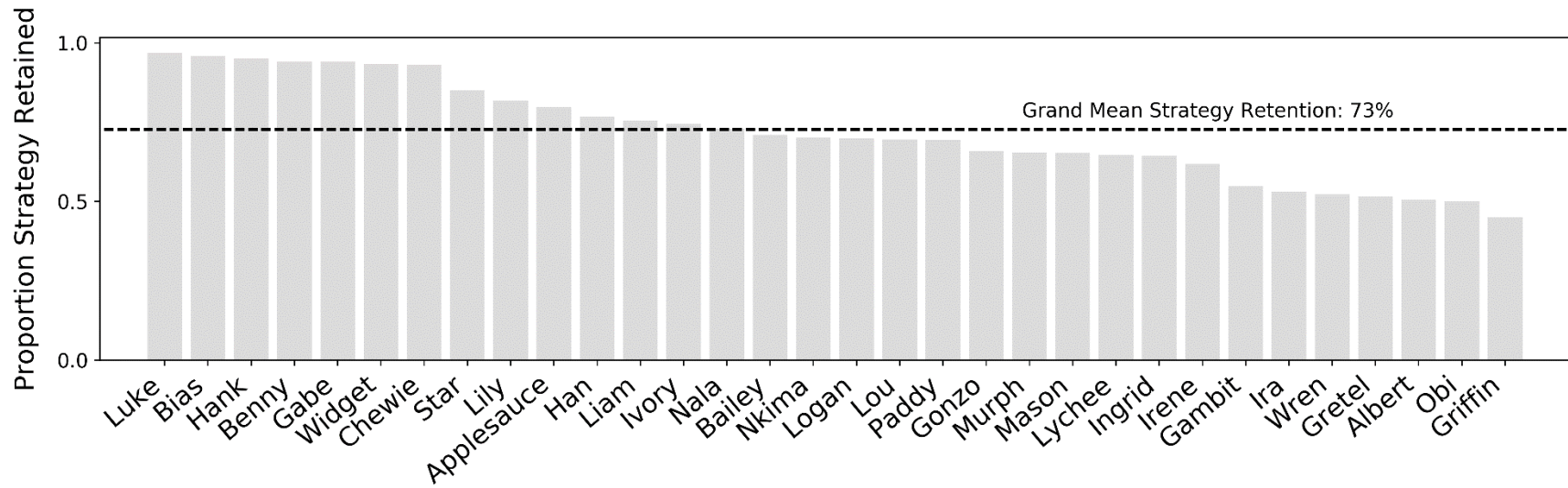
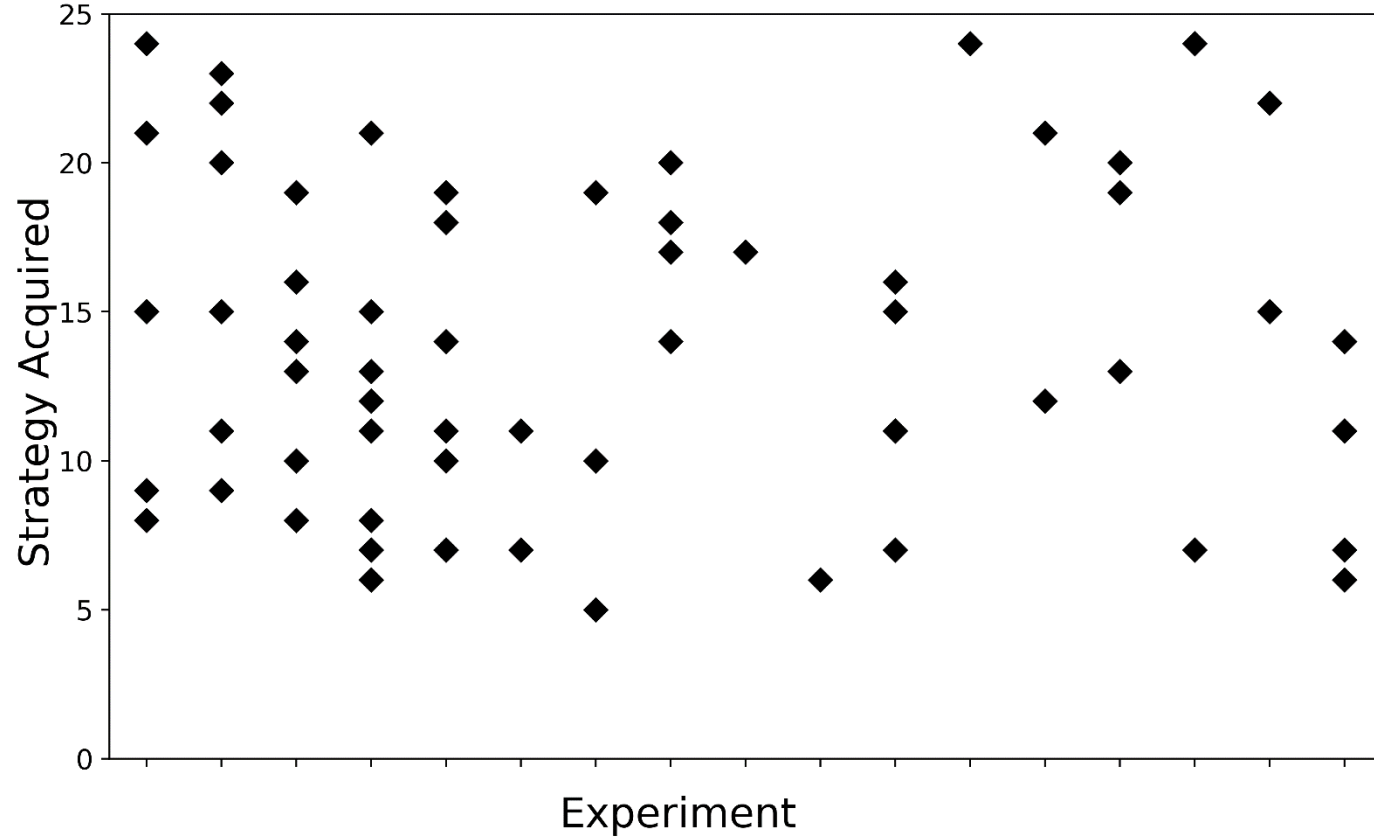*Figure 4.8 Strategy retention rates by individual monkeys and overall.*
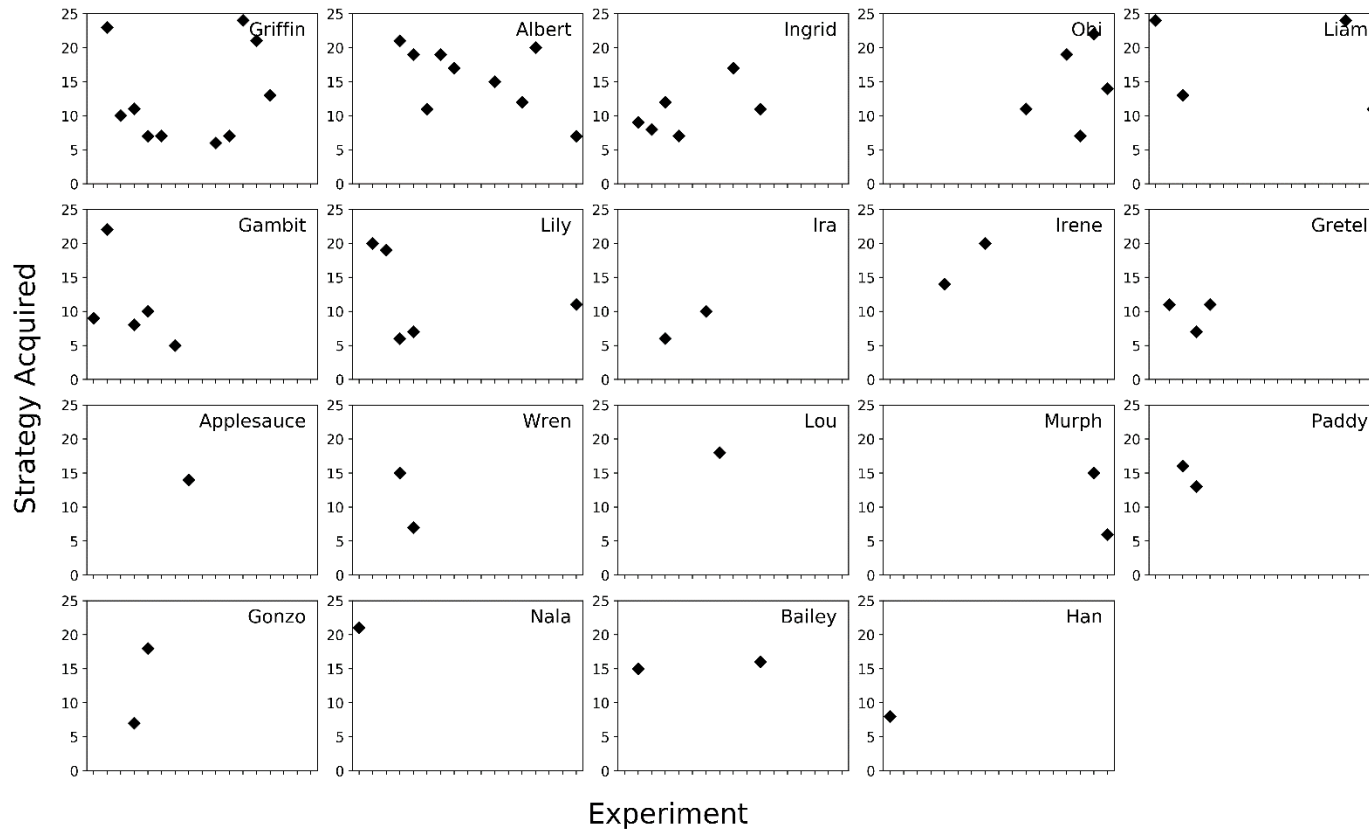
Figure 4.9 Strategy acquisition by experiment for each monkey. Black diamonds denote the block at which a majority of the highest likelihood strategies for the previous 5 blocks were the same, cue-based strategy. Experiments are depicted along the x-axis, with each tick one of the 1000 trial experiments (e.g., Experiment 1a, Experiment 1b, etc.). No obvious relationship exists between experiments and strategy acquisition. Monkeys did not use strategies earlier in experiments as they acquired more experience with the task.

Figure 4.10 Strategy acquisition by Experiment for all monkey. Black diamonds denote the block at which a majority of the highest likelihood strategies for the previous 5 blocks were the same, cue-based strategy for a monkey, in an experiment. These data are the same as in Figure 4.9, and many monkeys contributed multiple points to this plot by demonstrating strategy acquisition multiple times across the many experiments. Experiments are depicted along the x-axis, with each tick one of the 1000-block experiments. No obvious relationship exists among the points, such that animals did not acquire strategies more quickly across repeated exposures to the various weather predictions tasks.
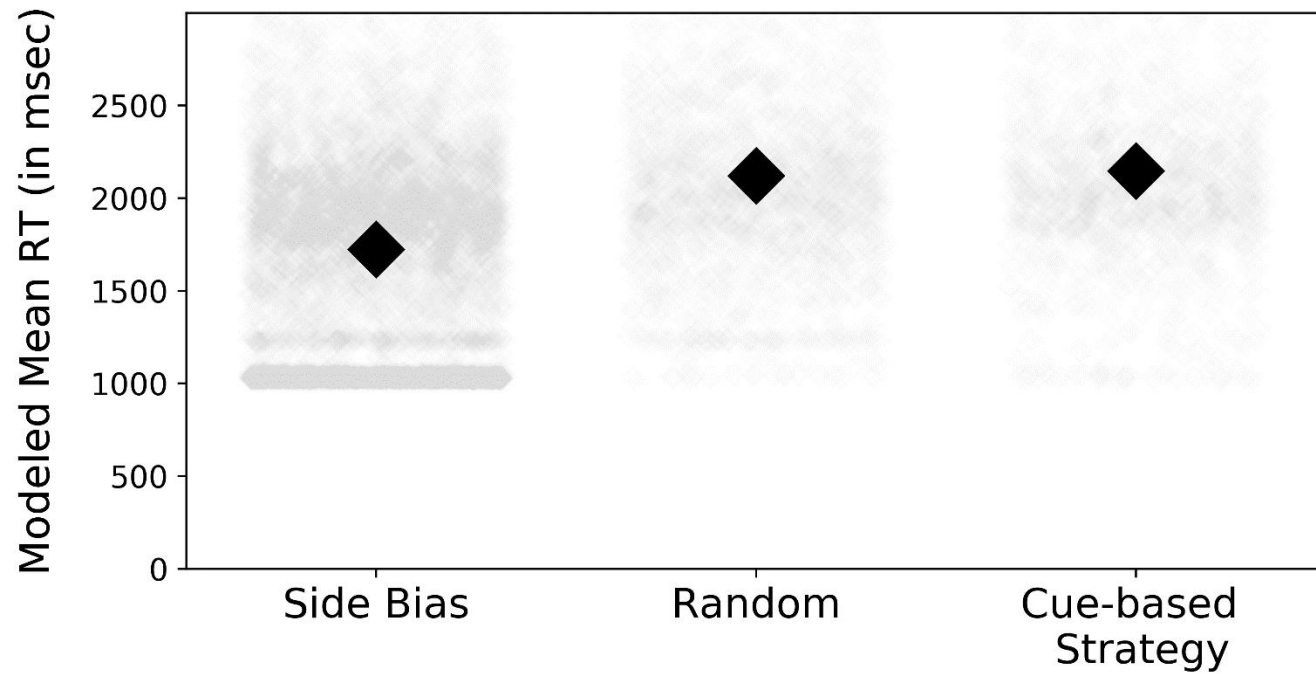
Figure 4.11 Central tendency of response latency distributions. Blocks were clustered as the product of side-biased, random, or cue-based strategies. The black diamond denotes the mean central tendency. Transparent grey diamonds are average monkey response latencies in a block inferred to be the product of the strategy group on the x-axis, included only as a reference to the distribution of the raw data. Random jitter along the x-axis was added to the grey diamonds to make a greater portion of the mean response latencies from blocked trials visible.
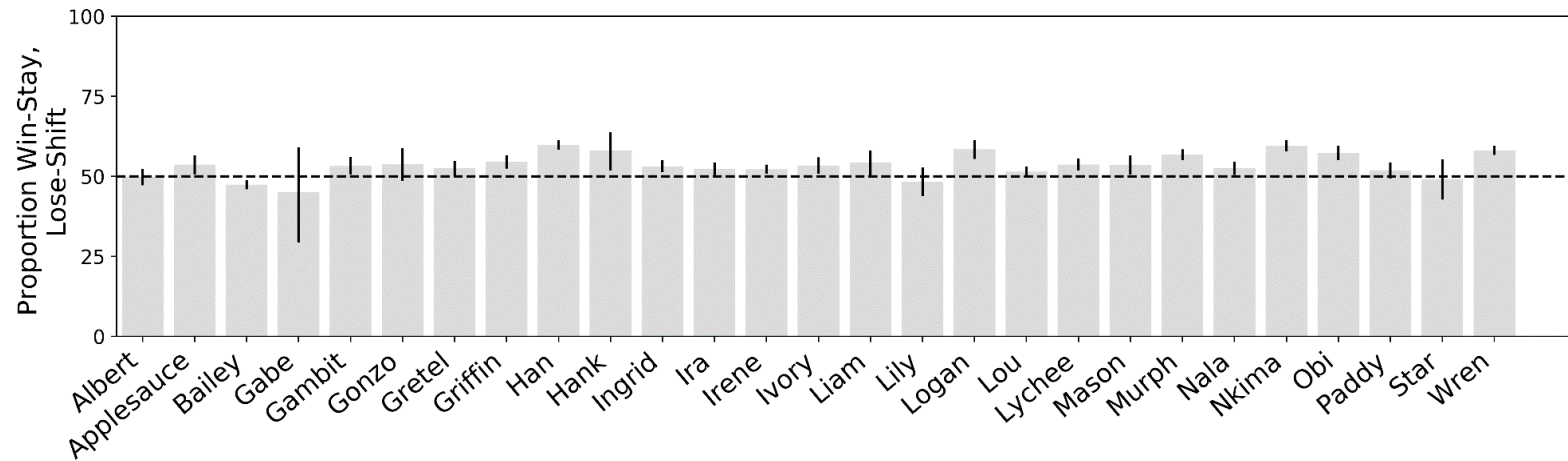
Figure 4.12 Win-stay or lose-shift strategy use. The height of each bar is the proportion of trials by each monkey in which a response in accordance with a win-stay or lose-shift strategy was observed in blocks of random strategy use. The dashed line indicates chance use of a win-stay or lose-shift strategy.

5 Discussion

This dissertation was conceived of as an attempt to identify the cognitive, perceptual, and motivational determinants of monkey behavior in a probabilistic categorization task. The unique challenges to learning posed by both probabilistic and categorization tasks made the effective sample size (i.e., the number of unique individuals that contributed informationally rich data) of this series of experiments lower than anticipated across all experiments. Despite the sample size, conclusions from this series of studies are clear. Evidence for each of the potential determinants of the monkeys' behavior will be evaluated and discussed in turn.

5.1 Cognitive Determinants

Hypothesis I: Strategy use during probabilistic categorization is a stable trait of an individual animal, and the same strategies used by an animal in one weather prediction task will be those used in another task.

Strategy use in probabilistic categorization tasks by individual monkeys was highly stable across blocks and experiments. Of critical concern to Hypothesis I, monkey strategy use did not at all resemble a series of random draws from the bank of specified strategies, or from any subset thereof with a length of more than one or two. Moreover, this was true to a similar extent for each monkey that demonstrated cue-based strategy use. No monkey, with the possible exception of Obi (rhesus) in Experiment 4, sampled widely among the available strategies within or across experiments. Hypothesis I was overall well-supported.

However, contra Hypothesis I, monkeys did not evidence the rich interindividual variability in clustered individual differences that would suggest that use of "trait" would be appropriate. Instead, the various plausible determinants of nonhuman primate strategy use (e.g., selective attention, associative learning ability) seemingly do little to avert species-wide convergence to the same few, highly similar strategies. These results predict that the same strategies used by a species in one weather prediction task are strongly predictive of strategy use by that species in every weather prediction task.

Hypothesis II: Behavior in weather prediction tasks is guided by factors other than an individual's preferred strategy, and titration of overall reward rate in the task will titrate strategy use accordingly. Specifically, the animals that used Random or Side-Bias strategies in Experiment 1 will use cue information to guide responding in Experiments 2a, 2b, and 2c.

In accordance with Hypothesis II, strategy use was sensitive to experimental manipulations to reward rate. However, the sensitivity was only demonstrated to affect adoption of a Single-Cue strategy. When many animals that had not previously demonstrated cue-based strategy use began using cue-based strategies at the most extreme reward rates tested, in Experiment 4, they were observed only to adopt those strategies that other individuals were already using. These manipulations had no demonstrable effect on the observed distribution of cue-based strategies, the complexity of strategy use, or the speed of strategy adoption.

Hypothesis III: Behavior in weather prediction tasks is guided by factors other than an individual's preferred strategy, and a set of more egalitarian cue probabilities will beget greater multicue responding in the animals that readily use strategies. Animals that do not readily use strategies will revert to Random or Side-biased strategies.

Hypothesis III was not supported. Experiments 2d and 2e presented animals with a set of probabilistic relationships among cues, responses, and outcomes in which the most highly predictive cues were less predictive than in other experiments, and the least predictive cues were more predictive than in other experiments. This manipulation did not cause more diverse, complex, or rapid strategy use.

Trends in the individual animals' strategy use seem to suggest that individual differences in learning rate could bear strongly on the observed rates of strategy adoption by the animals. Although the different animals largely arrived at use of the same strategy if they were going to use cues, how many trials were required before they used the cues reliably seemed to vary across individuals. Here, learning rate is used in the most rote associative and parametric of senses. Because such a learning rate is exclusively based on observation of behavior, it could be alternatively conceptualized not as reflecting the rate of internalization of cue information but instead as the rate of utilization of what information was learned (Speekenbrink et al., 2008). The source of any differences in learning rate could reasonably originate with differences in leveraging of attention, memory for what the animal did in previous trials, or any number of other, equivalently plausible cognitive constructs.

Consider the strategy analysis results from Experiment 2 for Ingrid (capuchin), Albert (capuchin), and Chewie (rhesus) (Figure 4.4). Ingrid reliably used Single-Cue

strategies in Experiment 2a and 2e from a very early block through the end of that experiment. Albert adopted use of the same Single-Cue strategy more fitfully, and nearer to the end of Experiments 2a, 2c, and 2d. Chewie did not evidence use of any strategy, but did reliably use Single-Cue strategies at higher block numbers in Whitham and Washburn (2020). Each of these observations fits neatly within an account of strategy use as dependent on a more global learning rate parameter. In such a conception, Ingrid learned one cue's predictive validity rapidly and utilized it into Single-Cue responding for the remainder of the experiment. Instantiation of such a parameter may also explain Ingrid's absent/decremented strategy use in other experiments, if what was internalized at an early trial was that no cue has predictive validity to either response. If Albert's hypothetical learning rate parameter was lower, cue information would be learned more slowly, and cue-based strategy use would be evident only late in experiments. If Chewie's learning rate was lower still, 1,000 trials of cue exposure could be insufficient for any demonstration of cue-based strategy use to emerge.

The apparent ceiling on the complexity of strategy use in probabilistic categorization may be underpinned by cognitive and associative mechanisms with a much longer history of experimental study. Conditioned blocking designs describe the remarkable degree to which cue information that does not contribute uniquely to a known cue/response relationship will not be acquired despite extensive opportunity for learning (Kamin, 1968; Kruschke, 2003). In such a design a participant first learns that cue A is strongly predictive of response 1 being rewarded, then is repeatedly exposed to cue A and cue B simultaneously, with response 1 still rewarded. Even though the participant experiences that cue B is also perfectly predictive of response 1 being rewarded, probe

trials with cue B reveal that participants do not behave as if cue B is perfectly predictive in this way.

A conception of probabilistic categorization in which utilization of single cues acts as a prerequisite for use of multiple cues, or even that learning of single cues is simply more likely than use of multiple cues, may inadvertently transform any weather prediction task into an approximation of a conditioned blocking design. In such a conception, learning of the first cue precludes learning of additional cues, even though the additional cues may be just as valid (and certainly contribute unique information relevant to optimal performance on the task). This synergy between observed strategy use and conditioned blocking designs is not perfect. Conditioned blocking designs typically require that the first cue learned is perfectly (or very highly) predictive of the response, and by definition no weather prediction task has this feature. That the first cue is not perfect, and thus does not perfectly block learning of other cues, may be what allows many humans to develop multicue responding. Across training, human participants may come to learn that the single cue-response relationship that they have learned is imperfect and "unblock" attention to other cues. However, animal learning is not prevented by initial learning in many other tasks in which animals spontaneously evidence new or superior cognition as performance matures. Moreover, in the best evidence for multicue strategy use among these experiments (Obi's Experiment 4), use of multiple cues was not preceded by use of less complex strategies. If, when animals do use multicue strategies, they do so right away, the initial conditioning phase of the blocking design would not occur. Although these notes prevent conditioned blocking from entirely explaining the observed pattern of results, any conditioned gating of attention to other cues may have contributed to the overall lack of use of multiple cues by the monkeys.

Information theory, more broadly, can present the problem of probabilistic categorization in similarly clarifying terms. The information yielded by a highly predictive cue was very high – offering an animal reward rates that were nearly at ceiling in most experiments from what would otherwise be the floor of chance responding. But once a valid cue was learned and utilized, learning a second, equally highly predictive cue offered the animal very little new information relevant to its decision-making. Although the superior reward rates yielded by the breadth of multicue strategies and the many patterns in which cue information might have appeared ensured that the informational and reward value of multicue learning was never literally absent, it may have been functionally so in many or all experiments.

It is also true that while experimenter and reviewer have omniscient insight into the structure of the task and the minor superiority of multicue strategy use, the monkeys likely lacked any such insight. To highly experimentally savvy animals, a task in which rewards are being dispensed at above-chance levels is solved. Moreover, successfully ignoring cue information that is not part of the selected, above-chance solution is a skill in many tasks that may feature irrelevant or incongruous cue information. Note that these considerations from information theory are important only to accounts in which strategic use of multiple cues develops incrementally from strategic use of single cues. Although this progression is what has been previously demonstrated in weather prediction tasks with humans, and was hypothesized in this study, it is not any exclusive way in which multiple items of cue information may be utilized. Presentation of weather prediction tasks to other species, or of weather prediction cues to primates in other experimental or naturalistic contexts, may reveal other routes through which strategic use of multiple cues develops.

The discrepancy between the cues with the highest predictive validity and the cues with lower predictive validity was intended, and is part of the challenge of situations in which multiple sources of information can be productively synthesized. To return briefly to examples presented in Chapter I, the metrics that might commonly be highly predictive of student success or scholarly quality are imperfect. Some students with high standardized test scores will fall short of predicted achievement, and some impressive citation metrics do not reflect incredible contributions to science. Other metrics might have predicted these discrepancies. Nuance in evaluating the relative validity of multiple competing sources of information, and facilitating use of multiple sources of information most effectively when a single source would seemingly suffice, is a very desirable set of human behaviors. The present results do little to provide direct recommendations for producing this desirable type of learning and responding. To the extent that the observed strategy use of the capuchin monkeys and rhesus macaques might model that of humans in these novel manipulations to the task, none of primary reinforcement manipulation, cue manipulation, practice effects, or naturally occurring individual differences in attention, memory, or other relevant cognitive competencies are seemingly fruitful predictors of multicue strategy use. A set of experiments with human participants that manipulated weather prediction tasks in the same ways as this dissertation would bear much more strongly on this more applied dimension of the study of probabilistic categorization.

Any direct mapping between the performance of monkeys and humans on the task should be made with extreme caution. Humans have evidenced markedly different, and markedly more multicue, strategy use than did the monkeys of this series of experiments. As reviewed in Chapter 1, some humans come to evidence some use of multicue strategies

in as few as 200 total trials. Differences in the framing of the task may explain part of this distinction, as in many probability evaluation tasks (Nickerson, 2004). Those tasks that frame the cues as symptoms and the responses as fictional diagnoses, or the cues as human features and the responses as preferred ice cream flavors, would seem to make clear to human participants in very explicit, very linguistically-embedded way that each cue offers something unique to prediction of the response. This distinction aside, this series of experiments gave the animals dozens of opportunities, and unique incentives, to do what humans do in mechanically identical tasks relatively spontaneously, yet the animals never did so. This result, alongside the demonstration by Newell and colleagues (2011) of the primary role of explicit cognition in probabilistic categorization and related comparative data reviewed by Hampton and colleagues (2020), suggests much more strongly that differences in ability to leverage explicit cognition drives the observed species differences in probabilistic categorization. Animals do learn to use multiple informative cues to guide their responses, as robust literatures on contextual cuing and cognitive control demonstrate exhaustively (e.g., Smith & Beran, 2018). It is in the domain of this dissertation, the rare (in published cognition literature) or common (in human and monkey ecology) situation in which cues conflict and underspecify the outcome, that such a demonstration remains elusive.

If these results reflect a fundamentally cognitive difference between humans and monkeys, the specific nature of that difference likely reflects a unique, qualitative difference in learning (Rumbaugh, 1971). Comparative cognition is the study of many such differences, qualitative and quantitative alike, yet so severe a distinction among these species in a task with seemingly little language engagement is more unusual. A design in which human participants completed a standard weather prediction task under

conditions of articulatory suppression (e.g., verbally repeating "1, 2, 3, 4, a, b, c, d") might reveal some latent language engagement in the task. That the task not only engages explicit learning faculties more than it does implicit learning faculties, but also goes further still to engage high-level use of language, would be an unreported dimension to the task. This would add additional color to previously reported relationships between language impairments and probabilistic categorization. Those reports were motivated by the ostensible mediation of both language impairment and the weather prediction task by implicit, procedural processes, but may instead reflect the direct influence of language on weather prediction performance.

These data and analyses facilitate only speculation about the source of individual differences in strategy adoption across monkeys. No obvious effects of animal age or sex were observed, but formal analyses of any effect of these parameters are precluded by the small sample sizes. The visualizations of the strategy analysis outputs (Figures 4.3-6) suggest a clear difference between the strategy use of capuchins and rhesus macaques in these experiments. That the capuchin monkeys were demonstrably superior in leveraging cognition of probable explicit nature is noteworthy, given that the species is often less adept with other tasks described in such terms (e.g., Smith et al., 2009, Smith et al., 2018). This superiority is largely relative, however; the capuchin monkeys did not evidence any rule-based cognition that they have not previously demonstrated in discrete categorization tasks (e.g., Smith et al., 2012) or the probabilistic categorization task of Whitham and Washburn (2020). Instead, it was the rhesus macaques that sharply diverged from species norms on categorization and decision-making tasks of this kind. Absent other behavioral evidence from the rhesus macaques, unanticipated tumult in their housing, uncharacteristically low-quality responses to this and other concurrent

computerized cognitive-behavioral tasks, and other unaccounted motivational factors (described below) likely drove the apparent species difference to a much greater degree than any purely cognitive factor.

5.2 **Perceptual Determinants**

Hypothesis IV: Utilization of cues is, in part, a perceptual process, and design considerations that have been demonstrated to engender multicue perception will engender multicue strategy use.

The data from Griffin (capuchin), Ingrid (capuchin), and Obi (rhesus) in Experiment 3a support Hypothesis IV. Presentation of all cue information as one, spatially-contiguous percept seemed to engender cue-based responding on the basis of more than one cue among a subset of animals that were already using single-cue strategies effectively. Further methodological considerations could do much to shrink the gulf between the strategy use of monkeys in this series of tasks and what is reported of humans. Multiple discrete cues (i.e., individual clipart images) were exclusively used in these and other weather prediction tasks purposefully. Multiple separable, and potentially opposing, cues must be synthesized in the ecologically relevant problems that the task purports to model. In the canonical example, of predicting rainy weather, one uses the discrete and separable cues of overhead cloud cover, the recent weather report, and whether passersby have raincoats to estimate whether rain is likely. Despite this deep embedding of use of discrete cues in the probabilistic categorization literature, it is possible that this format is simply incongruous with how monkey perception receives cue information. A weather prediction task in which the four cues were configured as

dimensional properties of a single stimulus might more faithfully detect their probabilistic categorization abilities without any corrupting influence of local precedence. For example, a single cue that varies in shape (e.g., rectangle vs trapezoid), orientation (e.g., upright vs horizontal), x-axis position onscreen (e.g., left vs right side of screen), and y-axis position onscreen (e.g., upper vs lower portion of screen) could represent each of the cue configurations of conventional weather prediction tasks. Such an arrangement is not without precedent in study of probabilistic categorization (e.g., Kruschke & Johansen, 1999), and may be uniquely well-suited for comparative study of monkeys' cognition.

It is possible that monkeys' predisposition to process perceptual information more locally did not cause the observed pattern of results in Experiment 3a. With only 13 total cue patterns, and therefore 13 unique percepts, the strategies used by these animals may be equally well-described as an "explicit memorization of unique category memberships" strategy. Animals are known to excel at learning how to map far greater numbers of unique stimuli to category memberships in ways that are not necessarily subject to the algorithmic cue weighting that probabilistic categorization ostensibly demands (e.g., Herrnstein et al., 1976). A requisite limitation of the manipulations of Experiment 3 was the lack of any mechanism with which to validate that the manipulation had functioned as intended. That is, although available literature suggested that attempting to subvert global/local biases or increase stimulus salience were worthwhile manipulations given the species and cognitions under study, no data from this study can estimate either the magnitude of global/local biases or the degree to which they were subverted (or whether order effects not captured by Experiment 1 went on to contribute to the observed strategy elaboration in Experiment 3a). Instead, these data provide compelling evidence that

further explorations of the role of primate perceptual norms will be important to any further exploration of probabilistic categorization.

Hypothesis V: Utilization of cues is, in part, a perceptual process, and cues that are qualitatively different than those of other experiments (socially-interesting rather than asocial; discrete rather than abstract; inherently salient rather than sterilely controlled) should engender qualitatively different strategy use.

Hypothesis V was not supported. Manipulation of neither the orientation nor the identity of the cue information caused strategy use to become more frequent or more complex. In contrast, those animals that completed previous experiments strategically often failed to do so in the same way in Experiment 3. Those animals that did not use cue-based strategies in previous experiments were not prompted to do so by the ostensibly perceptual manipulations of Experiment 3. The salience manipulations of Experiment 3c and 3d were perhaps particularly fraught. Sourcing images from public databases for use on spectrographically uncalibrated computer monitors ensured that the color and luminance of the images were only inconsistently faithful to the color and luminance information in the scene being depicted, and highly unfaithful to the species under study (Stevens et al., 2009). The highly artificial nature of those images likely made them much more akin to the clipart images of the other experiments than was originally intended (although see Talbot et al., 2016, for a demonstration of capuchin monkeys' ability to recognize and adaptively to use similarly formatted images). Any corrupting influence of these perceptual considerations was likely minor, but one of many contributors to lower than expected data quality. If feasible, any future exploration into this manipulation

should maximize the fidelity to real color and luminance of any digital images that are expected to function via their fidelity to real color and luminance.

The near-exclusive use of single cues among monkeys that used cue-based strategies is not without precedent in other perceptual tasks. In comparative perception, the same predisposition of monkeys to perceive local stimulus features rather than more global properties that motivated the manipulations of Experiments 3a and 3b can be readily conceptualized as a more modular form of Single-cue strategy use. The experimental evidence that was used to justify Experiment 3b, specifically, in which faces are sometimes found to be a class of stimuli that are processed more globally and holistically than other manner of stimuli, are party to their own flavor of nonhuman primates' single-cue superiority. In human gaze perception research, participants use both the head orientation and ratio of iris to visible sclera to make gaze judgments (Emery, 2000). In limited testing nonhuman primates have not been demonstrated to do so, and instead seem to ignore head direction in favor of the single cue of iris/sclera (Tomonaga & Imura, 2010).

5.3   Motivational Determinants

Of least theoretical interest to the study of probabilistic categorization were motivational determinants. In few situations would we expect complex cognition to increase as motivation decreases. Moreover, motivation may have been directly dependent on strategy use in the task: some of the animals that in Experiment 1 appeared most primed to demonstrate variable or complex strategy use in later experiments nearly universally failed to complete all later experiments.

Lower, but still high, levels of responding on the basis of no cue information were reported in the mechanically identical task of Whitham and Washburn (2020). Many animals in each study, including some animals that participated in both projects, seemingly evaluated the rate of pellet reward of side-biased or random responding as worthwhile. The manipulations to reward rate that were already included in the planned series of tasks reduced the rate of this behavior only inconsistently. Unfortunately, correction trials after incorrect responses could not be incorporated into the design (as recommended by Smith et al., 2012) without thwarting the meticulously controlled cue/response relationships on which weather prediction tasks are based.

The discrepancy between the rates of responding on the basis of no cue information in Whitham and Washburn (2020) and this series of tasks is more curious. Because 1,000 trials represented the extent of animals' experience with each experiment in each series of the present tasks, but represented "early learning" in the task of Whitham and Washburn, it is possible that this dissertation oversampled animals' earliest learning of a new task. Random or side-biased responses should be evident in these earliest trials, in the absence of any learning of cue information that could direct responding otherwise. However, those animals in both datasets that demonstrated use of cue based responding did so as early as in the first few blocks of a new experiment. A priori, few interindividual differences in learning rate within a species are of such magnitude as to require dozens-fold differences in exposure to cue information for equivalently-complex responses to be made. This suggests even more strongly that the disparity in cue-based strategy use is one downstream from motivational factors of the monkeys and unquantified factors of their testing. Many of the same animals, tested two years later on a mechanically identical set

of tasks, demonstrated markedly different appetites for using their cognitive and perceptual powers to bolster their reward rate in the task.

It is difficult, but not impossible, for future work to account better for these considerations. Because Experiment 4 demonstrated that (for some monkeys) there is some reward rate at which responses that are not based on cue information will give way to cue-based responding, it is perhaps reasonable to port these more extreme ITI and timeout durations to every experiment, for every monkey. However, just as the same parameter values of Whitham and Washburn (2020) were not sufficient to encourage the same degree of strategic responding in this dissertation, there is little reason to believe that any combination of parameter values will suit different animals (or the same animal at two different times) equally well. This challenge is common throughout comparative psychological study of animals with as many individual differences (and as rich a nontesting life, complete with variable diet, hormonal cycles, social forces, etc.) as nonhuman primates. The specific questions of this dissertation, ones of experimental manipulations' effect on individual monkeys' ability and inclination to deploy strategies of heretofore undocumented complexity, make the moving target of subjective motivation especially relevant. The design of this dissertation, in which reinforcement factors were rigorously controlled for but subjective motivation seemingly varied in uncontrolled and unmeasured fashion, leaves open the possibility that monkey would not demonstrate multicue strategy use, not necessarily that they could not. Evidence from Experiment 4 falls short of justifying this last point, as the most extreme adjustments to reward rate only attenuated the most motoric, least cognitive of response strategies; they did not elicit any unique strategy use.

5.4 General Discussion

This dissertation builds on the work of Whitham and Washburn (2020) to suggest much more strongly that monkeys are Single-Cue strategy users to an exceptional degree. Primates have proven to be accomplished rule-users in a variety of previous categorization tasks, dating to the earliest laboratory-based descriptions of rhesus monkey behavior (Hoffman et al., 2009; Kinnaman, 1902). That they were repeatedly demonstrated to lean heavily on simple rules when relationships among cues and responses were complex, probabilistic, and, perhaps, perceptually difficult to discern, fits neatly alongside existing scholarship in comparative and cognitive psychologies.

The niche modeling used to produce this result is but one of many published procedures for analyzing probabilistic categorization data. The models largely produce similar results, given the same inputs, and the explanatory virtues that drove selection of the strategy analysis perspective employed in the preceding analyses held true. That said, every model is imperfect, and other parameters and models might offer additional insights into this design and others of its kind.

In the few cases in which strategic use of multiple cues was identified by the strategy analysis above, analyzing blocks of greater than or fewer than 40 trials might have been clarifying. Use of 40 trial blocks was informed but arbitrary. An analysis of smaller block sizes would be better able to detect strategies that are changing rapidly, whereas an analysis of larger block sizes would provide the model with better information with which to identify strategy use. An extension of the strategy analysis presented in this manuscript might win both of these advantages simultaneously. Many strategy analyses could be run with the same data, but variable block sizes. For example, the same data could be analyzed three times with blocks made up of 25, 50, and 100 trials. Moreover,

the same analysis could use each of 25, 50, and 100 trial blocks in the same analysis (as actual strategy use surely changes much more irregularly than once every n trials). The analyst could then report the extent to which these analyses agree, and the distribution of model fits across the individual strategy analyses.

Two parameters in particular, learning rate and previous strategy use, would seem to be obvious candidates for inclusion in future modeling. A virtue of the analyses reported above is that they are memoryless – the strategy identification for a given block is not dependent on any previous or future blocks. For questions that are fundamentally focused on description, rather than on prediction of future trends in a specific animal's behavior, this feature maintains model sensitivity to what was often a dynamic shift from cueless to single-cue responding. The results presented in Chapter 4 make clear, however, that strategy identifications for consecutive blocks are very highly autocorrelated (Figure 4.7). This autocorrelation, known to the experimenter but not to the model, would ideally be incorporated into the modeling process as an additional predictor of strategy use in a block. Such an incorporation would do much to sharpen strategy identifications above the rates of correct identification depicted in Figure 4.1.

Bayesian model selection, psychologist adoption and interpretation of Bayesian techniques, and the tools for Bayesian mixture model specification and fitting have each advanced precipitously in the previous decade. Many Bayesian techniques would incorporate information from serialized blocks naturally via specification of a prior probability that is then updated with observed data (Kruschke & Liddell, 2018b). As an ancillary benefit, estimation of strategy adherence rates would be incorporated into the model by default and provide an additional dependent measure on which hypotheses about strategy use and individual differences might be tested. Constraints would be

required to ensure that the model retains sensitivity to new data, and does not dismiss strategic responding because previous blocks were not strategic. Without these constraints, the model might get "stuck" with a strategy that was highly evidenced in early trials but was an inappropriate depiction of mature responding. With the proper constraints in place, an evolved, Bayesian version of the present strategy analysis (akin to the Bayesian strategy analysis of Lee, 2016, but on a much larger scale), could do much to sharpen detection of strategy use in weather prediction and offer additional description of its deployment.

A principal advantage to the strategy analysis technique used above is that it can test an arbitrary number of strategy specifications against each other. Following Meeter and colleagues (2006), the strategies specified above ostensibly included all those that may reasonably have been employed by human participants. It remains possible, however, that another specification might better capture monkeys' (or humans') behaviors in the task than those represented in Table 4.1. Because of this reality, as well as these results' apparent suitability for future synthesis with allied data on individual differences, the rich potential for alternative analysis strategies (Bayesian or otherwise), and the more far-reaching virtues of transparent science, all data and code to run analyses are publicly available in Open Science Foundation's repositories.

The profound, consistent null result of single cue strategy use is not sufficient to discriminate between more cognitive sources of the strong predisposition to single cues or more perceptual ones. To both the researcher interested in why monkey strategy use is not more like that of humans, and the allied researcher interested in what drives the complexity of human strategy use, a design that can make this distinction remains a key goal of future research. In much of the scholarship of comparative cognition, individual

animals, and primates specifically, are exemplary in the most literal of senses. Each individual evidences a unique suite of cognitive competencies and motivational dispositions that are dictated in part by their species, but often to an equal or greater degree by their other traits. In the series of experiments that comprises this dissertation this individuality may have been true in the motivational sense, but, uncharacteristically, not in the cognitive competencies evidenced by the animals. In the specific set of cognitions required to complete weather prediction tasks monkeys have, to this point, demonstrated that they are impressively of a kind. This result has held in spite of a diversity of task manipulations of theoretical or hypothetical import. Such a monkey monolith has few parallels in comparative psychology, and represents a fertile ground for continued exploration of what procedural determinants and species differences drive this unusual invariance.

REFERENCES

Abbas, A., Belloy, M., Kashyap, A., Billings, J., Nezafati, M., Schumacher, E. H., & Keilholz, S. (2019). Quasi-periodic patterns contribute to functional connectivity in the brain. *Neuroimage*, *191*, 193-204.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716-723.

Ashby, F. G., Alfonso-Reese, L. A., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*, 442.

Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences*, *5*, 204-210.

Ashby, F. G., & Ennis, J. M. (2006). The role of the basal ganglia in category learning. *Psychology of Learning and Motivation*, *46*, 1-36.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, *56*, 149-178.

Babyak, M. A. (2004). What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine*, *66*, 411-421.

Barsalou, L. W. (1992). Cognitive psychology: An overview for cognitive scientists. Hillsdale, NJ: Erlbaum.

Bartlema, A., Lee, M., Wetzels, R., & Vanpaemel, W. (2014). A Bayesian hierarchical mixture approach to individual differences: Case studies in selective attention and representation in category learning. *Journal of Mathematical Psychology*, *59*, 132-150.

Basile, B. M., Hampton, R. R., Suomi, S. J., & Murray, E. A. (2009). An assessment of memory awareness in tufted capuchin monkeys (*Cebus apella*). *Animal Cognition, 12*, 169 –180.

Basile, B. M., Moylan, E. J., Charles, D. P., & Murray, E. A. (2015). Two-item same/different discrimination in rhesus monkeys (*Macaca mulatta*). *Animal Cognition*, *18*, 1221-1230.

Beran, M. J., Decker, S., Schwartz, A., & Smith, J. D. (2012). Uncertainty monitoring by young children in a computerized task. *Scientifica*, *2012*.

Beran, M. J., Perdue, B. M., Futch, S. E., Smith, J. D., Evans, T. A., & Parrish, A. E. (2015). Go when you know: Chimpanzees' confidence movements reflect their responses in a computerized memory task. *Cognition*, *142*, 236-246.

Beran, M. J., Perdue, B. M., Rossettie, M. S., James, B. T., Whitham, W., Walker, B., ... & Parrish, A. E. (2016). Self-control assessments of capuchin monkeys with the rotating tray task and the accumulation task. *Behavioural Processes, 129*, 68-79.

Beran, M. J., Perdue, B. M., & Smith, J. D. (2014). What are my chances? Closing the gap in uncertainty monitoring between rhesus monkeys (*Macaca mulatta*) and capuchin monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Learning and Cognition*, *40*, 303-316.

Beran, M. J., & Smith, J. D. (2011). Information seeking by rhesus monkeys (*Macaca mulatta*) and capuchin monkeys (*Cebus apella*). *Cognition*, *120*, 90-105.

Blanchard, T. C., & Hayden, B. Y. (2015). Monkeys are more patient in a foraging task than in a standard intertemporal choice task. *PloS One*, *10*, e0117057.

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

Carleton, R. N., Norton, M. P. J., & Asmundson, G. J. (2007). Fearing the unknown: A short version of the Intolerance of Uncertainty Scale. *Journal of Anxiety Disorders*, *21*, 105-117.

Church, B. A., Valdez, G. E., Boomer, J., & Smith, J. D. (2018). Implicit and explicit category learning: Independence or competition? Poster presentation at the 59th Annual Meeting of the Psychonomic Society, New Orleans, LA.

Cook, R. G., Brown, M. F., & Riley, D. A. (1985). Flexible memory processing by rats: use of prospective and retrospective information in the radial maze. *Journal of Experimental Psychology: Animal Behavior Processes*, *11*, 453-469.

Cooksey, R. W. (1996). The methodology of social judgement theory. *Thinking & Reasoning, 2*(2-3), 141-174.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12, 671–684.

Crupi, A. (2019, January 3). *Network TV can't survive without the NFL*. Ad Age. https://adage.com/article/media/top-50-u-s-broadcasts-2018/316102

Deaner, R. O., Isler, K., Burkart, J., & Van Schaik, C. (2007). Overall brain size, and not encephalization quotient, best predicts cognitive ability across non-human primates. *Brain, Behavior and Evolution*, *70*, 115-124.

Dulany, D. E., & Hilton, D. J. (1991). Conversational implicature, conscious representation, and the conjunction fallacy. *Social Cognition*, *9*, 85-110.

Dunbar, R. I., & Shultz, S. (2007). Understanding primate brain evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*, 649-658.

Duncan, K., Doll, B. B., Daw, N. D., & Shohamy, D. (2018). More than the sum of its parts: a role for the hippocampus in configural reinforcement learning. *Neuron*, *98*, 645-657.

Dutilh, G., Annis, J., Brown, S. D., Cassey, P., Evans, N. J., Grasman, R. P., ... & Kupitz, C. N. (2019). The quality of response time data inference: A blinded, collaborative assessment of the validity of cognitive models. *Psychonomic Bulletin & Review,* *26*, 1051-1069.

Eldridge, L. L., Masterman, D., & Knowlton, B. J. (2002). Intact implicit habit learning in Alzheimer's disease. *Behavioral Neuroscience*, *116*, 722-726.

Eling, P., Derckx, K., & Maes, R. (2008). On the historical and conceptual background of the Wisconsin Card Sorting Test. *Brain and Cognition*, *67*, 247-253.

Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews, 24*, 581-604.

Etz, A., & Vandekerckhove, J. (2018). Introduction to Bayesian inference for psychology. *Psychonomic Bulletin & Review*, *25*, 5-34.

Evans, T. A., Beran, M. J., Chan, B., Klein, E. D., & Menzel, C. R. (2008). An efficient computerized testing method for the capuchin monkey (*Cebus apella*): Adaptation of the LRC-CTS to a socially housed nonhuman primate species. *Behavior Research Methods*, *40*, 590-596.

Evans, T. A., Perdue, B. M., Parrish, A. E., & Beran, M. J. (2014). Working and waiting for better rewards: Self-control in two monkey species (*Cebus apella* and *Macaca mulatta*). *Behavioural Processes*, *103*, 236-242.

Finn, A. S., Kalra, P. B., Goetz, C., Leonard, J. A., Sheridan, M. A., & Gabrieli, J. D. (2016). Developmental dissociation between the maturation of procedural memory and declarative memory. *Journal of Experimental Child Psychology*, *142*, 212-220.

Flemming, T. M., Thompson, R. K., Beran, M. J., & Washburn, D. A. (2011). Analogical reasoning and the differential outcome effect: Transitory bridging of the conceptual gap for rhesus monkeys (*Macaca mulatta*). *Journal of Experimental Psychology: Animal Behavior Processes*, *37*, 353-360.

Foote, A. L., & Crystal, J. D. (2007). Metacognition in the rat. *Current Biology*, *17*, 551-555.

Fragaszy, D., Johnson-Pynn, J., Hirsh, E., & Brakke, K. (2003). Strategic navigation of two-dimensional alley mazes: comparing capuchin monkeys and chimpanzees. *Animal Cognition*, *6*, 149-160.

Fragaszy, D. M., Visalberghi, E., & Fedigan, L. M. (2004). *The complete capuchin: The biology of the genus Cebus*. Cambridge University Press.

Gabay, Y., Vakil, E., Schiff, R., & Holt, L. L. (2015). Probabilistic category learning in developmental dyslexia: Evidence from feedback and paired-associate weather prediction tasks. *Neuropsychology*, *29*, 844-854.

Gallistel, C. R., & Gibbon, J. (2000). Time, rate, and conditioning. *Psychological Review*, *107*, 289-344.

Garner, W. R. (1962). *Uncertainty and Structure as Psychological Concepts*. Wiley.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases". *European Review of Social Psychology, 2*, 83-115.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), *Subjective Probability* (pp. 129-161). Wiley.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review, 103*, 592-596.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*, 650-669.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*, 506-528.

Giner-Sorolila, R., & Chaiken, S. (1997). Selective use of heuristic and systematic processing under defense motivation. *Personality and Social Psychology Bulletin, 23*, 84-97.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: an adaptive network model. *Journal of Experimental Psychology: General, 117*, 227-247.

Gluck, M. A., Shohamy, D., & Myers, C. (2002). How do people solve the "weather prediction" task?: Individual variability in strategies for probabilistic category learning. *Learning & Memory, 9*, 408-418.

Goodhart, C. A. (1984). Problems of monetary management: the UK experience. In C. A. Goodhart (Ed), *Monetary Theory and Practice* (pp. 91-121). Palgrave, London.

Grayson, D. (2004). Some myths and legends in quantitative psychology. *Understanding Statistics, 3*, 101-134.

Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, *100*, 253-258.

Hampton, R. R., Engelberg, J. W. M., & Brady, R. J. (2020). Explicit memory and cognition in monkeys. *Neuropsychologia*, 107326.

Herbranson, W. T., Fremouw, T., & Shimp, C. P. (1999). The randomization procedure in the study of categorization of multidimensional stimuli by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *25*, 113-135.

Herrnstein, R. J., Loveland, D. H., & Cable, C. (1976). Natural concepts in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *2*, 285-302.

Hertwig, R., & Gigerenzer, G. (1999). The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, *12*, 275-305.

Hilbig, B.E., & Moshagen, M. (2014). Generalized outcome-based strategy classification: Comparing deterministic and probabilistic choice models. *Psychonomic Bulletin & Review*, *21*, 1431– 1443.

Hoffman, M. L., Beran, M. J., & Washburn, D. A. (2009). Memory for "what","where", and "when" information in rhesus monkeys (*Macaca mulatta*). *Journal of Experimental Psychology: Animal Behavior Processes, 35*, 143.

Hopkins, W. D., & Washburn, D. A. (2002). Matching visual stimuli on the basis of global and local features by chimpanzees (*Pan troglodytes*) and rhesus monkeys (*Macaca mulatta*). *Animal Cognition*, 5, 27–31.

Jack, A. I., Dawson, A. J., Begany, K. L., Leckie, R. L., Barry, K. P., Ciccia, A. H., & Snyder, A. Z. (2013). fMRI reveals reciprocal inhibition between social and physical cognitive domains. *NeuroImage*, *66*, 385-401.

Jaworska, N., & Chupetlovska-Anastasova, A. (2009). A review of multidimensional scaling (MDS) and its utility in various psychological domains. *Tutorials in Quantitative Methods for Psychology*, *5*, 1-10.

Lee, M. D. (2016). Bayesian outcome-based strategy classification. *Behavior Research Methods*, *48*, 29-41.

Lee, M. D., & Newell, B. R. (2011). Using hierarchical Bayesian methods to examine the tools of decision-making. *Judgment and Decision-Making, 6*, 832-842.

Lindburg, D. G. (1971). The rhesus monkey in North India: An ecological and behavioral study. *Primate Behavior: Developments in Field and Laboratory Research*, *2*, 1-106.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds), *Heuristics and Biases: The Psychology of Intuitive Judgment*, (pp. 49-81). Cambridge University Press.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, 103, 582-591.

Kamin, L.J. (1968). 'Attention-like' processes in classical conditioning. In M.R. Jones (Ed.), *Miami Symposium on the Prediction of Behavior: Aversive Stimulation* (pp. 9–33). Coral Gables, FL: University of Miami Press.

Kangas, B. D., Bergman, J., & Coyle, J. T. (2016). Touchscreen assays of learning, response inhibition, and motivation in the marmoset (*Callithrix jacchus*). *Animal Cognition*, *19*, 673-677.

Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science*, *4*, 533-550.

Kinnaman, A. J. (1902). Mental life of two Macacus rhesus monkeys in captivity. II. *The American Journal of Psychology, 13*, 173-218.

Klein, E. D., Evans, T. A., Schultz, N. B., & Beran, M. J. (2013). Learning how to "make a deal": Human (*Homo sapiens*) and monkey (*Macaca mulatta*) performance when repeatedly faced with the Monty Hall Dilemma. *Journal of Comparative Psychology*, *127*, 103-108.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*(5280), 1399-1402.

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory*, *1*, 106-120.

Krishnan, S., Watkins, K. E., & Bishop, D. V. (2016). Neurobiological basis of language learning difficulties. *Trends in Cognitive Sciences*, *20*, 701-714.

Kruschke, J. K. (2003). Attention in learning. *Current Directions in Psychological Science, 12*, 171-175.

Kruschke, J. K. (2005). Category learning. *The Handbook of Cognition*, *7*, 183-201.

Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in Methods and Practices in Psychological Science, 1*, 270-280.

Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1083.

Kruschke, J. K., & Liddell, T. M. (2018a). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review, 25,* 178-206.

Kruschke, J. K., & Liddell, T. M. (2018b). Bayesian data analysis for newcomers. *Psychonomic Bulletin & Review, 25,* 155-177.

Lee, M. D. (2016). Bayesian outcome-based strategy classification. *Behavior Research Methods, 48,* 29-41.

Lee, M. D., & Newell, B. R. (2011). Using hierarchical Bayesian methods to examine the tools of decision-making. *Judgment and Decision Making, 6,* 832-842.

Leopold, D. A., & Rhodes, G. (2010). A comparative view of face perception. *Journal of Comparative Psychology, 124,* 233-251.

Levine, M. (1959). A model of hypothesis behavior in discrimination learning set. *Psychological Review, 66,* 353–366.

Little, D. R., Lewandowsky, S., & Heit, E. (2006). Ad hoc category restructuring. *Memory & Cognition, 34,* 1398-1413.

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology, 6.*

Malassis, R., Gheusi, G., & Fagot, J. (2015). Assessment of metacognitive monitoring and control in baboons (*Papio papio*). *Animal Cognition, 18,* 1347-1362.

Marsh, H. L., & MacDonald, S. E. (2012). Information seeking by orangutans: A generalized search strategy?. *Animal Cognition, 15,* 293-304.

Mazur, J. E., & Kahlbaugh, P. E. (2012). Choice behavior of pigeons (*Columba livia*), college students, and preschool children (*Homo sapiens*) in the Monty Hall dilemma. *Journal of Comparative Psychology*, *126*, 407-420.

McElreath, R. (2020). Statistical rethinking: A Bayesian course with examples in R and Stan. CRC press.

McKinney, W. (2011). pandas: a foundational Python library for data analysis and statistics. *Python for High Performance and Scientific Computing, 14.*

Meeter, M., Myers, C. E., Shohamy, D., Hopkins, R. O., & Gluck, M. A. (2006). Strategies in probabilistic categorization: Results from a new way of analyzing performance. *Learning & Memory*, *13*, 230-239.

Meeter, M., Radics, G., Myers, C. E., Gluck, M. A., & Hopkins, R. O. (2008). Probabilistic categorization: How do normal participants and amnesic patients do it? *Neuroscience & Biobehavioral Reviews*, *32*, 237-248.

Melnikoff, D. E., & Bargh, J. A. (2018). The insidious number two. *Trends in Cognitive Sciences*, *22*, 668-669.

Newell, B. R., Dunn, J. C., & Kalish, M. (2011). Systems of category learning: Fact or fantasy? In B. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 54, pp. 167-215). Academic Press.

Newell, B. R., Lagnado, D. A., & Shanks, D. R. (2007). Challenging the role of implicit processes in probabilistic category learning. *Psychonomic Bulletin & Review*, *14*, 505-511.

Nickerson, R. S. (2004). *Cognition and chance: The psychology of probabilistic reasoning*. Psychology Press.

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology*, *43*, 25-53.

Palmeri, T. J., & Flanery, M. A. (1999). Learning about categories in the absence of training: Profound amnesia and the relationship between perceptual categorization and recognition memory. *Psychological Science*, *10*, 526-530.

Parrish, A. E., Brosnan, S. F., & Beran, M. J. (2015). Do you see what I see? A comparative investigation of the Delboeuf illusion in humans (*Homo sapiens*), rhesus monkeys (*Macaca mulatta*), and capuchin monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Learning and Cognition*, *41*, 395-405.

Parrish, A. E., Brosnan, S. F., Wilson, B. J., & Beran, M. J. (2014). Differential responding by rhesus monkeys (*Macaca mulatta*) and humans (*Homo sapiens*) to variable outcomes in the assurance game. *Animal Behavior and Cognition*, *1*, 215-229.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*, 353-363.

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83*, 304-308.

Pothos, E. M., & Busemeyer, J. R. (2009). A quantum probability explanation for violations of 'rational' decision theory. *Proceedings of the Royal Society B: Biological Sciences*, *276*, 2171-2178.

Qadri, M. A., Ashby, F. G., Smith, J. D., & Cook, R. G. (2019). Testing analogical rule transfer in pigeons (*Columba livia*). *Cognition*, *183*, 256-268.

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

Richardson, W. K., Washburn, D. A., Hopkins, W. D., Savage-Rumbaugh, E. S., & Rumbaugh, D. M. (1990). The NASA/LRC computerized test system. *Behavior Research Methods, Instruments, & Computers*, *22*, 127-131.

Ross, C. T., & Winterhalder, B. (2015). Sit-and-wait versus active-search hunting: A behavioral ecological model of optimal search mode. *Journal of Theoretical Biology*, *387*, 76-87.

Rumbaugh, D. M. (1971). Evidence of qualitative differences in learning processes among primates. *Journal of Comparative and Physiological Psychology*, *76*, 250-255.

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science, 2*, e55.

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin, 134*, 207-222.

Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making, 15*, 233-250.

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, *22*, 325-345.

Shepard, R. N. (1958). Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology*, *55*, 509–523.

Shohamy, D., Myers, C. E., Kalanithi, J., & Gluck, M. A. (2008). Basal ganglia and dopamine contributions to probabilistic category learning. *Neuroscience & Biobehavioral Reviews*, *32*, 219-236.

Sidorkin, A. M. (2016). Campbell's law and the ethics of immensurability. *Studies in Philosophy and Education*, *35*, 321-332.

Smith, J. D. (2014). Prototypes, exemplars, and the natural history of categorization. *Psychonomic Bulletin & Review*, *21*, 312-331.

Smith, J. D., Beran, M. J., Couchman, J. J., Coutinho, M. V., & Boomer, J. B. (2009). The curious incident of the capuchins. *Comparative Cognition and Behavior Reviews, 4*, 61–64.

Smith, J. D., Berg, M. E., Cook, R. G., Murphy, M. S., Crossley, M. J., Boomer, J., ... & Grace, R. C. (2012). Implicit and explicit categorization: A tale of four species. *Neuroscience & Biobehavioral Reviews*, *36*, 2355-2369.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1411-1436.

Smith, J. D., Redford, J. S., Beran, M. J., & Washburn, D. A. (2010). Rhesus monkeys (*Macaca mulatta*) adaptively monitor uncertainty while multi-tasking. *Animal Cognition*, *13*, 93-101.

Smith, J. D., Zakrzewski, A. C., & Church, B. A. (2016). Formal models in animal-metacognition research: the problem of interpreting animals' behavior. *Psychonomic Bulletin & Review*, *23*, 1341-1353.

Speekenbrink, M., Channon, S., & Shanks, D. R. (2008). Learning strategies in amnesia. *Neuroscience & Biobehavioral Reviews*, *32*, 292-310.

Spinozzi, G., De Lillo, C., & Truppa, V. (2003). Global and local processing of hierarchical visual stimuli in tufted capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology*, *117*, 15-23.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*, 645-665.

Stevens, M., Stoddard, M. C., & Higham, J. P. (2009). Studying primate color: Towards visual system-dependent methods. *International Journal of Primatology, 30*, 893-917.

Talbot, C. F., Leverett, K. L., & Brosnan, S. F. (2016). Capuchins recognize familiar faces. *Animal Behaviour, 122*, 37-45.

Taubert, J., Wardle, S. G., Flessert, M., Leopold, D. A., & Ungerleider, L. G. (2017). Face pareidolia in the rhesus monkey. *Current Biology, 27*, 2505-2509.

Thompson, V. A., Evans, J. S. B., & Campbell, J. I. (2013). Matching bias on the selection task: It's fast and feels good. *Thinking & Reasoning, 19*, 431-452.

Todd, P. M., & Gigerenzer, G. (2000). Précis of simple heuristics that make us smart. *Behavioral and Brain Sciences*, *23*, 727-741.

Tomonaga, M., & Imura, T. (2010). Visual search for human gaze direction by a chimpanzee (*Pan troglodytes*). *PloS one, 5*.

Trueblood, J. S., & Busemeyer, J. R. (2011). A quantum probability account of order effects in inference. *Cognitive Science*, *35*, 1518-1552.

Trueblood, J. S., Yearsley, J. M., & Pothos, E. M. (2017). A quantum probability framework for human probabilistic inference. *Journal of Experimental Psychology: General*, *146*, 1307.

Tu, H. W., Pani, A. A., & Hampton, R. R. (2015). Rhesus monkeys (*Macaca mulatta*) adaptively adjust information seeking in response to information accumulated. *Journal of Comparative Psychology*, *129*, 347-355.

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

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, *41*, 399-433.

Washburn, D. A. (1992). Analyzing the path of responding in maze-solving and other tasks. *Behavior Research Methods, Instruments, & Computers*, *24*, 248-252.

Washburn, D. A., & Rumbaugh, D. M. (1992). Testing primates with joystick-based automated apparatus: Lessons from the Language Research Center's Computerized Test System. *Behavior Research Methods, Instruments, & Computers*, *24*, 157-164.

Washburn, D. A., Smith, J. D., & Taglialatela, L. A. (2005). Individual differences in metacognitive responsiveness: Cognitive and personality correlates. *The Journal of General Psychology*, *132*, 446-461.

West, G., Vadillo, M. A., Shanks, D. R., & Hulme, C. (2018). The procedural learning deficit hypothesis of language learning disorders: we see some problems. *Developmental Science*, *21*, e12552.

Whitham, W. & Washburn, D. A. (2020). Strategy use in probabilistic categorization by rhesus macaques (*Macaca mulatta*) and Capuchin monkeys (*Cebus [Sapajus] apella*). *Journal of Comparative Psychology*.

Wills, A. J., & Pothos, E. M. (2012). On the adequacy of current empirical evaluations of formal models of categorization. *Psychological Bulletin*, *138*, 102-125.

Wilson, B., Slater, H., Kikuchi, Y., Milne, A. E., Marslen-Wilson, W. D., Smith, K., & Petkov, C. I. (2013). Auditory artificial grammar learning in macaque and marmoset monkeys. *Journal of Neuroscience*, *33*, 18825-18835.

Wilson, B., Smith, K., & Petkov, C. I. (2015). Mixed-complexity artificial grammar learning in humans and macaque monkeys: Evaluating learning strategies. *European Journal of Neuroscience*, *41*, 568-578.