

7-15-2016

Predicting Choices in Bumblebees (*Bombus impatiens*): Learning Rules and The Two-Armed Bandit

Isabel Lucia Rojas-Ferrer

University of Missouri-St. Louis, ilrdm6@mail.ums.edu

Follow this and additional works at: <http://irl.ums.edu/thesis>

Recommended Citation

Rojas-Ferrer, Isabel Lucia, "Predicting Choices in Bumblebees (*Bombus impatiens*): Learning Rules and The Two-Armed Bandit" (2016). *Theses*. 43.
<http://irl.ums.edu/thesis/43>

This Thesis is brought to you for free and open access by the Graduate Works at IRL @ UMSL. It has been accepted for inclusion in Theses by an authorized administrator of IRL @ UMSL. For more information, please contact marvinh@ums.edu.

PREDICTING CHOICES IN BUMBLEBEES (*BOMBUS IMPATIENS*):
LEARNING RULES AND THE TWO ARMED BANDIT

Isabel Rojas-Ferrer

B.S. Industrial Microbiology, University of Puerto Rico-Mayagüez, 2014

A Thesis Submitted to The Graduate School at the University of Missouri- St. Louis in partial
fulfillment of the requirements for the degree
Master of Science in Biology with emphasis in Ecology, Evolution and Systematics

August 2016

Advisory Committee

Aimee Dunlap, Ph.D.
Chairperson

Patricia Parker, Ph.D.

Zuleyma Tang-Martinez, Ph.D.

Yuefeng Wu, Ph.D.

Dedication

I dedicate this thesis to my parents, but above all to the loving memory of my grandmother, María Mercedes Ramírez-Rojas who showed me the value of strength, perseverance and intellect.

Table of Contents

Abstract.....	1
Chapter 1.....	2
Foraging and Choice Mechanisms.....	3
Information Source.....	5
Sampling and The Two Armed Bandit.....	8
Rules of Thumb.....	11
Learning and Memory.....	13
Learning Rules.....	15
Bayes' Theorem.....	16
Linear Operator.....	17
Relative Payoff Sum.....	19
Perfect Memory.....	19
Chapter 2.....	21
Introduction.....	22
Methods.....	24
Results.....	33
Discussion.....	48
References.....	51

ABSTRACT

Animals must make estimates about possible resources in order to choose the resource which will save them time and energy while conferring high energetic content. In order to make the most optimal decision, foragers must use various parameters to come up with an accurate estimate for each possible alternative. Learning rules allow us the possibility of analyzing which parameters animals may be using in order to make the best decision. We use compare known learning rules (i.e. Linear Operator Rule, Relative Payoff Sum Rule, Perfect Memory) and experimental data extracted from bumblebees (*Bombus impatiens*) subjected to a two armed bandit scenario in order to find what learning rule best describes their foraging choices in a changing environment. Our findings suggest that bumblebees seem to be using parameters consistent with the Linear Operator Rule and the Relative Payoff Rule. More importantly, our results suggest that there is great variance in learning rule use between individuals.

CHAPTER 1

I. Foraging and Choice Mechanisms

Animals are constantly exposed to resources that vary in energetic content. Each resource will also have corresponding predation risks, commuting time and energetic expenditure costs to it. Consequently, foraging individuals should seek high rewarding resources that can compensate for the aforementioned acquisition costs. The energetic content for each resource can be thought of as a currency that an individual forager must gain and invest at different times (Pyke 1979). This biological currency will allow foragers to compare between alternative resources in order to make the best possible decision, when choosing between the alternatives (Stephens and Krebs 1986).

Finding a high value food resource is imperative for foragers due to the link between foraging efficiency and fitness (Schoener 1971). Foragers that obtain higher energetic resources will not have to constantly invest in the acquisition costs involved in needing to forage more. Compared to foragers that only attain low rewarding resources and have to concurrently seek resources that can compete with the demands of their current environment, successful foragers are more likely to survive and reproduce (Pyke et al. 1977). This is because animals saving foraging time can then invest that time into alternative behaviors such as avoiding predation, finding mates, and defending resources. In order to increase the amount of food attained over time, also known as the food intake rate, foragers must have knowledge of the reward values for all of the possible alternatives.

Although foraging itself can seem incredibly complex, the analysis and expectations for behavior while foraging can be simplified through the use of

optimal foraging theory approaches. The optimal foraging models that arose after 1966 proposed that animals should maximize their food intake by making optimal decisions. Robert H. MacArthur, Eric Pianka and John Merrit Emlen were the first to propose optimal foraging theory in 1966 by comparing foraging strategies to economic theories. In MacArthur and Pianka (1966), they propose that a forager should consider staying or defecting to other alternatives depending on the available prey choices. When prey abundance is low, foragers should switch to other prey types. Alternatively, Emlen (1966) suggested that prey choice should be dependent on prey value such as caloric intake that would then serve to measure the appropriate handling time appointed to the alternative. These models consider how animals analyze and evaluate prospective choices while including the costs of each choice. The theory takes into consideration a forager's possible resource values, a cost-benefit analysis, and its effects on an optimal behavior (Pyke et al. 1977). A basic assumption of these early foraging models is that in order for foragers to make optimal decisions, they would have information on various parameters pertaining to each resource alternative. Optimal foraging strategies would then incorporate known parameters that would guide the forager to the best alternative. Specifically, foragers would have to know the energetic content and distribution of each possible resource choice.

Theoretically, animals select a behavior based on the information they have obtained from the environment (Blumstein 1996). In order to simplify this, scientists assumed that animals intuitively knew about their resources so as to optimize their food intake. It was assumed that animals should have some knowledge of the quality of each choice and behave accordingly in order to

optimize their food intake. Foragers would have knowledge of patch value to the extent of hierarchically categorizing resources and visiting them accordingly (Charnov 1973). The question remained as to what information were foragers using in order to make the optimal choice. For instance, animals that are foraging for the first time have yet to experience the given value of each resource.

Further work on optimal foraging theory has brought to light various frameworks on how animals use information. Essentially, animals could use various parameters such as handling time, caloric value, distance to resource, different resource types, abundance of resource and more, to make optimal foraging decisions. For instance, the Marginal Value Theorem (Charnov 1976) utilizes most of the previously mentioned parameters in order to estimate how animals should behave in order to optimize their food intake rate in a stochastic environment. Though Charnov's (1976) work required an animal to know about every aspect of its environment and possible resources, we now believe that animals need not know about all of the parameters about their environment. Foragers must learn from their environment in order to compile the necessary information (Iwasa et al. 1981).

II. Information Sources

Given that animals are not omniscient, information can alter the behavior of a decision maker and influence their fitness (Danchin 2013), presumably because information is valuable. For instance, reliable information should influence the response of an individual by decreasing the uncertainty surrounding a certain choice (Schmidt et al. 2010). Foragers find themselves faced with a variety of

decision types, where they must decide on various alternatives starting from where to forage, to when to leave a patch, to electing the best alternative out of an array of choices (Stephens 2008). Given so many choices and types of choices, effective foraging is reliant on reliable information in order to maximize foraging efficiency and thus help increase potential fitness.

Animals have many sources of information that can drive their behavioral responses. For solitary foragers, prior information of a resource can be gained via their genes or from aspects of their environment. Without ever experiencing the environment, foragers can inherit preferences that may bias their decisions. *Drosophila melanogaster* display an innate bias for oviposition substrate site that can be changed after generations of experimental selection (e.g. Mery & Kawecki 2004; Dunlap & Stephens 2009), but can also revert to their original preference when not being selected upon (Mery and Kawecki 2004). Examples such as these lead researchers to propose that though some foragers may rely heavily on inherited information, this inherited preference can theoretically be extinguished or even renew itself under selective pressures. Preferences might become fixed when the world remains unchanging and learning is not a reliable source of information. The information inherent to these preferences will be passed on through generations and remain relevant when deciding between possible resources (Dunlap and Stephens 2009).

Because the choices of the individual allowed it to survive to reproduction, the offspring should adopt behaviors similar to its predecessor (Danchin 2013). However information need not only be transmitted genetically. Importantly,

information can also be gained through, for example, cultural transmission by means of social learning through the parents (Danchin 2013). Still, many animals do not have the benefit of socially acquiring information about their environment and must gain prior information exclusively through their genes. Individuals can display preferences based on a sensory bias, where an individual will display a stronger response to a specific stimulus (Raine and Chittka 2007). Preferences that result in adaptive behaviors can then be strengthened through generations (Fuller et al. 2005), causing naïve foragers to commonly display an innate preference for certain resource. When given the option between the color blue or the color yellow, bumblebees of different species displayed varying preferences despite being associated to a negative payoff (Ings et al. 2009).

In the case of low change within the environment, information is more valuable since the individuals will be able to accurately estimate the true state of their environment (Dewitt 1998; Koops 2004; McNamara and Dall 2010). Still, it is unrealistic to think of natural environments as being stable. In fact, animals must adapt to varying environments with ephemeral resources. Resources can be depleted by stochastic events in the environment, causing the environment to vary in its reliability. When the environment becomes less persistent, it is less predictable and fixed behaviors are no longer adaptive (Dunlap and Stephens 2009). In this case animals must act on information gained from the environment itself in order to modify their behavior accordingly. Thus animals rarely have complete information about their environment, despite the early work on optimal foraging theory that assumed that foragers knew the value of important parameters such as prey size, handling time and encounter rates (Charnov 1973).

Finally, when one adds the random variation inherent in each parameter, this variation combines to produce imperfect information, as even inherited information in a relatively fixed environment can become no longer accurate (Stephens et al. 2007).

III. Sampling and The Two Armed Bandit

Stochastic environments present the problem of incomplete information (Stephens and Charnov 1982). In stochastic environments, foragers may find themselves having to choose from a set of alternatives with unknown value. The only way for the forager to learn this value is through sampling the available resources to acquire new information about rewards and distributions of resources. Sampling rate itself can be optimized, and this sampling is necessary for high performance in a foraging task (Krebs et al. 1978).

Incomplete information demands that a forager experience all possible alternatives before deciding on a possible choice (Pyke 1984). Sampling all possible alternatives can prove costly to foragers in terms of energy expenditure and time investment. To begin with, animals must invest in acquisition costs, as they must sample to acquire reliable information about a possible resource (Stephens 1987). Second, there are sampling error and overrun error costs that consist of missed productive opportunities and sampling unproductive options (Dechaume-Moncharmont 2005; Stephens 1987). Similar to both of these costs, foragers have to consider opportunity costs of choosing one alternative over another, and the probability of losing the profits of that unchosen alternative (Winterhalder 1983). Furthermore, sampling available information sources has

inherent predation risks attached to it. Still, in most types of environments, sampling should reduce uncertainty sufficiently to make better adapted decisions (Schmidt et al. 2010).

For optimal foraging, sampling should be done consistently since most environments vary unpredictably. As the forager experiences various parameters it will be able to better track its environment and accurately adjust to its current state (Dunlap and Stephens 2012). Every experience with each parameter will accumulate to provide an estimate of the actual value of the resource. As previously mentioned, because resources are ephemeral their value is constantly oscillating between depleting and regenerating. This means that every visit that a forager makes to a specific resource may result in a different value along a normal distribution curve. We expect that every resource has a different curve with its own specific mean. Foragers must then make an assumption on the mean value for each resource (Mangel 1990).

Foragers are constantly exposed to various alternatives which they then may choose to sample or not. Given all the aforementioned costs and risks associated with sampling, foragers might choose a random resource in which to forage from and hope for the best. In a multi-armed bandit scenario, decision-makers must choose from multiple alternatives (i.e. arms) that have an unknown reward (Reverdy et al. 2015). The multi-armed bandit refers to a scenario with multiple arms or alternatives to choose from. At each time point, the forager will choose an arm and experience its value. From then on, the forager can continue to experience various arms or continue with a preferred arm. As a note, the first

choice can be random or chosen due to some prior preference that the individual may have. The multi-armed bandit looks to understand how foragers sample their environment. The objective of the forager caught in a multi-armed bandit dilemma is to increase its gain of food, which matches the assumptions of the optimal foraging theory (Srivastava et al. 2013). Consequently, multi-armed bandits serve as a great tool to study optimal foraging.

It is not always in the best interest of a forager to sample in high frequencies, since sampling errors and overrun errors increase accordingly (Stephens 1987). When a forager settles on a reliable pattern for the depletion and renewal of a resource, the variance of the estimated mean value for the resource should decrease and the forager should no longer sample its environment at high frequencies (Pyke 1984). Foragers are then faced with the problem of acting on new information that may be gained through sampling or foraging based on information that has already been collected (Sherratt 2011; Stephens 1991; Shettleworth et al. 1988). The central problem within the multi-armed bandit revolves around the concepts of exploration versus exploitation.

Theoretically, foragers should choose the exploitation strategy when it encounters a high quality resource and an uncertain environment, as sampling can be costly (Dall and Johnstone 2002). In contrast, experimental work showed that blue jays exposed to a highly unpredictable environment will choose to sample their environment at higher frequencies and use an exploration strategy (Dunlap and Stephens 2012). To simplify this problem, let's consider that a forager only has to choose between two options (i.e. a two-armed bandit scenario) and the

proportion of certainty of the world is unknown. The two-armed bandit scenario has only two options to choose from. When recreating a two-armed bandit scenario using great tits, researchers found that foragers will display an optimal balance between both strategies and suggest that foragers use simple rules to establish the optimal proportion of exploitation and exploration (Krebs et al. 1978).

IV. Rules of Thumb

The multi-armed bandit measures how foragers make optimal decisions when foraging. In order to maximize their profits, foragers must exhaust a known resource or keep sampling their environment. Additionally, foragers may increase a pre-existing preference through exploitation or learn through their environment via exploration (Scott 2010). As exploitation and exploration are mutually exclusive actions that can't be done simultaneously, foragers should employ simple rules to cope with their environment.

Let's assume that while sampling their environment foragers employ simple mechanisms instead of investing in expensive neurophysiological costs associated to learning, retaining and retrieving memories. If this were the case, we would expect that natural selection would evolve simplistic mechanisms of choice that would result in effective foraging. Rules of thumb provide foragers with mechanisms that use minimal amounts of information about the environment and produce suitable behaviors while foraging (Real 1994).

Rules of thumb provide foragers with low cost, low risk behaviors that allow it to handle complex foraging tasks (Naug and Arathi 2007). Some of these rules can

include Herrnstein's matching law, where pigeons would continue choosing a specified alternative in correlation to the number of times the alternative was reinforced (Herrnstein 1961). We continue by mentioning a fixed giving up time rule, which was developed out of Charnov's Marginal Value Theorem, where animals should depart a patch after a set amount of time (Krebs et al. 1974). Additionally, the number rule states that the forager should depart after collecting a set number of prey items (Stephen and Krebs 1986). Finally, we conclude by mentioning the ε -sampling rule where foragers continuously sample between resources at a constant probability, and the failures departure rule where foragers should switch food after experiencing a payoff under a set threshold (Thuijsman et al. 1995). Ultimately, the goal of many rules of thumb is to decrease the costs of sampling.

Similar to the failures rule, foragers can also employ a win-stay/lose-shift rule. One of the simplest rules of thumbs, this rule states that foragers should switch to an alternative resource when the resource sampled proves unrewarding (e.g. Randall and Zentall 1997). This proves to be a reliable strategy with transitory resources, as foragers could potentially avoid visiting unrewarding patches. More so, foragers such as tamarins and free ranging titi monkeys have been observed to indulge in a win-stay/lose-shift rule when resources are predictable (Bicca-Marques 2005). On the other hand, marmosets did not adopt this rule of thumb (Platt et al. 1996) which leads us to look at other rules such as the win-shift rule which states that forager should switch to an alternative resource, even when the sampled resource proves to be rewarding. This rule can impede the forager's ability to exhaust good resources (Olton and Schlosberg 1978).

One of the many benefits of rules of thumb is the lack of information required and the non-existent need for memory. Still, rules of thumb rarely take into account a dynamic environment and the associated costs of sampling (Naug and Arathi 2007). Accordingly, rules of thumb result in perceptual errors that may impede optimal decisions (Bouskila and Blumstein 1992). In addition, bumblebee foraging behavior has been proposed to relate better to more complex optimal models when compared to rules of thumb (Birnaskie 2009).

V. Learning and Memory

Learning rules bring into question long-term versus short-term memory. Foragers can make predictions of food rewards based on past experiences (e.g. Lewis 1986; McNamara and Houston 1987; McNamara et al. 2006). For instance, bumblebees exposed to a two-armed bandit experimental design were found to use prior choices when making a decision and not solely relying on their last choice (Kearar et al. 2002).

In order for foragers to increase the frequency of correct choices and decrease the number of errors, they must learn to track their environment (Morand-Ferron and Giraldeau 2010). Though we've mentioned tracking experiences before, environmental tracking also implies the retrieval of past experiences that can be applied to current events (Dunlap et al. 2009). Though rules of thumb may prove to be successful in relatively fixed environments, stochastic environments call for an updating estimate of a resource value that can only be gained through learning and, thus, memory (Eliassen et al. 2009).

Animals need to retain information in order to make optimal foraging decisions (Papaj and Lewis 1993). Yet, foragers are also constantly sampling their environment and must integrate their newly acquired information to their previously experienced information. Animals must find a way in which to balance both past and present information. One way that animals may do this is by weighing both types of information.

Past information deteriorates and becomes unreliable with time, but can also be a good predictor of future conditions (McNamara and Houston 1987). If sufficiently sampled, animals can come up with a consensus about the current conditions of their environment and have a good estimation about resource values. To this point, animals must establish a rate at which past information is discounted in order to use reliable information. We can use α as the memory variable, also known as the discount rate. In other words, it is the rate by which past information is being deducted when making current choices. This means that the alpha for prior information can range from one to zero. If animals were to exponentially weight the past, they would attribute higher value to either the past or the present information and will remain fixed throughout their sampling trials (McNamara and Houston 1987). In this scenario, a smaller alpha would put less weight on past information and more weight on present events.

Now let's assume that the value of alpha is not fixed and will fluctuate with the amount of observations the forager is experiencing. According to McNamara and Houston (1987), the α should decrease with uncertainty and foragers should place

more value on current information. Thereby the value α should increase as the number of observations increases. Logically, foragers should not place a high value on past experiences when they start foraging as they have no past experience to rely on (Dunlap et al. 2009). As their experience increases, then they may place more value on past observations. Respectively, jays rely more heavily on past information when exposed to a highly variable environment (Dunlap and Stephens 2012).

VI. Learning Rules

Foragers exposed to an unpredictable environment should evolve learning mechanisms that surpass the limited capacities of the rules of thumb (Lea et al. 2012). For this reason, researchers have moved to study optimality models and construct a specific model with relevant parameters thought to influence foraging behavior (Parker and Maynard Smith 1990). Optimality models would then take into account time, the state of the environment, and the actions of conspecifics (McNamara et al. 2001). Dynamic environments reflect a change of environmental state from one unit of time to another. As a result, simulations comparing optimality models and rules of thumb strengthen the claim that optimality models are better solutions for foraging problems (Janetos and Cole 1981). Optimality models are classically represented as mathematical models that include parameters that seem relevant to the behavior being optimized.

Naïve foragers may use strategies similar to rules of thumb that may be adjusted as the individual experiences its environment (Morand-Ferron and Giraldeau 2010). Bumblebees, for instance, adjust their patch leaving rule with their

experience of the environment (Biernaskie et al. 2009). At this point, rules of thumb are no longer fixed and animals start utilizing learning rules. Learning rules are represented as equations that predict the likelihood of foragers selecting between possible alternatives (Mookherjee 1997). These rules will also determine how individuals will adjust their decisions as they sample their environment (Hamblin 2009). Most of the well-known learning rules integrate prior information to current observations. Successful learning rules have the ability to approximate optimal foraging strategies to the point of invading a population and becoming the predominant strategy used in said population (Beauchamp 2000).

i. **Bayes' Theorem**

Probably the best known learning rule is Bayes' theorem, where prior information is combined with current information to make a proper assessment of the environmental state (Hamblin et al. 2009). Assume that a forager has a choice between resource A and resource B; per Bayes' theorem, we should be able to calculate the conditional probability of the forager choosing either alternative. Using the equation $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ we can calculate the probability of the forager choosing A given that it could also choose B, while also taking into account the prior information of resource A and resource B. Keep in mind that a prior can be genetically acquired or through sampling the environment. The animal does not require any more information, other than its past experiences and will develop an average rate of the environment (McNamara et al. 2006).

Bayes' theorem does not directly employ the use of memory parameters or sampled value. We can view Bayes' equation as a strengthening or dilution of a

given preference. The first choice a naïve forager makes will, theoretically, be driven by a preference (Valone 2006). If the first choice based on a preference proves to be rewarding, said preference will keep increasing. If the preferred choice proves to be unrewarding or becomes unrewarding at any given point, then the preference should start to fade and even go to extinction. Under this assumption, Bayes' theorem may also be referred to as Bayesian updating, as each new choice will update the probability of choosing the same reward in the future.

This model has gained support from both simulations and empirical studies. For instance, *Apis cerana* were proposed to follow Bayesian processes that allowed them to successfully complete a complex foraging task with correct behavioral responses (Naug and Arathi 2007). Bayesian foragers have been predicted to reduce the cost of incomplete information, making this model one of the primary learning rules proposed to be used by animals (Olsson and Brown 2006). As *Apis* is not the only genera to follow the expectations of Bayesian processing, researchers have proposed that natural selection has driven animals to follow the parameters of Bayes' theorem (Trimmer et al. 2011). Still, though many foragers closely follow the assumptions made by the model, bumblebees have been identified as less than perfect Bayesian foragers while still not following patterns that would fit known rules of thumb (Biernaskie et al. 2009).

ii. **Linear Operator**

Let's suppose that instead of foragers updating previous preferences, they come up with an estimate about the value of each possible alternative. Not forgetting that the value for each resource may change at different time points, foragers

should then update each estimate at every time point. Under these assumptions we arrive at the linear operator learning rule which integrates observed values, past estimates, and weighted memory using the equation $\mu_n = \alpha_n \mu_{n-1} + (1 - \alpha_n) x_n$ (McNamara and Houston 1986).

The given equation results in an estimation of the value of a resource at time n (μ_n). From here, the equation combines the rate in which the past is discounted with the estimated value at the previous time unit ($\alpha_n \mu_{n-1}$). This first half of the equation will determine the weight that will be given to past estimates. Finally, the second half of the equation will generate the weight given to the observed value at the current time. The value $(1 - \alpha_n)$ will denote the residual of what was discounted from past information, and x_n represents the observed value at time n . The result will also reflect current information. With more observations, the estimates for a resource should increase or decrease exponentially (Beuchamp 2000).

Similar to Bayes' theorem, linear operator allows for the forager to update its information in accordance to its sampling. Unlike Bayes' theorem, this model doesn't take into account preferences or prior information. When both models were compared in terms of performance Grob et al. (2005) demonstrated that though both models adequately chose the best course of action in the short-term, neither could perform adequately over a longer period of time.

iii. **Relative Payoff Sum**

The relative payoff sum integrates both of the prominent features of the linear operator and Bayes' theorem. Mathematically similar to the previous learning rule, this model integrates rates of discounting the past, observed values at the current time, past estimates and priors. Distinct from the linear operator, $\mu_n = \alpha_n \mu_{n-1} + (1 - \alpha_n) r_A + x_n$, this equation includes a prior (r_A) for a given resource. By adding this new variable, the estimated value of a resource cannot be reduced to zero due to the prior (Hamblin and Giraldeau 2009).

First proposed by Harley (1981), the relative payoff sum gets its name due to the probability of choosing a given alternative in correlation to the payoff gained from it during previous trials. Seemingly the learning rule that yields the most optimal behaviors, relative payoff sum has moved to fixation faster in computer simulations when compared to linear operator and perfect memory (Hamblin and Giraldeau 2009).

iv. **Perfect Memory**

Lastly, the perfect memory learning rule does not integrate either current observed values or memory windows. Instead it employs the use of priors and cumulative payoffs for one alternative averaged over the number of times the forager has selected that alternative. The equation $\mu_n = \frac{\alpha + R_{A_n}}{\beta + N_{A_n}}$, starts with an estimated value for resource A at time n. For this equation α will no longer represent the rate of discount for past information. Instead α will represent the prior or preference expressed for the alternative A ranging from zero to one, while β will equal the residual value of the prior. After which point the first part R_{A_n}

will represent the total amount of payoff gained from alternative A up to time n . Subsequently, N_{A_n} symbolizes the amount of times alternative A was chosen up to time n .

Contrary to the Linear Operator Rule and the Relative Payoff Sum rule, the Perfect Memory Rule does not discount memory. In fact, this rule gives equal importance to all its past rules and integrates a preference for each choice (Houston and Sumida 1987). In there lies the problem with the Perfect Memory Rule according to Hamblin and Giraldeau (2009). Their work suggests that due to its mathematical parameters, the Perfect Memory Rule cannot keep up with sudden changes consistent with a stochastic environment.

CHAPTER 2

I. Introduction

Nature is composed of ephemeral food resources that are constantly fluctuating between high and low values of reward. Consequently, foragers have to make decisions based on the variance exhibited by potential resources. The goal of said foragers is to enhance their food intake rate and their overall fitness (Schoener 1971). Foragers must accurately estimate the reward value associated to each resource by using various environmental parameters (McNamara and Houston 1987). Accurate estimates of resource values will lead to successful foraging decisions, which is the basis of optimal foraging theory.

Early works on optimal foraging theory assumed that foragers had complete information about every environmental parameter (Papaj and Lewis 1993, Charnov 1973). Yet, animals are not omniscient about the state of their current environment and must learn how to best exploit it. Furthermore, foragers do not require all of the available information in their environment to make optimal foraging decisions (Iwasa et al. 1981). Instead, individuals should rely either on inherited information or information gained through sampling in order to modify their foraging decisions to best suit their current environmental state. When given the choice between flowers of differing colors, bees display an innate predisposition for colors such as blue and yellow which have been associated to flowers with high nectar rewards (Giurfa et al. 1995). Thus, color preferences result from inherited information that can be favored by natural selection and used to make adequate foraging decisions (Raine and Chitka 2007). Still, as the environment becomes less predictable and preferences may no longer be adaptive, foragers must learn by sampling their environment (Dunlap and

Stephens 2009).). Bees have been shown to disregard colors when color preference is no longer the optimal foraging strategy (Hill et al. 1996). However, sampling all the available alternatives would be time consuming, energetically expensive and prone to predation risks (Dechaume-Moncharmont 2005, Dornhaus et al. 2005, Stephens 1978, Winterhalder 1983).

In a multi-armed bandit scenario, foragers must extract information from the environment itself by sampling from an unknown set of alternatives (Reverdy et al. 2015). With each new choice, the decision-maker will attempt to maximize its food intake rate by combining known environmental parameters in order to make accurate estimates about their environmental conditions (Krebs et al. 1978). It was originally suggested that foragers should employ simple rules of thumb in order to make the most optimal decisions with a minimal amount of information (Real 1994). It has now been suggested that foragers, such as bees, follow foraging patterns consistent with more complex learning rules (Biernaskie et al. 2009).

If foragers are indeed learning with each new choice, then we would expect them to be incorporating past information to current events. Under the assumptions of known learning rules, foragers should update information with each new choice and have a choice mechanisms based on estimated values for each possible alternative (Beuchamp 2000). Most notable, Baye's theorem has been often used to study how foragers incorporate past experiences to their current knowledge in order to make adaptive decisions (McNamara et al. 2006). In this case, innate preferences could serve as a prior that can tend towards fixation, extinction or

remain in fluctuation (Valone 2006). Although bees follow trends similar to the assumptions established by Baye's theorem there may be other learning rules that better describe how bumblebees make foraging decisions (Biernaskie et al. 2009).

By exposing individual bumblebees (*Bombus impatiens*) to a two-armed bandit scenario where the reward values of two resource are known and changing, we aim to find what parameters bees are using to make optimal foraging decisions. We compare well established learning rules: Baye's Theorem, Linear Operator, Relative Payoff Sum, and Perfect memory to find the one that best describes the bees foraging patterns. By comparing the learning rules various parameters, we hope to find how animals are processing information. We believe that foragers should utilize more than just preference and integrate the use of current values sampled as well as prior estimates consistent with the relative payoff sum rule. Furthermore, we compare the weight that foragers are placing on current observations versus past experiences.

II. **Methods**

a. **Study Design**

We exposed ten individual foraging bees, *Bombus impatiens*, to a T-maze mirroring a two-armed bandit dilemma in order to measure how bees make continuous decisions in a changing environment. The maze consisted of a total of ten nodes and two branches per node representing a possible alternative to choose from. At the start of the maze, bees would have no prior knowledge about the reward value for each alternative. After making a choice between the left or right branch, panels would be placed to prevent the bees from back tracking or

sample the other alternative. This was repeated after each choice made. The bees would then have to experience the completed maze a total of eight times for a total of 80 choices.

The first 20 trials experienced by the forager would always be a bad state, meaning that the bee would have to choose between a stable resource equal to six microliters of 20% sugar solution and a fluctuating resource of six microliters of 5% sugar solution. The next 20 trials would then switch to a good state, where the fluctuating resource will now equal to six microliters of 50% sugar solution. This pattern would then repeat itself once more, for a total of 40 choices in a bad state and 40 choices in a good state. New maze liners were placed after each run.

We provided two different stimulus of colors, blue and yellow, to work as sensory cues for the bees. Each alternative would be randomly assigned to one of the colors for each individual bee. Color stimuli did not remain attached to a specific alternative for all bees. Additionally, placement of the alternative resources was randomized for left or right for each trial. The first choice for each individual was used to average the strength of preference for each color and compared it to a generalized model to quantify the strength of color preference for each individual.

b. Optimizing Memory Weight

Using Beuchamp (2000) algebraic equations for the learning rules, we simulated and compared each equation using various memory weights. Memory weight has been proposed to be either constantly weighted in favor of past experiences or present observations, or be a memory window. Memory window has gained more

support based on McNamara and Houston (1987) claim that individuals give more weight to past experience as they gain more experience of their environment. We ran the linear operator rule and the relative payoff sum using constantly weighted alphas ranging from zero to one. Furthermore, we also simulated both learning rules using a memory window following the equation $\alpha_n = \frac{n-1}{n}$, where n equals the number of trials experienced by the individual bee. After running each simulation and obtaining the estimates for the values of the stable resource and the fluctuating resource at each time unit, we calculated the expected choice for each time point and compared it to what was actually observed. The expected choice is the proportion that the individual bee will choose a given resource based on the past estimate, $\frac{\mu_i}{\mu_i + \mu_j}$. The difference between the expected choice and the observed choice determine a new value, which we call the behavioral consistency to model.

An repeated measures univariate ANOVA was ran to see if there were differences between each alpha, including the memory window (α_n). We ran two statistical tests per learning rule, for the values of the stable resource and the fluctuating resource. Both resources assume either a stable environment or a fluctuating environment. Each alpha was a treatment and the subject and trial number were considered as factors influencing the behavioral consistency to model. Following the repeated measures univariate ANOVA, we ran a post-hoc Tukey test to compare each group.

Finally, we also simulated the Perfect Memory rule using Beauchamp (2000), but as there is no use of memory weight, we solely compared estimated choice to the

observed choice. Similar to our other simulations, we used a repeated measures ANOVA while only taking into consideration individuals and trials as factors.

c. The Role of Preference

To uncover each individual's color preference, we used $Y=1$ or 0 to denote which alternative the bee chose. We averaged the first choice of all ten foraging individuals to get a proportion of 0.6 strength of preference for the color blue and 0.4 strength of color preference for the color yellow. By using a generalized model, we tested the preference for each color for each individual in order to obtain the exact preference displayed by the bee. In which case, r_i and r_2 will correspond to the residuals for each color and i will now equal the bees from one to ten.

d. Simulations and Comparing Learning Rules

We used the algebraic equations used in Beuchamp (2000) for the Linear Operator Rule, the Relative Payoff Sum Rule and the Perfect Memory Rule. As the bees were introduced to a two armed bandit scenario, they had to choose between two possible options consistent with the assumptions of the learning rules. Essentially, bees had to make an estimate for each of the alternatives and choose accordingly. Our first aim was to simulate the estimates made by the foraging bees at each time point for both alternatives. The stable alternative will be representative of a stable environment, while the fluctuating alternative will equal an unpredictable environment.

i. Linear Operator Rule

Following the equation $\mu_n = \alpha_n \mu_{n-1} + (1 - \alpha_n) x_n$, where the μ_n equals the estimate for the reward value of an alternative and x_n equals the observed values. We will discuss α_n , memory weight, in the next section. For the stable environment, x_n will always equal 20, unless it is not chosen and then it will be zero. The fluctuating resource will equal 5 in a bad state or 50 in a good state, if not chosen then it will equal zero. μ_n will have a value for each trial or time unit determined by the observed value chosen, the memory weight and the past estimate. Additionally, every estimate for the stable alternative, μ_A , will have a corresponding estimate for the fluctuating alternative, μ_B .

ii. Relative Payoff Sum Rule

Similar to the equation for the Linear Operator Rule, the Relative Payoff Sum now includes a residual which we equaled to a prior preference exhibited by the individuals. The residual was calculated by averaging the first choice for color for all ten individuals, concluding in a residual of 0.6 for blue and 0.4 for yellow. Under these assumptions we follow the equation $\mu_n = \alpha_n \mu_{n-1} + (1 - \alpha_n) r_A + x_n$ and repeat the same process as with the Linear Operator Rule. In contrast to the Linear Operator Rule, we apply the residual for every individual forager. As colors are not consistently associated with one resource alternative, we made sure to combine every individual with their corresponding residual for each alternative.

iii. Perfect Memory Rule

In the case of the Perfect Memory Rule, $\mu_n = \frac{\alpha + R_{i_n}}{\beta + N_{i_n}}$, we make use of the corresponding residuals similar to how we did for the Relative Payoff Sum.

Furthermore, we use the same values for the observed values, but these are joined into the total payoff for that alternative, R_i .

e. Modelling and Comparing Learning Rules

In collaboration with Dr. Yuefeng Wu, we used the observed data to compare various models, each with their own parameter to uncover how bees are weighing memory and their learning process. The individual foraging bees must follow the assumptions of the learning rules and estimate the value of the alternatives during that time unit. Each learning rule discussed carries its own set of assumptions and parameters used. Similar to Baye's theorem, the linear operator and the relative payoff sum update their current estimates about the value of a resource using previous estimates for that same resource. In contrast, Linear Operator excludes the use of preference for some sensory cue and includes the actual reward value obtained during that time unit. The Relative Payoff Sum also uses previous estimates, current observed values, and includes a residual value that represent preference for color. For each of these rules there is the added component of a rate of memory discount where the forager should place weight on past estimates or current observations. Under the assumptions of McNamara and Houston (1987), foragers could have a weighted memory and place a constant weight to the past and the present observations or have a memory window the weight placed on past observations should increase with every new observation. Finally, the Perfect Memory rule assumes that all past experiences are remembered and equally weighted by the forager. Due to the nature of our experimental study, there was a need to reparametrize the original equations suggested in Hamblin and Giraldeau (2009).

i. Relative Payoff Sum

$$(1) \mu_i = \alpha\mu_{i-1} + (1 - \alpha)r_i + x_i$$

here the μ_i is the estimate where $i=1$ or 2 corresponding to the two alternatives, x_i equals the reward value of alternative i at time t , r_i is the residual corresponding to the information or preference that the bees had before it obtains any "new" information in current experiments for alternative i , and α_i reflects the rate of memory discount.

Re-parametrize Relative Payoff Sum

We let $\mu_i(1) = r_i$ before acquiring any reward during the first time point.

We assume that the first choice made by the individual bee will be determined by the residual term. This gives us the equation

$$(2) \mu_i(t) = x_i(t) + \alpha x_i(t-1) + \alpha^2 x_i(t-2) + \dots + \alpha^{t-1} x_i(1) + r_i$$

Assume $x_i(t)$ are identically distributed with μ , then

$$(3) E(\mu_i(t)) \rightarrow \frac{1}{1-\alpha} \mu + r_i$$

as $t \rightarrow \infty$

ii. Linear Operator

$$(4) \mu_i = \alpha\mu_{i-1} + (1 - \alpha)x_i$$

the equation removes the residual from the relation payoff sum rule which causes the current observations to be directly multiplied by the weight placed on current observations.

Re-parametrize Linear Operator

$$^{(5)}\mu_i(t) = (1 - \alpha)[x_i(t) + \alpha x_i(t-1) + \alpha^2 x_i(t-2) + \dots + \alpha^{t-1} x_i(1)]$$

Assume $x_i(t)$ are identically distributed with mean μ , then

$$^{(6)}E(\mu_i(t)) \rightarrow \mu$$

as $t \rightarrow \infty$

The linear operator would be the same as the relative payoff sum while excluding the residual term. The difference would be the asymptotic expected values of the alternatives due to the absence of the residual term, which allows the value to fall to zero. This would have no meaning for the decision-making mechanisms of the bees.

iii. Perfect Memory

$$^{(7)}\mu_i(t) = \frac{\alpha + R_i}{\beta + N_i}$$

in which case the ratio of μ_1 and μ_2 is the only thing that will affect the bees choice, α will no longer represent a discounting memory factor and will now equal the residual for i , R_i will be the total payoff from alternative i till time t , and N_i is the total payoff till time t .

Re-parametrize Perfect Memory

α will now equal to r_1 and β will equal to $r_1 + r_2$. $\sum_{j=1}^t X_i(j)$ will be the total payoff from alternative i till time t . This will be a special case of the relative payoff sum with $\alpha = 1$.

$$^{(8)}\mu_i = r_i + \sum_{j=i}^t X_i(j)$$

Another type of Perfect Memory model:

$$^{(9)}\mu_i = \alpha_t \mu_i(t-1) + (1 - \alpha_t) X_i(t)$$

where $\alpha_t = (t - 1)/t$. This model has the same form as the Linear Operator with alpha serving as a memory window denoted as α_t . Using a simple transformation, we can equal the Linear Operator equation to

$$_{(10)}\mu_i = \frac{\sum_{j=1}^t x_i(j)}{t}$$

This new equation can be used as a special case model where the residual is 0 and up to some multipliers of $1/t$.

Perfect Memory and Linear Operator are special cases of the Relative Payoff Sum.

iv. Iterative form and memory window

The general memory window type of model

$$_{(11)}\mu_i(t) = x_i(t) + \alpha_1 x_i(t-1) + \alpha_2 x_i(t-2) + \dots + \alpha_{t-1} x_i(1) + r_i$$

$$\text{with } \alpha_1, \dots, \alpha_w \neq 0 \text{ and } \alpha_{w+1}, \dots, \alpha_{t-1} = 0$$

Because the alpha can be seen as the strength of memory, so we assume that the weight of memory decreases as the time lag increases

$$_{(12)} \alpha_t = \prod_{j=1}^{t-1} A_j$$

and estimate $\{A_j\}$ will show the pattern if it exists. Notice that $0 < A_j < 1$ for $j=1, \dots, t$.

This model should not be used at the individual level as it will cause over fitting.

A larger sample size is required to find the memory patterns.

III. Results

A. Optimizing Memory Weight

a. Linear Operator

A univariate repeated measures ANOVA of 11 memory weights from zero to one (0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1) showed that alpha is a statistically significant factor ($F_{(11,1719)}=251.04, p<0.00000001$). Except for when the alpha is one, the difference between the observed values and the predicted values of ten out of the 11 memory weights used tend to zero, and thus follow the predictions of the model more closely. We used the difference between observed choices and the predictions of the model to analyze how closely the behaviors of the bees followed the assumptions of the Linear Operator rule. The predicted choices were calculated using the equation $\frac{\mu_i}{\mu_i + \mu_j}$. As the values calculated using each of the memory weights tend to zero, the observed choice and the predicted choice are equal to one another. Negative values will indicate that the bees are choosing the predicted choice less and positive value indicate that bees are over-choosing the predicted alternative. When alpha is equal to one, we see an extreme value that diverges from the observed trend. Once the memory weight of one is dropped from the analysis, we observe that memory weight is no longer statistically significant ($F_{(10,1575)}=1.33, p=0.21$).

We also included a memory window ($\frac{1}{1-\alpha}$) in our analysis, which also tends towards zero, but differs from the 11 memory weights used. In an individual contrast analysis within the ANOVA, we observe the memory window is significantly different from the other memory weights used ($F_{(1,1719)}=7.99, t=2.83, p=0.005$).

Effect	Effect (F/R)	SS	Degrees of Freedom	MS	F	P
Intercept	Fixed	4.42652	1	4.426518	18.1175	0.002122
Bee	Random	2.1981	9	0.244323	10.1135	0.000000
Alpha	Fixed	66.71027	11	6.064570	251.0361	0.000000
Trial	Fixed	13.53926	15	0.902617	37.3628	0.000000
Alpha*Trial	Fixed	6.67781	165	0.040472	1.6753	0.000001
Error		41.52788	1719	0.024158		

Table 1. Linear Operator Univariate Repeated Measures ANOVA

Effect	Effect (F/R)	SS	Degrees of Freedom	MS	F	P
Intercept	Fixed	0.11474	1	0.114737	1.27904	0.287312
Bee	Random	0.80735	9	0.089705	4.47227	0.000008
Alpha	Fixed	0.26605	10	0.026605	1.32639	0.210464
Trial	Fixed	14.80415	15	0.986943	49.20432	0.000000
Alpha*Trial	Fixed	4.69693	150	0.031313	1.56111	0.000039
Error		31.59144	1575	0.020058		

Table 2. Univariate Test of Significance for the Linear Operator Rule excluding alpha equal to one.

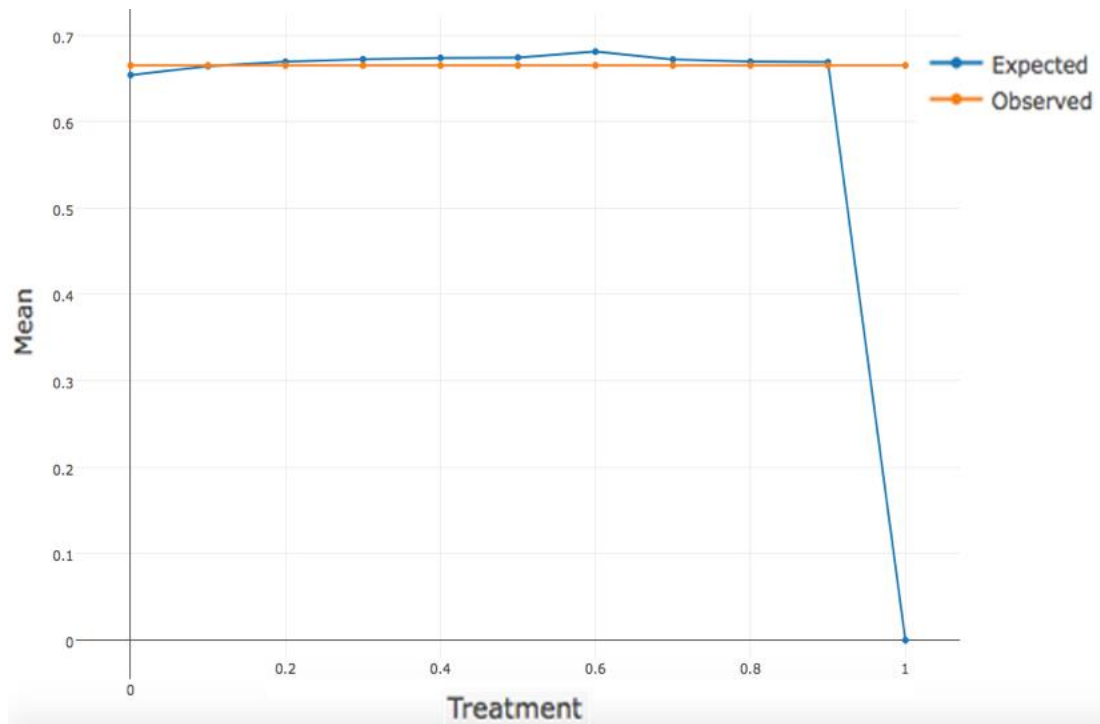


Figure 1. Linear Operator Rule Consistency to Data for the Stable Resource

The x-axis displays the alphas used from zero to one. The y-axis displays the mean of the predicted values and for the observed choices. At each of the points that the lines intersect the value of the observed choice overlaps with the values of the predicted choices and are equal to each other. Overlapping points reflect the efficiency of the model in predicting the choices of the bees for the stable resource.

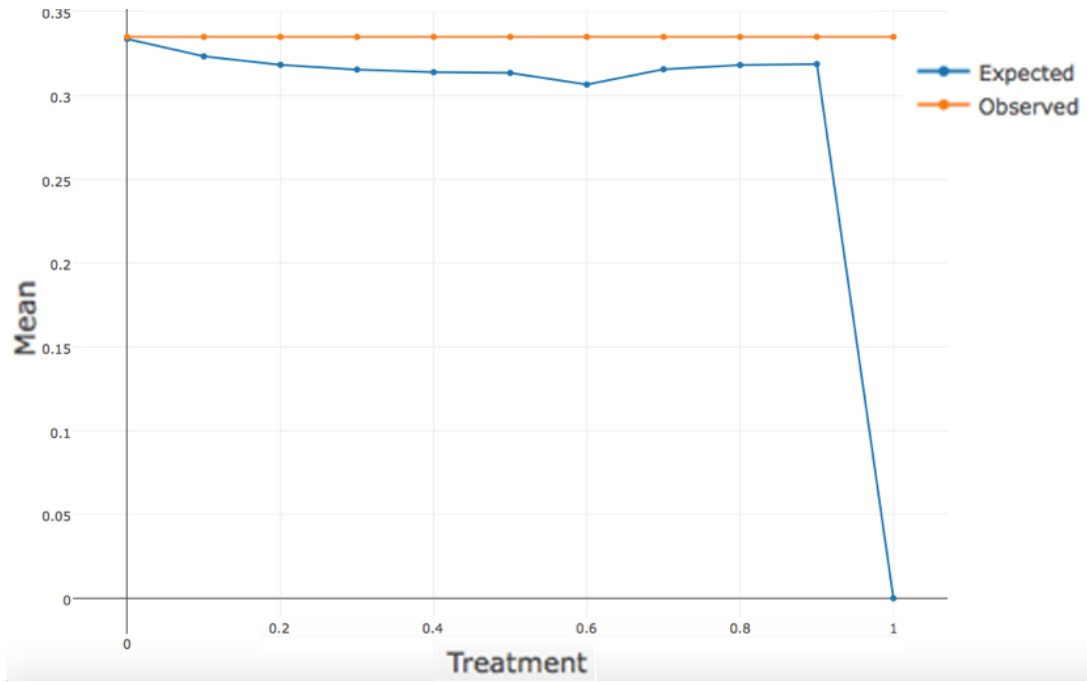


Figure 2. Linear Operator Rule Consistency to Data for the Fluctuating Resource

The x-axis displays the alphas used from zero to one. The y-axis displays the mean of the predicted values and for the observed choices. At each of the points that the lines intersect the value of the observed choice overlaps with the values of the predicted choices and are equal to each other. Overlapping points reflect the efficiency of the model in predicting the choices of the bees for the fluctuating resource.

i. Function of memory weight in trials

Trials are statistically significant for the Linear Operator Rule ($F_{(15,1719)}=37.36$, $p<0.00000001$), even when alpha equal to one has been removed ($F_{(15,1575)}=49.20$, $p<0.00000001$). Furthermore, there was a statistically significant trial by alpha interaction ($F_{(165,1719)}=1.67$, $p<0.00000001$).

ii. Function of memory weight with resource type

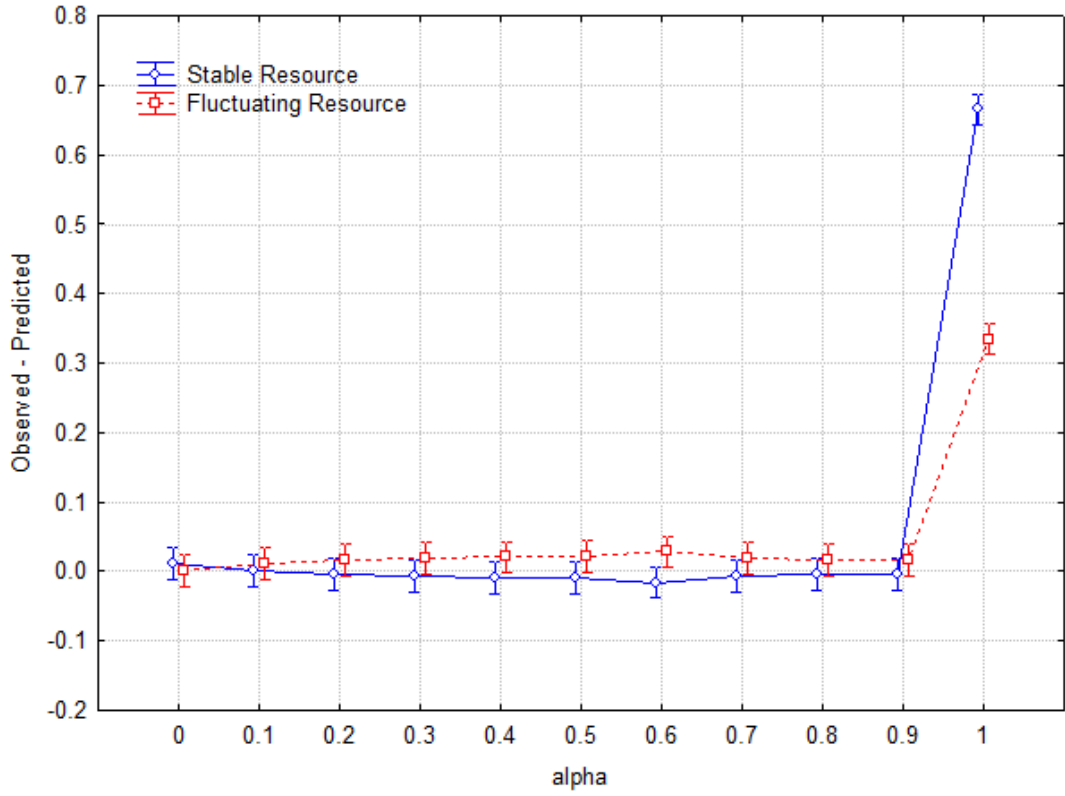


Figure 3. Linear Operator Rule Consistency to Data by Resource Type

The x-axis displays the alphas used for the Linear Operator rule ranging from zero to one. The y-axis displays the difference between the observed choices and the predicted choices, where zero means that there is no difference. The blue line reflects the stable resource which continuously has the same reward value, the red line is the fluctuating resource that changes value every 20 trials. Values higher than zero equal to over-choosing the resource. Bees seem to be choosing stable and fluctuating resources more than predicted.

b. Relative Payoff Sum

Our univariate repeated measures ANOVA showed that memory weights was a statistically significant factor ($F_{(10,189)}=34.22, p<0.00000001$). We compared 11 memory weights from zero to one (0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1). Similar to the Linear Operator Rule, alpha equal to one diverges from the trend towards zero. In contrast to the Linear Operator Rule, the Relative Payoff Sum

with a memory weight of 0.8 is closest to zero, following the predictions of the model the closest.

Effect	Effect (F/R)	SS	Degrees of Freedom	MS	F	P
Intercept	Fixed	0.385083	1	0.385083	3.80490	0.082885
Bee	Random	0.910864	9	0.101207	26.92548	0.000000
Resource	Fixed	0.044905	1	0.044905	11.94662	0.000676
Alpha	Fixed	1.286203	10	0.128620	34.21859	0.000000
Resource*Alpha	Fixed	3.115407	10	0.311541	82.88334	0.000000
Error		0.710411	189	0.003759		

Table 3. Relative Payoff Sum Rule Univariate Repeated Measures ANOVA

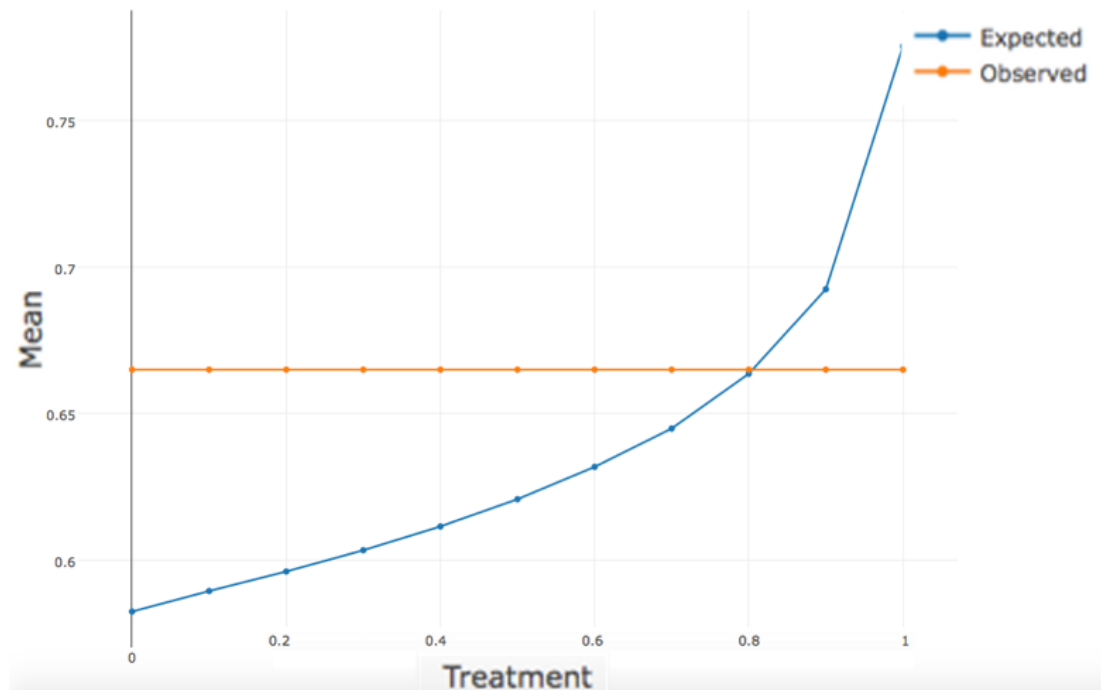


Figure 4. Relative Payoff Sum Rule Consistency to Data for the Stable Resource

The x-axis displays the alphas used from zero to one. The y-axis displays the mean of the predicted values and for the observed choices. At each of the points that the lines intersect the value of the

observed choice overlaps with the values of the predicted choices and are equal to each other. Overlapping points reflect the efficiency of the model in predicting the choices of the bees for the stable resource.

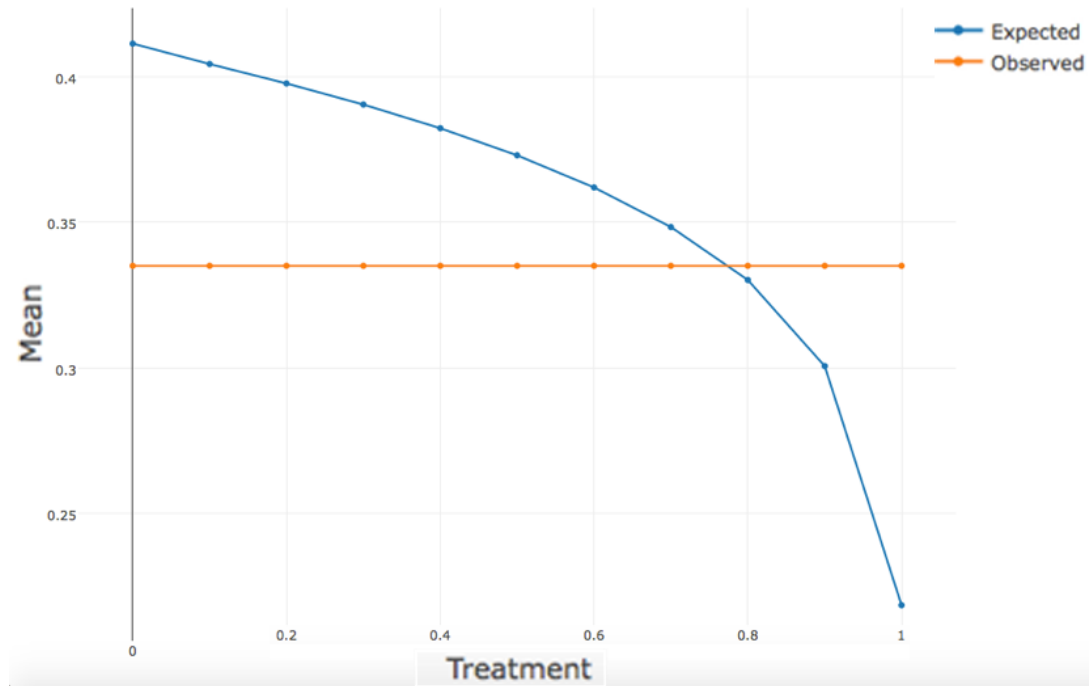


Figure 5. Relative Payoff Sum Rule Consistency to Data for the Fluctuating Resource

The x-axis displays the alphas used from zero to one. The y-axis displays the mean of the predicted values and for the observed choices. At each of the points that the lines intersect the value of the observed choice overlaps with the values of the predicted choices and are equal to each other. Overlapping points reflect the efficiency of the model in predicting the choices of the bees for the fluctuating resource.

i. Function of memory weight with resource type

The stable resource (μ_A) and the fluctuating resource (μ_B) were analyzed separately, both displaying a interaction with alpha ($F_{(10,189)}=82.88$, $p<0.00000001$). When compared to each other, we observe differences in alpha equal to zero ($t=2.60$, $p=0.01$), alpha equal to 0.1 ($t=2.72$, $p=0.007$), alpha equal to 0.2 ($t=2.66$, $p=0.008$), alpha equal to 0.3 ($t=2.50$, $p=0.013$), and alpha equal to 0.4 ($t=2.26$, $p=0.025$). Memory weights of 0.5 and above showed no difference

between the stable resource and the fluctuating resource. As the memory weight increases, the difference between the observed values and the predicted values declines towards zero. The fluctuating resource displays a closer trend towards zero, in comparison to the values of the stable resource.

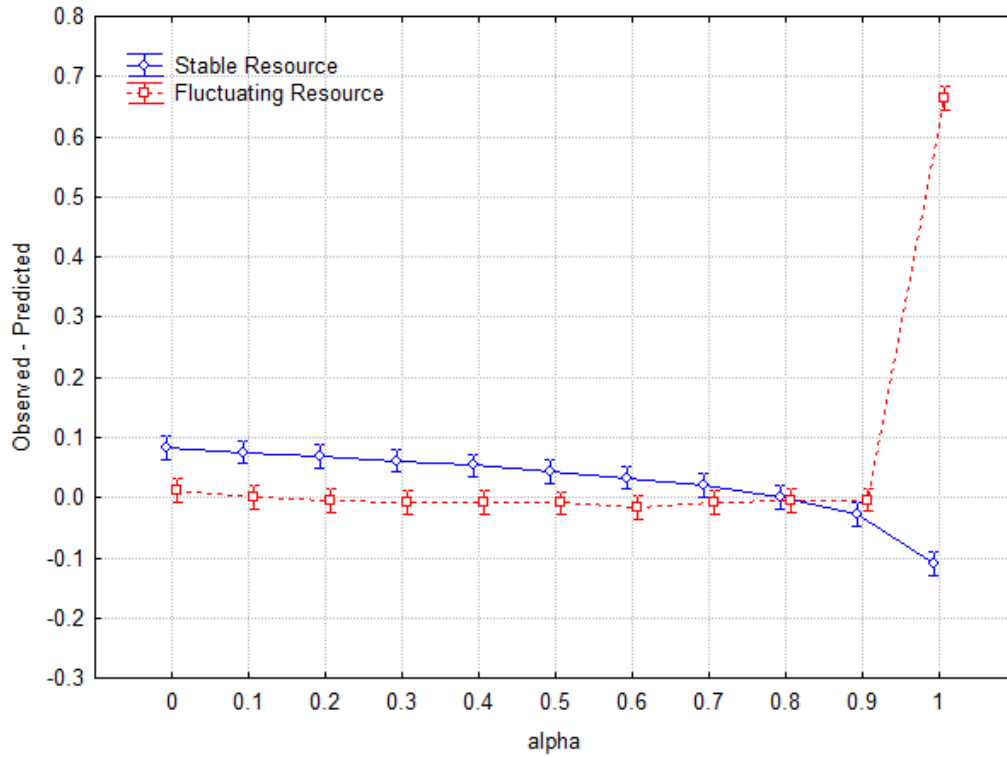


Figure 6. Relative Payoff Sum Consistency to Data by Memory Weight for Resource Type

The x-axis displays the 11 different memory weights used ranging from zero to one. The y-axis displays the observed choices of the bees to the predicted choices based on the model. We can observe that as memory weight increases, the data tends towards zero which means that the predicted choice and the observed choice are the same. More so, we observe that there is a significant difference between the memory weights and the memory weight equal to one. Additionally, the values of the fluctuating resource are more consistent with the predictions of the Relative Payoff Sum.

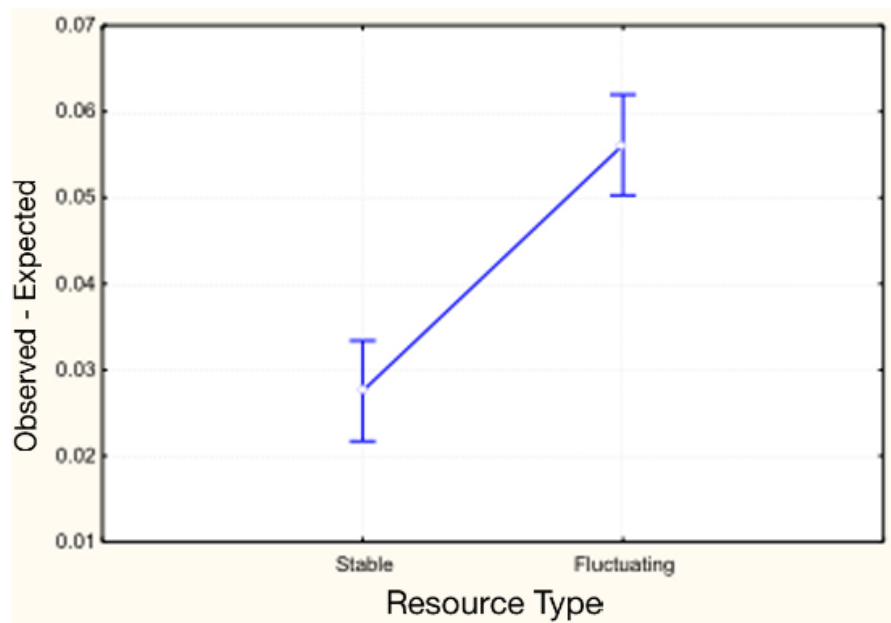


Figure 7. Consistency to the Relative Payoff Sum based on Resource Type

The x-axis displays the two possible resources encountered by a bee: the stable resource and the fluctuating resource that changes every 20 trials. The y-axis displays the difference between the observed choice and the predicted choice of the model. The closer the values are to zero, the better the model predicted the choice of the bee. The choices made in the stable resource seem the more consistent to the model (~0.027) than that of the fluctuating resource (~0.055).

B. The Role of Preference

Using the first choice of the ten bees, we got an average of 0.6 for the color blue and a 0.4 for the color yellow. We then confirmed this preference using the using of 30 different bees exposed to the T-maze, were we found that the preference for the initial choice of the color blue is 0.63 and 0.36 for the color yellow. The generalized model showed that nine out of 10 bees had a strong preference for the color blue, while subject eight showed a strong preference for the color yellow. Essentially, the ratio for each bee demonstrated a flexible preference, except for subject #8 whose ratio reflected a strong preference for the color blue.

C. Simulating and Comparing Learning Rules

a. Perfect Memory

A univariate repeated measures ANOVA was ran for the Perfect Memory rule and resulted in statistical significance for subject and trial ($F_{(9,303)}=8.32$, $p<0.00000001$; $F_{(15,303)}=11.89$, $p<0.00000001$), but no statistical significance for the interaction between trial and subject.

Effect	Effect (F/R)	SS	Degrees of Freedom	MS	F	P
Intercept	Fixed	1.73291	1	1.732914	15.61214	0.00334
Bee	Random	0.99898	9	0.110998	8.32604	0.000000
Trial	Fixed	1.96385	14	0.140275	11.89447	0.000000
Bee*Trial	Fixed	1.48596	126	0.011793	0.16818	1.000000
Error		23.14071	330	0.070123		

Table 4. Perfect Memory Rule Univariate Repeated Measures ANOVA

b. Comparing Learning Rules

We chose the memory weight that best explained the data for the Relative Payoff Sum and the Linear Operator. Due to all the memory weight stability in the Linear Operator rule, we chose a memory weight of 0.8 for both learning rules. We then compared the rules to the Perfect Memory rule. Neither individual variation nor learning rule came out as significant. Yet there is a significant interaction between individual variation and learning rules ($F_{(18,30)}=9.44$, $p<0.00000001$).

A Tukey HSD Post-hoc Test showed that the Linear Operator and the Relative Payoff Sum don't differ from each other, but both are significantly different to

the Perfect Memory Rule. Additionally, when analyzing each rule based on trials, there is no statistical significant difference in any of the 16 blocked trials.

Effect	Effect (F/R)	SS	Degrees of Freedom	MS	F	P
Intercept	Fixed	1.73291	1	1.732914	15.61214	0.003349
Bee	Random	0.99898	9	0.110998	0.31516	0.957730
Rule	Fixed	2.24418	2	1.122091	2.95051	0.077930
Trial	Fixed	1.96385	14	0.140275	11.89447	0.000000
Bee*Rule	Random	6.84548	18	0.380304	9.43856	0.000000
Bee*Trial	Random	1.48596	126	0.011793	0.29006	1.000000
Rule*Trial	Fixed	2.69216	28	0.096149	2.36480	0.000243
Bee*Rule*Trial	Random	10.24586	252	0.040658	1.51739	0.085392
Error		0.80384	30	0.026795		

Table 5. Comparison of Learnings Rules Univariate Repeated Measures ANOVA

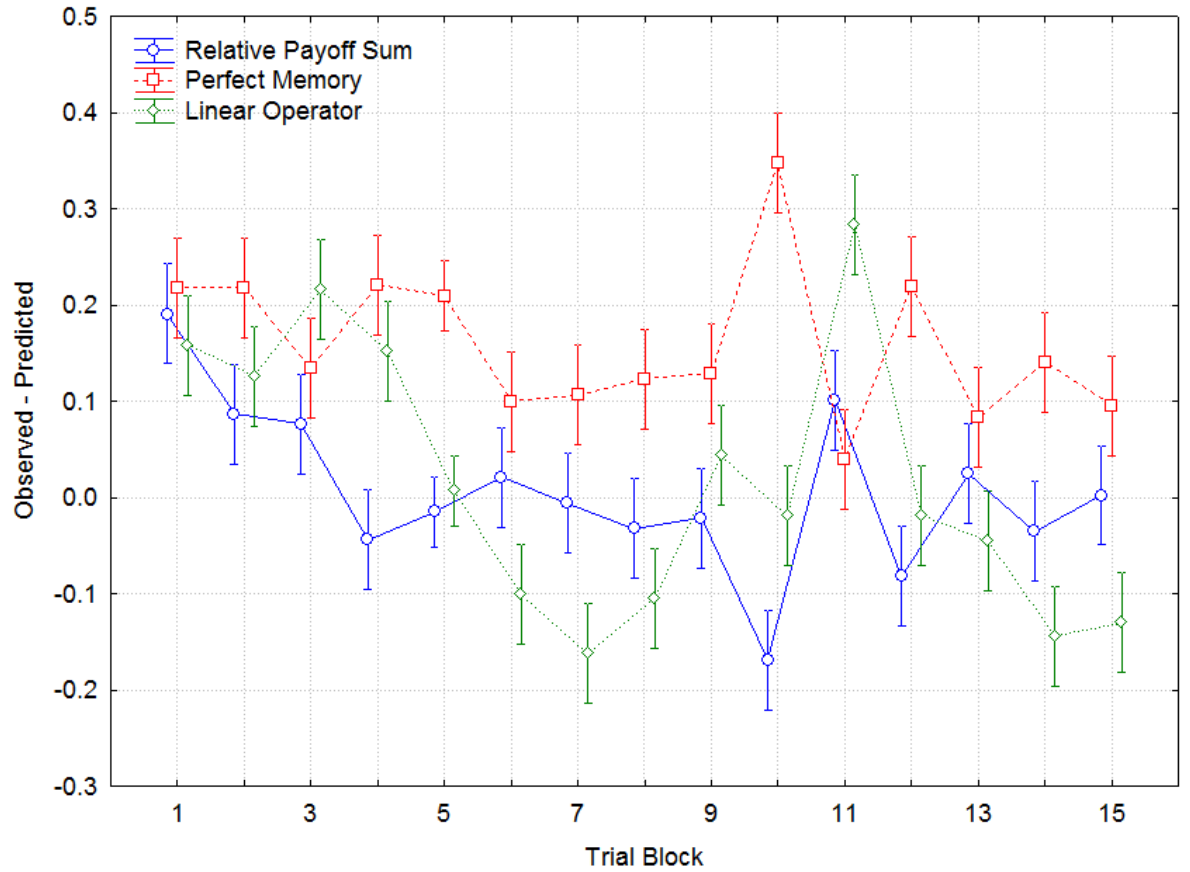


Figure 8. Consistency of the Learning Rules based on Number of Choices

The x-axis displays the number of choices made by the bees blocked in five choices for a total of 16 blocks of the 80 trial choices. By block four, 20 choices have been made and the learning rules are similar. From block four to block 12 there is more variation between the learning rules. After block 12, 55 choices have been made, the learning rules converge once more. The closer the values are to zero, the more the observed choices are following the predictions of the model. Negative values reflect that bees are choosing the predicted alternative less than expected. Positive values reflect that bees are choosing the predicted alternative more than expected.

i. Individual Variation in Learning Rules

Due to the statistical significance for the interaction of subject and the learning rule that best describes the choice patterns observed. By getting the difference of the observed to the predicted choice and comparing it to zero, which means that there is no difference between both values, we looked for values that were not

significant. As we are comparing the values to zero, any value that differs from zero assumes that the bee is not following the predictions of the model and the model does not best describe the behavior of the bee. We looked at the p-values for each bee for each learning rule and found that there is no statistical difference for any bee when using the Linear Operator rule. Subject 11 was the only statistically significant bee when evaluating the Relative Payoff Sum rule. Nine out of ten individuals showed no difference when using the Relative Payoff Sum rule. Only three individuals followed the predictions made by the Perfect Memory rule, the other six individuals were statistically significant for the Perfect Memory rule. Individuals showed a consistent trend in statistically significant values in all the learning rules for both of the resource types.

D. Modeling and Comparing Learning Rules

We used a deviance information criterion for the Bayesian model selection. This criterion allows us to examine which model best fits the observed data. As the deviance information criterion increases, the better the model. We found that a linear operator that uses a memory window, α_n , has a DIC of 1043. A linear operator model with a constant alpha and constant has r_i a DIC of 1051. The relative payoff sum with individual r_i and a constant alpha for every individual had a DIC of 900. A relative payoff sum with a memory window, α_n , and common r_i had a DIC of 912. When the alpha remains constant and the r_i varies for each individual the DIC is 891. When both the r_i vary and the memory windows are used the DIC is 886. For the perfect memory model with constant preferences, r_i , the DIC is 896. The perfect memory model with different r_i for each individual bee, we observe a DIC of 893.

E. Individual Differences

All learning rules displayed bees as being a statistically significant factor. Comparably, the Linear Operator rule ($F_{(9,1719)}=10.11$, $p<0.00000001$), the Relative Payoff Sum rule ($F_{(9,189)}=26.93$, $p<0.00000001$) and the Perfect Memory ($p=0.04$), which is why individual subjects were used as a random factor for the analysis done. Simulations showed higher variation per individual in the Perfect Memory rule in comparison to the remaining rules. In contrast, when modeling and comparing learning rules models with varying preferences for each bee displayed a large margin of error. Furthermore, the relative payoff sum demonstrated the largest margin of error due to its individual differences in color preferences and the memory weights.

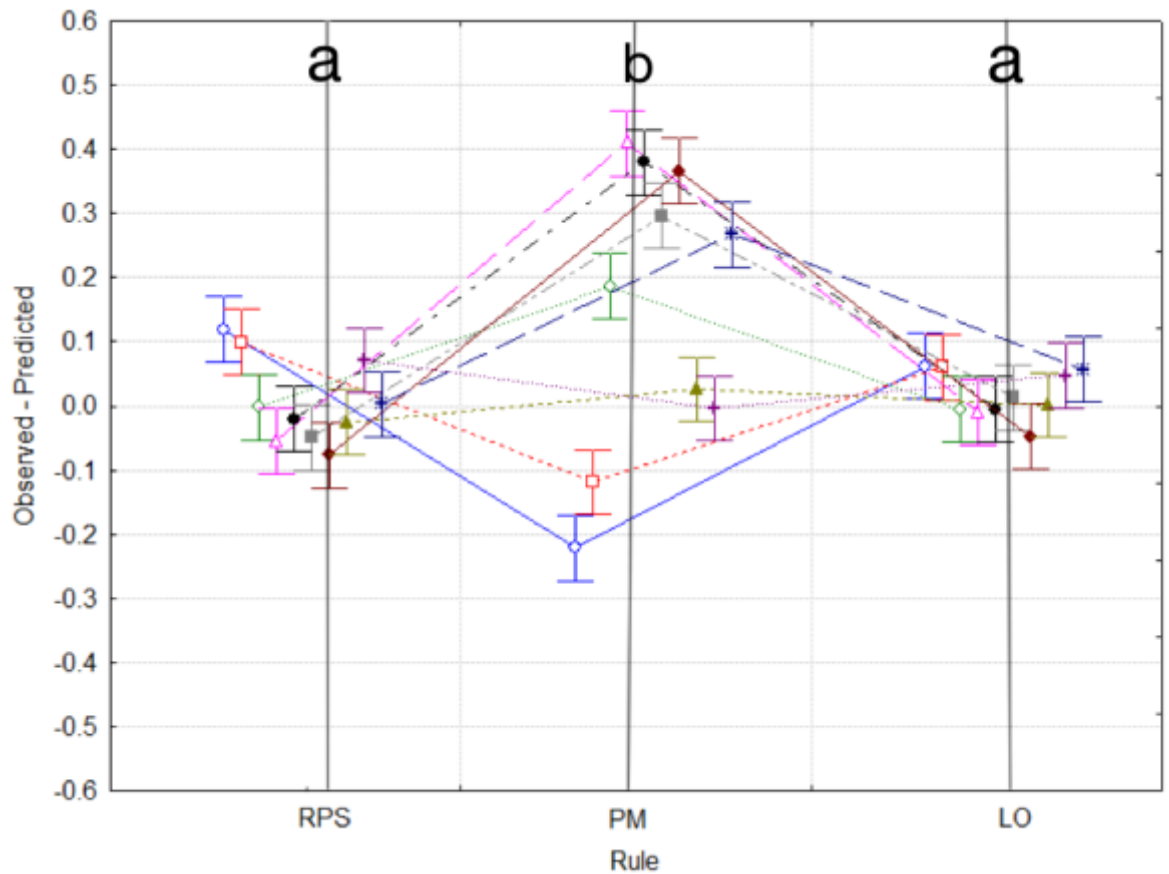


Figure 10. Individual Differences in Constancy to the Learning Rules

The x-axis displays the three different learning rules and the y-axis displays the difference between the observed choices and the predicted choices of the model. When the values are at zero, there will be no difference between the observed and the predicted choices. Negative values represent under-choosing the predicted resource and the positive values represent the bees choosing the predicted alternative more often than expected. Each line represents of one subject and how well its choice patterns are predicted by each learning rule. Perfect Memory displays more variation in comparison to the clustered groups in Relative Payoff Sum and Linear Operator.

IV. **Discussion**

Bees seem to be using more than just color preferences to make estimates about resources. Our data shows that when given the choice between placing more weight on past experiences or present observations, bees tend to place more weight on past choices. When simulating the Linear Operator rule bees will not really care about the weight they are placing on their memories, but instead modulate the strength they place on past experiences depending on the number of choices they've made. In other words, memory seems to vary with the number of choices when employing the Linear Operator rule. For the Relative Payoff Sum rule, we observed that bees should place more weight on past experiences. Theoretically, choice number will not matter in this case since the bees are already preferring to base their current estimates based on past estimates, which is a summary of all past choices. Furthermore, we see that it is very unlikely that bees are making estimates solely on past experiences. We propose that foraging bees are not likely to give equal weight to past and new information as suggested by the Perfect Memory rule and they should not only use past information as they may not have past information.

Contrary to our initial predictions, bees do not seem to be using a memory window to make decisions when using the Relative Payoff Sum rule. On the other hand, the Linear Operator factors in the possibility that foragers are modifying the strength they place on past experience depending on the number of choices they've made similar to a memory window. Still, our findings show us that bees show distinct individual variation for which learning rule best describes their choice patterns. There may be the possibility that individual foragers are

employing different parameters different times. Our analysis suggest that bees are using the parameters suggested by the Linear Operator rule, but some bees may also be incorporating color preference in their estimations and, thus, using the parameters predicted by the Relative Payoff Sum rule. Though Perfect Memory rule seems to be the rule that bees should be using, our simulations indicate that most of our bees are not using the parameters set by this learning rule.

Up to now, preference for color has been the only thing that has been considered in an animals decision-making process with the use of Bayes' Theorem. Our data suggests that preference for color is indeed important in a foragers decision to the point that there is great variation between individuals. Individual variation seems to be prevalent for color preference and even in the preference for the use of past or present information. For which reason, the Perfect Memory rule and the Relative Payoff Sum rule will be more prone to error, particularly the later. Though very similar, we believe that Linear Operator does not require a fixed memory weight for the individuals, since it does not take into consideration a prior such as color preference. Though there is variation in color preference displayed in individuals, there seems to be a consensus for color preference throughout the bees. As we know, foraging bees are genetically related, that strengthens the idea that color preference is inherited. Under the assumption that color preference is inherited, we believe that the stable memory weight in the Relative Payoff Sum is being used as a way to protect a color preference from going extinct in a population. Memory weight will also place weight on present observations and thus other color preferences that the forager may have. Similar

to Bayes' Theorem, Relative Payoff Sum ensures that a foraging bee will give take into consideration an alternative given another alternative.

When exposed to an unchanging world or a stable resource, foraging bees display more variation than when exposed to a fluctuating environment. Naïve foragers seem to initially over-estimate the values of the resources, but decline once they've experienced a low reward. Soon after the resource fluctuates to a higher reward, bees are observed to increase their estimations of the rewards once more. Once the world changes back to low reward, the bees will regress to under-estimating the resources' true reward value. Our data suggests that foraging bees adopt a cyclical pattern of over- and under- estimating their resources depending on their last experience.

Future studies should focus on individual variation during decision-making processes. We also encourage work done on the individual variation between the use of information sources, as our work demonstrated that some individuals place different weights on past and present experiences. Our own work will now focus on the use of learning rules when exposed to different information sources, where we would expect a larger difference between the Relative Payoff Sum and the Linear Operator as new priors will need to be taken into consideration. Finally, we believe that novel information sources may affect the memory weight as the current observation may now have more weight over past experiences.

REFERENCES

- Beauchamp, G. U. Y. "Learning rules for social foragers: implications for the producer–scrounger game and ideal free distribution theory." *Journal of Theoretical Biology* 207.1 (2000): 21-35.
- Bicca-Marques, Júlio César. "The win-stay rule in foraging decisions by free-ranging titi monkeys (*Callicebus cupreus cupreus*) and Tamarins (*Saguinus imperator imperator* and *Saguinus fuscicollis weddelli*)." *Journal of Comparative Psychology* 119.3 (2005): 343.
- Biernaskie, Jay M., Steven C. Walker, and Robert J. Gegear. "Bumblebees learn to forage like Bayesians." *The American Naturalist* 174.3 (2009): 413-423.
- Blumstein, Daniel T., and Amos Bouskila. "Assessment and decision making in animals: a mechanistic model underlying behavioral flexibility can prevent ambiguity." *Oikos* (1996): 569-576.
- Bouskila, Amos, and Daniel T. Blumstein. "Rules of thumb for predation hazard assessment: predictions from a dynamic model." *American Naturalist* (1992): 161-176.
- Charnov, Eric L. "Optimal foraging, the marginal value theorem." *Theoretical population biology* 9.2 (1976): 129-136.
- Charnov, Eric, and Gordon H. Orians. "Optimal foraging: some theoretical explorations." (2006).
- Dall, Sasha RX, and Rufus A. Johnstone. "Managing uncertainty: information and insurance under the risk of starvation." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 357.1427 (2002): 1519-1526.
- Danchin, Étienne. "Avatars of information: towards an inclusive evolutionary synthesis." *Trends in ecology & evolution* 28.6 (2013): 351-358.

- Dechaume-Moncharmont, François-Xavier, et al. "The hidden cost of information in collective foraging." *Proceedings of the Royal Society of London B: Biological Sciences* 272.1573 (2005): 1689-1695.
- DeWitt, Thomas J., Andrew Sih, and David Sloan Wilson. "Costs and limits of phenotypic plasticity." *Trends in ecology & evolution* 13.2 (1998): 77-81.
- Dunlap, Aimee S., et al. "Why some memories do not last a lifetime: dynamic long-term retrieval in changing environments." *Behavioral Ecology* 20.5 (2009): 1096-1105.
- Dunlap, Aimee S., and David W. Stephens. "Components of change in the evolution of learning and unlearned preference." *Proceedings of the Royal Society of London B: Biological Sciences* (2009).
- Dunlap, Aimee S., and David W. Stephens. "Tracking a changing environment: optimal sampling, adaptive memory and overnight effects." *Behavioural processes* 89.2 (2012): 86-94.
- Eliassen, Sigrunn, et al. "Quantifying the adaptive value of learning in foraging behavior." *The American Naturalist* 174.4 (2009): 478-489.
- Emlen, J. Merritt. "The role of time and energy in food preference." *American Naturalist* (1966): 611-617.
- Fuller, Rebecca C., David Houle, and Joseph Travis. "Sensory bias as an explanation for the evolution of mate preferences." *The American Naturalist* 166.4 (2005): 437-446.
- Giurfa, M., et al. "Colour preferences of flower-naïve honeybees." *Journal of Comparative Physiology A* 177.3 (1995): 247-259.
- Groß, Roderich, et al. "Simple learning rules to cope with changing environments." *Journal of The Royal Society Interface* 5.27 (2008): 1193-1202.

- Hamblin, Steven, and Luc-Alain Giraldeau. "Finding the evolutionarily stable learning rule for frequency-dependent foraging." *Animal Behaviour* 78.6 (2009): 1343-1350.
- Harley, Calvin B. "Learning the evolutionarily stable strategy." *Journal of theoretical biology* 89.4 (1981): 611-633.
- Herrnstein, Richard J. "Relative and Absolute strength of Response as a Function of Frequency of Reinforcement1, 2." *Journal of the experimental analysis of behavior* 4.3 (1961): 267-272.
- Hill, Peggy SM, Patrick H. Wells, and Harrington Wells. "Spontaneous flower constancy and learning in honey bees as a function of colour." *Animal Behaviour* 54.3 (1997): 615-627.
- Houston, Alasdair I., and Brian H. Sumida. "Learning rules, matching and frequency dependence." *Journal of Theoretical Biology* 126.3 (1987): 289-308.
- Ings, Thomas C., Nigel E. Raine, and Lars Chittka. "A population comparison of the strength and persistence of innate colour preference and learning speed in the bumblebee *Bombus terrestris*." *Behavioral Ecology and Sociobiology* 63.8 (2009): 1207-1218.
- Iwasa, Yoh, Masahiko Higashi, and Norio Yamamura. "Prey distribution as a factor determining the choice of optimal foraging strategy." *American Naturalist* (1981): 710-723.
- Janetos, Anthony C., and Blaine J. Cole. "Imperfectly optimal animals." *Behavioral Ecology and Sociobiology* 9.3 (1981): 203-209.
- Lea, Stephen EG, et al. "The cognitive mechanisms of optimal sampling." *Behavioural processes* 89.2 (2012): 77-85.
- Lewis, Alcinda C. "Memory constraints and flower choice in *Pieris rapae*." *Science* 232.4752 (1986): 863-865.

- Keasar, Tamar, et al. "Bees in two-armed bandit situations: foraging choices and possible decision mechanisms." *Behavioral Ecology* 13.6 (2002): 757-765.
- Koops, Marten A. "Reliability and the value of information." *Animal Behaviour* 67.1 (2004): 103-111.
- Krebs, John R., Alejandro Kacelnik, and Peter Taylor. "Test of optimal sampling by foraging great tits." *Nature* 275.5675 (1978): 27-31.
- Krebs, John R., John C. Ryan, and Eric L. Charnov. "Hunting by expectation or optimal foraging? A study of patch use by chickadees." *Animal Behaviour* 22 (1974): 953IN3-964.
- MacArthur, Robert H., and Eric R. Pianka. "On optimal use of a patchy environment." *American Naturalist* (1966): 603-609.
- Mangel, Marc. "Dynamic information in uncertain and changing worlds." *Journal of Theoretical Biology* 146.3 (1990): 317-332.
- McNamara, John M., and Alasdair I. Houston. "Memory and the efficient use of information." *Journal of theoretical biology* 125.4 (1987): 385-395.
- McNamara, J. M., A. I. Houston, and E. J. Collins. "Optimality models in behavioral biology." *Siam Review* 43.3 (2001): 413-466.
- McNamara, John M., and Sasha RX Dall. "Information is a fitness enhancing resource." *Oikos* 119.2 (2010): 231-236.
- McNamara, John M., Richard F. Green, and Ola Olsson. "Bayes' theorem and its applications in animal behaviour." *Oikos* 112.2 (2006): 243-251.
- Mery, Frederic, and Tadeusz J. Kawecki. "The effect of learning on experimental evolution of resource preference in *Drosophila melanogaster*." *Evolution* 58.4 (2004): 757-767.

- Mookherjee, Dilip, and Barry Sopher. "Learning and decision costs in experimental constant sum games." *Games and Economic Behavior* 19.1 (1997): 97-132.
- Morand-Ferron, Julie, and Luc-Alain Giraldeau. "Learning behaviorally stable solutions to producer–scrounger games." *Behavioral Ecology* 21.2 (2010): 343-348.
- Naug, Dhruva, and H. S. Arathi. "Sampling and decision rules used by honey bees in a foraging arena." *Animal cognition* 10.2 (2007): 117-124.
- Olsson, Ola, and Joel S Brown. "The foraging benefits of information and the penalty of ignorance." *Oikos* 112.2 (2006): 260-273.
- Olton, David S., and Peter Schlosberg. "Food-searching strategies in young rats: Win-shift predominates over win-stay." *Journal of Comparative and Physiological Psychology* 92.4 (1978): 609.
- Papaj, Daniel R., and Alcinda C. Lewis, eds. *Insect Learning: Ecology and Evolutionary Perspectives*. Springer Science & Business Media, 2012.
- Parker, Geoffrey A., and J. Maynard Smith. "Optimality theory in evolutionary biology." *Nature* 348.6296 (1990): 27-33.
- Platt, Michael L., et al. "Differences in feeding ecology predict differences in performance between golden lion tamarins (*Leontopithecus rosalia*) and Wied's marmosets (*Callithrix kuhli*) on spatial and visual memory tasks." *Animal Learning & Behavior* 24.4 (1996): 384-393.
- Pyke, Graham H. "Optimal foraging in bumblebees: rule of movement between flowers within inflorescences." *Animal Behaviour* 27 (1979): 1167-1181.
- Pyke, Graham H. "Optimal foraging theory: a critical review." *Annual review of ecology and systematics* 15 (1984): 523-575.

- Pyke, Graham H., H. Ronald Pulliam, and Eric L. Charnov. "Optimal foraging: a selective review of theory and tests." *Quarterly review of Biology* (1977): 137-154.
- Randall, Christopher K., and Thomas R. Zentall. "Win-stay/lose-shift and win-shift/lose-stay learning by pigeons in the absence of overt response mediation." *Behavioural Processes* 41.3 (1997): 227-236.
- Raine, Nigel E., and Lars Chittka. "The adaptive significance of sensory bias in a foraging context: floral colour preferences in the bumblebee *Bombus terrestris*." *PLoS One* 2.6 (2007): e556.
- Real, Leslie. *Behavioral mechanisms in evolutionary ecology*. University of Chicago Press, 1994.
- Reverdy, Paul, Vaibhav Srivastava, and Naomi Ehrich Leonard. "Satisficing in multi-armed bandit problems." *arXiv preprint arXiv:1512.07638* (2015).
- Sherratt, Thomas N. "The optimal sampling strategy for unfamiliar prey." *Evolution* 65.7 (2011): 2014-2025.
- Shettleworth, Sara J., et al. "Tracking a fluctuating environment: a study of sampling." *Animal Behaviour* 36.1 (1988): 87-105.
- Schmidt, Kenneth A., Sasha RX Dall, and Jan A. Van Gils. "The ecology of information: an overview on the ecological significance of making informed decisions." *Oikos* 119.2 (2010): 304-316.
- Schoener, Thomas W. "Theory of feeding strategies." *Annual review of ecology and systematics* (1971): 369-404.
- Scott, Steven L. "A modern Bayesian look at the multi-armed bandit." *Applied Stochastic Models in Business and Industry* 26.6 (2010): 639-658.

- Srivastava, Vaibhav, Paul Reverdy, and Naomi Ehrich Leonard. "Correlated multiarmed bandit problem: Bayesian algorithms and regret analysis." *arXiv preprint arXiv:1507.01160* (2015).
- Stephens, D. W. "On economically tracking a variable environment." *Theoretical Population Biology* 32.1 (1987): 15-25.
- Stephens, David W. "Decision ecology: foraging and the ecology of animal decision making." *Cognitive, Affective, & Behavioral Neuroscience* 8.4 (2008): 475-484.
- Stephens, David W., and John R. Krebs. *Foraging theory*. Princeton University Press, 1986.
- Stephens, David W., Joel S. Brown, and Ronald C. Ydenberg. *Foraging: behavior and ecology*. University of Chicago Press, 2007.
- Stephens, David W., and Eric L. Charnov. "Optimal foraging: some simple stochastic models." *Behavioral Ecology and Sociobiology* 10.4 (1982): 251-263.
- Thuijsman, Frank, et al. "Automata, matching and foraging behavior of bees." *Journal of Theoretical Biology* 175.3 (1995): 305-316.
- Trimmer, Pete C., et al. "Decision-making under uncertainty: biases and Bayesians." *Animal cognition* 14.4 (2011): 465-476.
- Valone, Thomas J. "Are animals capable of Bayesian updating? An empirical review." *Oikos* 112.2 (2006): 252-259.
- Winterhalder, Bruce. "Opportunity-cost foraging models for stationary and mobile predators." *American Naturalist* (1983): 73-84.