Working Paper Series

No. 03-2021

Making sense of monkey business: Re-examining tests of animal rationality

Roy Allen
University of Western Ontario
rallen46@uwo.ca

Pawel Dziewulski
University of Sussex
p.k.dziewulski@sussex.ac.uk

John Rehbeck
Ohio State University
rehbeck.7@osu.edu

Abstract: This paper re-examines research that studies economic rationality using experimental data generated by nonhuman animals (e.g. rats, pigeons, monkeys, etc.). The standard experimental methodology to elicit choices from nonhuman animals allows a researcher to test three types of economic rationality: standard deterministic utility maximization, average choice rationality, and random utility maximization. Most of the research has evaluated whether animals satisfy average choice rationality. We describe the difference between these models and check each type of rationality on capuchin monkey data from Chen et al. (2006). We reject standard deterministic utility maximization, but cannot reject either average choice rationality or random utility maximization. This paper is the first to provide a statistical test for average choice rationality.

Making sense of monkey business: Re-examining tests of animal rationality[*]

Roy Allen

Department of Economics
University of Western Ontario
rallen46@uwo.ca

Paweł Dziewulski

Department of Economics
University of Sussex
P.K.Dziewulski@sussex.ac.uk

John Rehbeck

Department of Economics
Ohio State University
rehbeck.7@osu.edu

February, 2021

Abstract

This paper re-examines research that studies economic rationality using experimental data generated by nonhuman animals (e.g. rats, pigeons, monkeys, etc.). The standard experimental methodology to elicit choices from nonhuman animals allows a researcher to test three types of economic rationality: standard deterministic utility maximization, average choice rationality, and random utility maximization. Most of the research has evaluated whether animals satisfy average choice rationality. We describe the difference between these models and check each type of rationality on capuchin monkey data from Chen et al. (2006). We reject standard deterministic utility maximization, but cannot reject either average choice rationality or random utility maximization. This paper is the first to provide a statistical test for average choice rationality.

1   Introduction

Experiments on decision making by nonhuman animals are thought to provide evidence on whether violations of economic rationality by humans have evolutionary foundations. Many experiments on nonhuman decision making follow the early research paradigm of Kagel et al. (1975), Battalio et al. (1981), Kagel et al. (1981), Battalio et al. (1985), and Kagel et al. (1995), which examine choice consistency of rats and pigeons using revealed preference techniques on average choices (Afriat, 1967; Varian, 1982).[1] However, previous work does not distinguish between three types of economic rationality: standard deterministic utility maximization, average choice rationality, and random utility maximization. While this may seem like a subtle point, if researchers are trying to understand evolutionary foundations of human decision-making, then it is critical that the same type of economic rationality is being studied on both human and nonhuman animals.

We describe the choice elicitation method for nonhuman animals at a high level to better understand the different ways this data can be used to evaluate the different types of economic rationality. The choices elicited from experiments on nonhuman animals often involve many choices from a few different environments. From here, a researcher could treat the choices from an environment as distinct choices following Afriat (1967) and Varian (1982), or treat the many choices as a distribution. The literature following Kagel et al. (1975) has often implicitly treated the data as a distribution of choices and looked at rationality criteria on *average choices*. Thus, this effectively evaluates average choice rationality (Allen et al., 2021). However, when one has a distribution of choices it is also sensible to model choices generated by random utility maximization following McFadden and Richter (1990). Thus, there are essentially two choices for a researcher to make:

1. whether to process the raw data to a distribution of choices and;

2. which model of economic rationality to examine.

To the best of our knowledge, these modeling choices have not been compared in the literature. This is important because this research seeks to find an evolutionary foundation for human decision making, but so far the literature on human decision making and nonhuman animal decision making have been using different data. In particular, research evaluating rationality of humans has subjects make a single choice in each budget and check whether the data is consistent with standard deterministic utility maximization.[2]

---

[1]There are other papers that study rationality of animals, e.g., Shafir (1994), Chen et al. (2006), Latty and Beekman (2010), Lea and Ryan (2015), Krasheninnikova et al. (2018), and Natenzon (2019) study rationality of choice distributions from bees, capuchin monkeys, amoeboids, frogs, and parrots.

[2]For example, Andreoni and Miller (2002), Choi et al. (2007), Choi et al. (2014), among many others ask humans for a single choice from each budget and check whether data satisfies standard deterministic utility maximization.

Since there is only a single choice in these environments, there is no difference between examining a distribution of choices or a single choice.

While the above conceptual distinctions are important for interpretation, this paper hopes to be a guide for researchers examining nonhuman rationality. First, we describe the standard method to elicit choices from nonhuman animals. After this, we define the dataset collected in these experiments and describe how to check whether the data is consistent with standard deterministic utility maximization (Afriat, 1967; Diewert, 1973; Varian, 1982), average choice rationality (Allen et al., 2021), and random utility maximization (McFadden and Richter, 1990; McFadden, 2005; Hoderlein and Stoye, 2015; Kitamura and Stoye, 2018). While average choice rationality has been examined in the literature, until this paper there was no statistical test. Finally, we provide an illustrative example checking whether the data on choices of capuchin monkeys from Chen et al. (2006) can be described by any of the models of economic rationality. We reject standard deterministic utility maximization, but cannot reject either the mean choice model or random utility. Thus, the choices from Chen et al. (2006) can be viewed as an evolutionary foundation for preferences over distributions of choices. We discuss the implications of this in Section 6.

The remainder of the paper is organized as follows. Section 2 explains the experimental methodology used to elicit choices from nonhuman animals and defines the data collected. Section 3 describes standard deterministic utility maximization, average choice rationality, and random utility maximization. Section 4 presents a statistical test of the average choice rationality. In Section 5 check whether the data from Chen et al. (2006) satisfies any of the three models of rationality and perform the statistical test developed in this paper. Section 6 provides our final remarks.

2 Experimental methodology and collected data

Most experiments examining rationality in animals look at choices from two linear budget sets with two goods. We will simplify the notation within the paper to account for this case. Thus, a consumption bundle is denoted $x \in \mathbb{R}^2_+$. For example, consumption good one could be apple slices while consumption good two could be grapes. These are two consumption goods from Chen et al. (2006).

The type of linear budgets often used in experiments is given in Figure 1. Here there is some "price" that defines a trade-off rate between these goods. For example, this trade-off rate could be defined by tokens that were traded to a human, as in Chen et al. (2006), or the number of times needed to interact with a button to deliver food as in Kagel et al. (1975). The relevant trade-off rates between the two goods are summarized by normalized prices $p \in \mathbb{R}^2_{++}$. Thus, the quantity of bundles available to be purchased is defined by the

budget

$$B(p) = \{x \in \mathbb{R}^2_+ \mid p_1 x_1 + p_2 x_2 \leq 1\}.$$

In Figure 1, the set is represented by the budget line (i.e., the set of bundles (x_1, x_2) for which the expenditure $p_1 x_1 + p_2 x_2$ is equal to 1) and all the points below it. Throughout the paper, we assume the two budget lines under consideration intersect.[3]

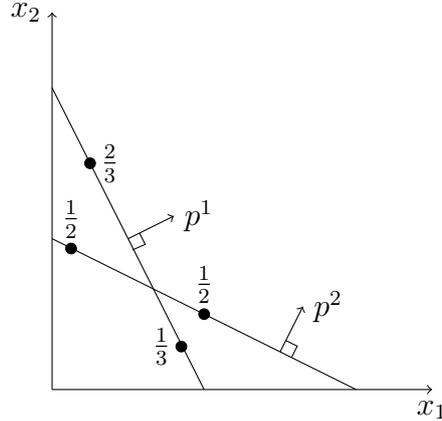Figure 1: Two overlapping budget sets with two consumption goods (x_1 and x_2). The consumption goods for animals are often food (e.g. apples and grapes, water and flavored water, etc.). The prices (p^1 and p^2) define a trade-off rate between two goods. Choices made on each budget line are in black and the proportion each bundle is chosen is next to it.

Unlike human experiments which describe the task to the subject, nonhuman animals cannot have the task explained to them. This has led researchers to have nonhuman animals make choices from a given budget several times. For example, a nonhuman animal may be asked to trade tokens for apples and grapes several times until the choices of the animal do not vary too much. While this is an *ex-ante* sensible thing to do, it also introduces the conceptual problem of trying to discern whether an individual has a preference over the distribution of their choices.

We denote the primitive dataset of choices in the two budgets by $\mathcal{D} = \{(p^1, x^{1,n})\}_{n=1}^{N_1} \cup \{(p^2, x^{2,m})\}_{m=1}^{N_2}$. Here p^1 is the vector of normalized prices from the first budget environment. The bundle $x^{1,n}$ is interpreted as the n-th repetition of the task with budgets defined by normalized prices p^1. The choices made from the second budget (under prices p^2) are indexed by $x^{2,m}$. The primitive dataset also admits a distributional dataset $\mathcal{D}_D = \{(p^1, \mu^1), (p^2, \mu^2)\}$ where μ^1 is the sample distribution of all choices from the first budget set. For example, given the first budget with the price p^1, the bundle x chosen $\mu^1(x) = 2/3$ of the time could result when $N_1 = 15$ and that bundle was chosen ten times. In general, we let the set of all distributions of choices in $\mathbb{R}^2_+$ be given by Δ.

---

[3] Unless the budget lines cross, none of the models we discuss in this paper produces any testable implications.

3 Models of rationality

This section defines the three relevant models of rationality.

3.1 Standard deterministic utility maximization

Utility maximization supposes that an individual makes choices that maximize some locally nonsatiated utility function defined by $u : \mathbb{R}^2_+ \to \mathbb{R}$ over the linear budget constraint. A utility function is defined as *locally nonsatiated* when there is always a better bundle "nearby."[4] Thus, we assume that the choice of bundles is generated by

$$\max_{x \in \mathbb{R}^2_+} \quad u(x)$$

$$\text{s.t.} \quad p_1 x_1 + p_2 x_2 \leq 1.$$

This model does not involve any distribution of choices and thus we work directly with the primitive dataset $\mathcal{D}$. For this model, one could only see two distinct bundles such as $x^{1,1} \neq x^{1,2}$ chosen from the same budget when the utility from the first bundle equals the utility of the second bundle.

The conditions to test utility maximization are well known from Afriat (1967) and Varian (1982). We record the results for the two budget case below.

Proposition 1. *A dataset* $\mathcal{D} = \left\{(p^1, x^{1,n})\right\}_{n=1}^{N_1} \cup \left\{(p^2, x^{2,m})\right\}_{m=1}^{N_2}$ *is consistent with deterministic utility maximization with local nonsatiation if and only if there are no observations* $\tilde{n} \in \{1, \dots, N_1\}$ *and* $\tilde{m} \in \{1, \dots, N_2\}$ *from the two budgets such that*

$$p_1^1 x_1^{2,\tilde{m}} + p_2^1 x_2^{2,\tilde{m}} \leq 1 \quad and \quad p_1^2 x_1^{1,\tilde{n}} + p_2^2 x_2^{1,\tilde{n}} \leq 1,$$

with one inequality strict.

To understand why this condition characterizes deterministic utility rationality, note that if $p_1^1 x_1^{2,\tilde{m}} + p_2^1 x_2^{2,\tilde{m}} \leq 1$, then this bundle could have been purchased at prices p^1 so that it reveals $u(x^{2,\tilde{m}}) \leq u(x^{1,\tilde{n}})$. Similarly, the inequality $p_1^2 x_1^{1,\tilde{n}} + p_2^2 x^{1,\tilde{n}} \leq 1$ reveals $u(x^{1,\tilde{m}}) \leq u(x^{2,\tilde{n}})$. If either inequality is strict, then the revelation of utility is strict (by local nonsatiation). For example, if there are bundles $x^{1,\tilde{n}}$ and $x^{2,\tilde{m}}$ with the first inequality strict in Proposition 1, then the two inequalities reveal that $u(x^{2,\tilde{m}}) < u(x^{1,\tilde{n}}) \leq u(x^{2,\tilde{m}})$, which is not possible for any utility function.

---

[4]Formally, a utility function is *locally nonsatiated* when for any $x \in \mathbb{R}^2_+$ and any $\varepsilon > 0$ there exists $y \in \mathbb{R}^2_+$ with $||y - x|| \leq \varepsilon$ such that $u(y) > u(x)$, where $|| \cdot ||$ denote the Euclidean norm.

3.2 Average rationality

Average rationality is essentially deterministic utility rationalization over the average bundle chosen. This is the type of rationality that has been examined in Kagel et al. (1975), Battalio et al. (1981), Kagel et al. (1981), Battalio et al. (1985), Kagel et al. (1995), and Chen et al. (2006). Since the average is a property of the distribution of choices, we consider a model where an individual can have a preference for randomization. In particular, we consider when the individual chooses an optimal distribution with certain properties.

We let Δ be the set of distributions over consumption bundles, i.e., over $\mathbb{R}_+^2$.[5] Here we assume a utility function $U : \Delta \to \mathbb{R}$ only depends on the average consumption of the distribution. Thus, we can write $U\big(\mathbb{E}_\nu[x]\big)$ where $\mathbb{E}_\nu[x]$ is the expected (average) consumption bundle (a vector in $\mathbb{R}_+^2$) generated by the distribution of choices ν. We also assume that U is locally nonsatiated over the expected consumption bundles.

Here, it is assumed that the distribution of choice $\mathcal{D}_D$ is generated by

$$\max_{\nu \in \Delta} \quad U\big(\mathbb{E}_\nu[x]\big)$$
$$\text{s.t.} \quad \nu\left(\{x \in \mathbb{R}_+^2 \mid p_1 x_1 + p_2 x_2 \leq 1\right) = 1.$$

Our interpretation of this model follows Machina (1985). Here, we view an individual as having a most-preferred distribution in mind and the individual randomizes according to that distribution whenever they make a choice. The restrictions of this model are very similar to deterministic utility maximization, even though the budget constraint takes a different form and requires the distribution to place positive probability only on bundles within the budget set. In the following proposition we assume that μ^1, μ^2 are true distributions. We discuss sampling errors and statistical testing in Section 4.

Proposition 2. *A dataset $\mathcal{D}_D = \big\{(p^1, \mu^1), (p^2, \mu^2)\big\}$ is consistent with average rationality for a local nonsatiated utility U if and only if it is not true that*

$$p_1^1 \mathbb{E}_{\mu^2}[x_1] + p_2^1 \mathbb{E}_{\mu^2}[x_2] \leq 1 \quad and \quad p_1^2 \mathbb{E}_{\mu^1}[x_1] + p_2^2 \mathbb{E}_{\mu^1}[x_2] \leq 1$$

with one inequality strict.

This model is flexible enough so that not all observed choices need to satisfy the rationality criteria of deterministic utility maximization. However, it still requires that the individual satisfies the rationality criteria *on average*. Also, it suggests the subject may have distributional considerations in mind when making choices. See Allen et al. (2021) for a theoretical study of these models.

---

[5] For formal details on modeling distributions of consumption bundles, see Allen et al. (2021).

3.3 Random utility

We now define the random utility maximization following Block and Marschak (1960), McFadden and Richter (1990), McFadden (2005), Hoderlein and Stoye (2015), and Kitamura and Stoye (2018). The idea behind random utility models is that an individual's choices might be governed by a distribution over different preferences. For example, one could interpret these preferences as different moods.

To represent random utility models mathematically, let $\mathscr{U}$ be the space of strictly quasiconcave locally nonsatiated utility functions $\tilde{u} : \mathbb{R}_+^2 \to \mathbb{R}$. A random utility model generates choices in the following manner. First, there is a probability measure ρ over the space of functions $\mathscr{U}$ such that the distribution of choices $\nu \in \Delta$ satisfies

$$\nu(O) \;=\; \rho\Big(\big\{\tilde{u} \in \mathscr{U} : \operatorname{argmax}_{x\in\mathbb{R}_+^2 \,|\, p_1 x_1 + p_2 x_2 \leq 1} \tilde{u}(x) \in O\big\}\Big), \tag{1}$$

for any measurable set $O \subseteq \mathbb{R}_+^2$. The argmax set is a singleton since $\mathscr{U}$ consists of strictly quasiconcave functions. In other words, the probability of choosing a bundle in the set O is equal to the probability of drawing a utility function that is maximized over the budget at some point in the set O.

The method to test random utility models significantly differs from the deterministic utility model and average rationality. The paper of Hoderlein and Stoye (2015) shows that for the random utility model, one only needs to check conditions on the probabilities of making choices in certain regions. The regions of choice are shown in Figure 2. Here $R^{1|1}$ is the region on the first budget line that lies above the second budget line in Figure 2. More generally, we refer to $R^{r|t}$ as the r-th region of the t-th budget as defined in Figure 2. Since we assume that preferences are locally nonsatiated, the only relevant regions are on the budget lines.
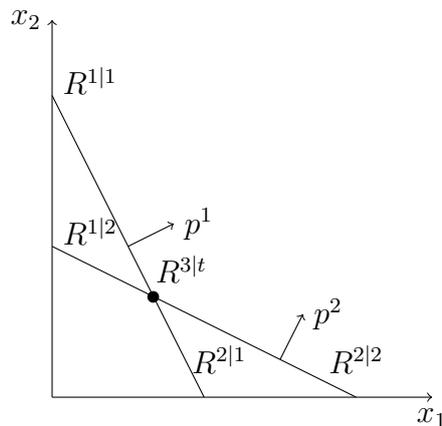

Figure 2: Relevant budget regions to examine random utility models with two budget sets and two goods

Hoderlein and Stoye (2015) show that random utility models can be checked by examining choice probabilities in these regions with a distributional data set $\mathcal{D}_D$. A similar result holds for higher dimensional consumption with multiple budgets as shown in Kitamura and Stoye (2018). The condition essentially requires that there cannot be a large proportion of violations of deterministic rationality. For an appropriately measurable set $S \subseteq \mathbb{R}_+^2$ and the probability measure $\nu \in \Delta$, we let $\nu(S)$ be the probability a choice is in the set S. We record the result from Hoderlein and Stoye (2015) below.

Proposition 3. *The distributional dataset $\mathcal{D}_D = \left\{(p^1, \mu^1), (p^2, \mu^2)\right\}$ is consistent with a random utility model if and only if $\mu^1(R^{1|1}) \geq \mu^2(R^{1|2})$ and $\mu^1(R^{3|1}) = \mu^2(R^{3|2})$.*

4 Statistical testing of average rationality

The method of inference we present is designed for application to the common datasets in nonhuman animal experiments mentioned throughout. Thus, we have two goods and two budget sets.[6] This empirical setting has several features that simplify statistical testing relative to a more general setup. First, prices and income are known exactly, so they do not need to be estimated. Second, there are several realized choices from each budget, which allows us to use the central limit theorem to justify a normal approximation of the means of the sampled distributions. Third, because we have two budgets, there is a single cycle to consider when checking average rationality. After describing the test, we apply the methods to the data from Chen et al. (2006) in Section 5.

We assume an analyst observes a collection of N_1 realizations from budget one $\{X^{1,n}\}_{n=1}^{N_1}$ and N_2 realizations of choices from budget two $\{X^{2,m}\}_{m=1}^{N_2}$. These are treated as random variables and we denote them in upper case. Each realization of the random variable is a vector of quantities, i.e., $X^{1,n} \in \mathbb{R}_+^2$. These realizations are draws from the appropriate distribution μ^t, which we interpret as the true choice. We interpret the population mean quantities $\mathbb{E}[X^{1,n}]$ as the mean of the distribution μ^1 and the population mean quantities $\mathbb{E}[X^{2,m}]$ as the mean of the distribution μ^2. For example, $\mathbb{E}[X^{1,n}] = \int x \, d\mu^1(x)$.

When $T = 2$, the null hypothesis of average rationality can be written as

$$
\begin{aligned}
H_0: \ & p^1 \cdot \mathbb{E}[X^{2,m}] \leq 1 \ \text{ implies } \ p^2 \cdot \mathbb{E}[X^{1,n}] \geq 1 \ \text{ and} \\
& p^2 \cdot \mathbb{E}[X^{1,n}] \leq 1 \ \text{ implies } \ p^1 \cdot \mathbb{E}[X^{2,m}] \geq 1.
\end{aligned}
$$

---

[6]To generate statistical tests for general situations, one can potentially use the results from Hsieh et al. (2018).

The alternative is the same as the statement in Proposition 2 and written as

$$H_a : p^1 \cdot \mathbb{E}[X^{2,m}] \leq 1 \quad \text{and}$$
$$p^2 \cdot \mathbb{E}[X^{1,n}] \leq 1 \quad \text{with one inequality strict.}$$

Recall that in the application, prices and income are measured without error, so we use lower case letters to indicate that they are nonrandom.

We reject the null hypothesis when sample averages are sufficiently far from the null. Formally, let the sample averages be given by

$$\overline{X}^1 := \frac{1}{N_1} \sum_{n=1}^{N_1} X^{1,n} \quad \text{and} \quad \overline{X}^2 := \frac{1}{N_2} \sum_{m=1}^{N_2} X^{2,m}.$$

First note that if $\overline{X}^t$ were nonrandom and equal to the true expectation, then we could reject the null hypothesis when

$$p^1 \cdot \overline{X}^2 \leq 1 \quad \text{and} \quad p^2 \cdot \overline{X}^1 \leq 1,$$

with at least one inequality strict. If we allow for sampling variability, the empirical means may be different from the theoretical ones. Thus, we propose a test of the form

$$\phi := \mathbf{1} \left\{ p^1 \cdot \overline{X}^2 \leq 1 - \hat{\sigma}^2 c_{\sqrt{1-\alpha}} \quad \text{and} \quad p^2 \cdot \overline{X}^1 \leq 1 - \hat{\sigma}^1 c_{\sqrt{1-\alpha}} \right\},$$

where $\mathbf{1}\{\cdot\}$ is the indicator function, $\phi = 1$ denotes rejection, and $\phi = 0$ denotes failure to reject. We denote the nominal size of the test by α (e.g. 0.05), and let $c_{\sqrt{1-\alpha}}$ be the $\sqrt{1-\alpha}$-quantile of the standard normal distribution. In addition,

$$\hat{\sigma}^1 := \sqrt{\frac{1}{N_1^2} \sum_{n=1}^{N_1} \left(p^2 \cdot X^{1,n} - p^2 \cdot \overline{X}^1 \right)^2} \quad \text{and} \quad \hat{\sigma}^2 := \sqrt{\frac{1}{N_2^2} \sum_{m=1}^{N_2} \left(p^1 \cdot X^{2,m} - p^1 \cdot \overline{X}^2 \right)^2},$$

are sample analogue estimators of the standard deviation of $p^2 \cdot \overline{X}^1$ and $p^1 \cdot \overline{X}^2$, respectively. The key assumption that justifies this test is that for $t = 1, 2$ there is a large number, N_t, of independent draws from each distribution μ^t.

To provide some intuition behind this test, note that each inequality in the argument of ϕ is motivated by a standard one-sided testing problem. Under the independence assumption and some mild conditions, we have the large sample approximation

$$\text{Prob} \left(p^1 \cdot \mathbb{E}[X^{2,m}] \leq p^1 \cdot \overline{X}^2 + \hat{\sigma}^2 c_{\sqrt{1-\alpha}} \right) \approx \sqrt{1-\alpha}.$$

Thus, on the event

$$p_1 \cdot \overline{X}^2 \leq 1 - \hat{\sigma}^2 c_{\sqrt{1-\alpha}},$$

or equivalently

$$p_1 \cdot \overline{X}^2 + \hat{\sigma}^2 c_{\sqrt{1-\alpha}} \leq 1,$$

we have strong evidence against

$$p^1 \cdot \mathbb{E}[X^{2,m}] \geq 1.$$

A similar argument holds when testing the other inequality. Thus, when the test function ϕ is 1, we have strong evidence against *both* inequalities

$$p^1 \cdot \mathbb{E}[X^{2,m}] \geq 1 \quad \text{and} \quad p^2 \cdot \mathbb{E}[X^{1,n}] \geq 1,$$

i.e., strong evidence against H_0. The choice of $c_{\sqrt{1-\alpha}}$ comes from the fact that we have to reject two inequalities. In more detail, this threshold is motivated by

$$\text{Prob}\left(p^1 \cdot \mathbb{E}[X^{2,m}] \leq p^1 \cdot \overline{X}^2 + \hat{\sigma}^2 c_{\sqrt{1-\alpha}} \quad \text{and} \quad p^2 \cdot \mathbb{E}[X^{1,n}] \leq p^2 \cdot \overline{X}^1 + \hat{\sigma}^1 c_{\sqrt{1-\alpha}}\right)$$
$$= \text{Prob}\left(p^1 \cdot \mathbb{E}[X^{2,m}] \leq p^1 \cdot \overline{X}^2 + \hat{\sigma}^2 c_{\sqrt{1-\alpha}}\right) \text{Prob}\left(p^2 \cdot \mathbb{E}[X^{1,n}] \leq p^2 \cdot \overline{X}^1 + \hat{\sigma}^1 c_{\sqrt{1-\alpha}}\right)$$
$$\approx \sqrt{1-\alpha}\sqrt{1-\alpha} = 1 - \alpha,$$

where the first equality follows from the assumption of independence between choices across budgets.

We note that the test we propose is not consistent against alternatives in which one of the inequalities is binding. That is, a configuration such as

$$p^1 \cdot \mathbb{E}[X^{2,m}] = 1 \quad \text{and} \quad p^2 \cdot \mathbb{E}[X^{1,n}] < 1$$

violates the null, but is not rejected with probability approaching 1 as the sample size increases.[7]

There are a few other features of this test to point out. First, note that if a deterministic test (e.g. using $\overline{X}^t$ in place of the expectations) is consistent with average rationality, then the statistical test fails to reject for any $\alpha < 0.75$. This is because if $\alpha < 0.75$ then

---

[7]This suggests a potential power issue with the design of Chen et al. (2006), which is constructed to detect violations of the compensated law of demand by pivoting the budget constraint. It is not a deficiency of the test we present. Intuitively, the alternative hypothesis is on the boundary of the null hypothesis, and so it is not possible to statistically distinguish it from nearby points in the null. One may report the minimal nominal size (described below) as a way to describe evidence against the null hypothesis.

$c_{\sqrt{1-\alpha}} > c_{0.5} = 0$.

One may not find the failure to reject an interesting way to differentiate between datasets. We suggest one way to differentiate between datasets in the spirit of the Afriat efficiency index (Afriat, 1973). Instead, one can compute the least nominal size α^* under which the null is rejected. To do this, one can compute the largest $c^*_{\sqrt{1-\alpha}}$ under which the test rejects the data and then find the corresponding α^*. The construction of α^* is related to a p-value but is distinct because there are configurations consistent with the null hypothesis in which rejection probabilities are not asymptotically equal to nominal size α.

High α^* gives less evidence against the null. To see why this is the case, note that if α^* is high ($\alpha^* > 0.75$), then $c_{\sqrt{1-\alpha^*}}$ is negative. Thus, this requires the data to not only satisfy the condition stated in Proposition 2 when failing to reject, but that the distance of the average bundles must also be far away from rejecting. The closer the α^* is to one, the lower the evidence is against the null since this suggests that the mean bundles are far apart.

5   Application to Chen et al. (2006)

We apply the above procedure to the data on choices by capuchin monkeys that was previously analyzed in Chen et al. (2006). We examine whether the capuchin monkeys are rational according to deterministic utility maximization, average rationality, and random utility models. Thus, we check the conditions of the earlier propositions.

The experiment was performed on three different capuchin monkeys. We refer to the three subjects under their abbreviations: AG, FL, and NN. The consumption bundles consisted of two goods: slices of apples and gelatin cubes or grapes (depending on the subject). In the experiment, capuchin monkeys traded tokens for food items under two different exchange rate regimes. In the first regime, one token could be exchanged for one slice of apple or one gelatin cube/grape. In the second regime, two slices of apple could be exchanged for one gelatin cube/grape. Per each subject, the data generated in the experiment consisted of multiple choices from the two budget sets. For additional information on the experiment, we refer the reader to Chen et al. (2006).

The details and results of the test are contained in Table 1. We note that two of the three subjects from Chen et al. (2006) refute utility maximization, while all three subjects could be described by a random utility model and average rationality. We note that all monkeys satisfy average rationality so the statistical test proposed earlier does not reject for any $\alpha < 0.75$. To try to better discern between the different datasets, we also report the least nominal size α^* that results in a rejection. We note that all α^* computed are indistinguishable from one. Thus, there is weak evidence against the hypothesis that

	Subject		
	AG	FL	NN
	Budget 1		
p^1	$(1/12, 1/12)$	$(1/12, 1/12)$	$(1/12, 1/12)$
N_1	12	11	6
$\bar{X}^1$	$(6.08,\ 5.92)$	$(5.64,\ 6.36)$	$(5,\ 7)$
	Budget 2		
p^2	$(1/18, 1/9)$	$(1/20, 1/10)$	$(1/20, 1/10)$
N_2	22	14	10
$\bar{X}^2$	$(9,\ 4.5)$	$(13.86,\ 3.07)$	$(12.8,\ 3.6)$
Deterministic	No	No	Yes
Random utility	Yes	Yes	Yes
Average choice	Yes	Yes	Yes
α^*	1.00	1.00	1.00

Table 1: Results of tests

subjects satisfy average rationality.

6   Conclusion

In this paper we compare the observable implications of three models of choice that can be used when studying economic rationality in nonhuman subjects: deterministic utility maximization, average choice rationality, and random utility maximization. We show that, given the primitive data generated in such experiments, it is possible to differentiate between the three models. We apply our results to the dataset in Chen et al. (2006) studying choices of capuchin monkeys and show that although deterministic utility maximisation is rejected for most subjects, the data is unable to distinguish between mean choice and random utility.

In addition to clarifying these distinct notions of rationality, this paper contributes to the literature on stochastic choice. In particular, the mean choice model follows the approach of Machina (1985) studying a preference over distributions. This perspective has become increasingly influential.[8] Thus, from the perspective of the stochastic choice literature, the evidence from the animal experiments of Kagel et al. (1975), Battalio

---

[8]For example, Swait and Marley (2013), Fudenberg et al. (2015), Freer and Martinelli (2016), Cerreia-Vioglio et al. (2019), and Allen et al. (2021) characterize different models where individuals express preferences over distributions. Similarly, there is evidence on human subjects in Sopher and Narramore (2000), Agranov and Ortoleva (2017), Agranov et al. (2020), Agranov and Ortoleva (2020), and Feldman and Rehbeck (2020) that also suggests individuals may prefer a distribution of choices.

et al. (1981), Kagel et al. (1981), Battalio et al. (1985), and Kagel et al. (1995) evaluating average rationality can also be interpreted as evidence for a preference over distributions. This complements an approach in Natenzon (2019) that studies stochastic choice for nonhuman subjects without studying an explicit preference over distributions.

References

Afriat, S. N. (1967). The construction of utility functions from expenditure data. *International economic review 8*(1), 67–77.

Afriat, S. N. (1973). On a system of inequalities in demand analysis: An extension of the classical method. *International economic review*, 460–472.

Agranov, M., P. J. Healy, and K. Nielsen (2020). Stable randomization. Working Paper.

Agranov, M. and P. Ortoleva (2017). Stochastic choice and preferences for randomization. *Journal of Political Economy 125*(1), 40–68.

Agranov, M. and P. Ortoleva (2020). Ranges of preferences and randomization. Working Paper.

Allen, R., P. Dziewulski, and J. Rehbeck (2021). Revealed statistical consumer theory. Working paper.

Andreoni, J. and J. Miller (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica 70*(2), 737–753.

Battalio, R. C., J. H. Kagel, and D. N. MacDonald (1985). Animals' choices over uncertain outcomes: Some initial experimental results. *The American Economic Review 75*(4), 597–613.

Battalio, R. C., J. H. Kagel, H. Rachlin, and L. Green (1981). Commodity-choice behavior with pigeons as subjects. *Journal of Political Economy 89*(1), 67–91.

Block, H. D. and J. Marschak (1960). Random orderings and stochastic theories of responses. *Contributions to probability and statistics*, 97–132.

Cerreia-Vioglio, S., D. Dillenberger, P. Ortoleva, and G. Riella (2019). Deliberately stochastic. *American Economic Review 109*(7), 2425–2445.

Chen, M. K., V. Lakshminarayanan, and L. R. Santos (2006). How basic are behavioral biases? evidence from capuchin monkey trading behavior. *Journal of Political Economy 114*(3), 517–537.

Choi, S., R. Fisman, D. Gale, and S. Kariv (2007). Consistency and heterogeneity of individual behavior under uncertainty. *American economic review 97*(5), 1921–1938.

Choi, S., S. Kariv, W. Müller, and D. Silverman (2014). Who is (more) rational? *American Economic Review 104*(6), 1518–1550.

Diewert, W. E. (1973). Afriat and revealed preference theory. *Review of Economic Studies 40*(3), 419–425.

Feldman, P. and J. Rehbeck (2020). Revealing a preference for mixing: An experimental study of risk. Working Paper.

Freer, M. and C. Martinelli (2016). A representation theorem for general revealed preference. GMU Working Paper in Economics.

Fudenberg, D., R. Iijima, and T. Strzalecki (2015). Stochastic choice and revealed perturbed utility. *Econometrica 83*(6), 2371–2409.

Hoderlein, S. and J. Stoye (2015). Testing stochastic rationality and predicting stochastic demand: The case of two goods. *Economic Theory Bulletin 3*(2), 313–328.

Hsieh, Y.-W., X. Shi, and M. Shum (2018). Inference on estimators defined by mathematical programming. *Available at SSRN 3041040*.

Kagel, J. H., L. Battalio, R. C. Battalio, L. Green, et al. (1995). *Economic choice theory: An experimental analysis of animal behavior*. Cambridge University Press.

Kagel, J. H., R. C. Battalio, H. Rachlin, and L. Green (1981). Demand curves for animal consumers. *Quarterly Journal of Economics 96*(1), 1–15.

Kagel, J. H., R. C. Battalio, H. Rachlin, L. Green, R. L. Basmann, and W. R. Klemm (1975). Experimental studies of consumer demand behavior using laboratory animals. *Economic Inquiry 13*(1), 22–38.

Kitamura, Y. and J. Stoye (2018). Nonparametric analysis of random utility models. *Econometrica 86*(6), 1883–1909.

Krasheninnikova, A., F. Höner, L. O'Neill, E. Penna, and A. M. von Bayern (2018). Economic decision-making in parrots. *Scientific reports 8*(1), 125–137.

Latty, T. and M. Beekman (2010). Irrational decision-making in an amoeboid organism: Transitivity and context-dependent preferences. *Proceedings of the Royal Society B: Biological Sciences 278*(1703), 307–312.

Lea, A. M. and M. J. Ryan (2015). Irrationality in mate choice revealed by túngara frogs. *Science 349*(6251), 964–966.

Machina, M. J. (1985). Stochastic choice functions generated from deterministic preferences over lotteries. *Economic Journal*, 575–594.

McFadden, D. and M. K. Richter (1990). Stochastic rationality and revealed stochastic preference. *Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz, Westview Press: Boulder, CO*, 161–186.

McFadden, D. L. (2005). Revealed stochastic preference: A synthesis. *Economic Theory 26*(2), 245–264.

Natenzon, P. (2019). Random choice and learning. *Journal of Political Economy 127*(1), 419–457.

Shafir, S. (1994). Intransitivity of preferences in honey bees: support for 'comparative' evaluation of foraging options. *Animal Behaviour 48*(1), 55–67.

Sopher, B. and J. M. Narramore (2000). Stochastic choice and consistency in decision making under risk: An experimental study. *Theory and Decision 48*(4), 323–350.

Swait, J. and A. Marley (2013). Probabilistic choice (models) as a result of balancing multiple goals. *Journal of Mathematical Psychology 57*(1), 1–14.

Varian, H. R. (1982). The nonparametric approach to demand analysis. *Econometrica*, 945–973.