(Ir)rational choices of humans, rhesus macaques, and capuchin monkeys in dynamic stochastic environments

Julia Watzek*, ᵃ & Sarah F. Brosnanᵃ, ᵇ

* Correspondence: j.watzek@gmail.com

ᵃ Department of Psychology, Language Research Center, Georgia State University
ᵇ Department of Philosophy, Neuroscience Institute, Center for Behavioral Neuroscience, Georgia State University

This is an author copy of the final manuscript published in *Cognition*.[1]

Abstract

Human and animal decision-making is known to violate rational expectations in a variety of contexts. Previous models suggest that statistical structures of real-world environments can favor such seemingly irrational behavior, but this has not been tested empirically. We tested 16 capuchin monkeys, 7 rhesus monkeys, and 30 humans in a computerized experiment that implemented such stochastic environments. Subjects chose among up to three options of different value that disappeared and became available again with different probabilities. All species overwhelmingly chose transitively ($A > B > C$) in the control condition, where doing so maximized overall gain. Most subjects also adhered to transitivity in the test condition, where it was suboptimal, but ultimately led to negligible losses compared to the optimal, non-transitive strategy. We used a modeling approach to show that differences in temporal discounting may account for this pattern of choices on a proximate level. Specifically, when short- and long-term goals are valued similarly, near-optimal decision rules can map onto rational choice principles. Such cognitive shortcuts have been argued to have evolved to preserve mental resources without sacrificing good decision-making, and here we provide evidence that these heuristics can provide almost identical outcomes even in situations in which they lead to suboptimal choices.

1   Introduction

Organisms make thousands of choices every day, choosing where to go, what to eat, with whom to mate, and whether to run away or fight. How can individuals make good decisions, particularly when survival depends on it? In real-life situations, it can be difficult to assess what the best available option is, particularly when the outcomes of our actions are uncertain. For example, should a female mate with the current male or wait for a better opportunity, which risks going empty-handed? Should an animal expend energy moving toward a known but almost depleted food patch or look for a new one? Understanding why and under what circumstances we make

adaptive decisions – or fail to do so – is a fundamental question to human and animal behavior.

Given the complexity of natural environments, it is not surprising that individuals often do not act entirely optimally, but deviate from rational expectations in certain contexts, e.g., switching their preferences depending on their past experiences, the current choice set, how options are presented, or their energetic state.[2] While the optimal solution is sometimes obvious and does not require much thought, other times it may be difficult to discern or may even appear irrational. In this study, we explored how subjects would respond in the latter case. We tested whether three primate species would employ seemingly irrational decision rules in environments in which doing so was actually optimal.

Rational choice theory[3] provides a useful starting point to predict decision makers' choices based on the assumption that individuals assign absolute values for each option and use them to form consistent and ordered preferences in order to maximize their long-term gain (and minimize their long-term loss). Two important principles that follow from this assumption are the transitivity axiom (if an individual prefers Option A given the choice {A, B} and Option B in choice {B, C}, then it should prefer A in choice {A, C}) and the principle of independence of irrelevant alternatives (adding or removing inferior options to a choice set should not influence decisions; if an individual prefers Option A given the choice {A, B, C}, then it should also prefer A in choice {A, B}).

However, both humans and animals are known to sometimes violate even these most basic principles of rational choice, particularly when options vary along multiple dimensions.[4] For example, hoarding grey jays consistently violated transitivity when choosing among three foraging options that placed them at different risks of predation:[5] The jays preferred one raisin at low risk over two raisins at intermediate risk ($A > B$) and preferred those two raisins over three raisins at high risk ($B > C$), but preferred three raisins at high risk over one raisin at low risk ($C > A$). Similarly, hummingbirds' preferences between two artificial flowers that varied in volume and sugar concentration changed when a lower-quality decoy was introduced, violating the assumption of independence of irrelevant alternatives.[6] Such studies suggest that decision makers often assign relative rather than absolute values to the options available in their environment, which can lead to inconsistent choices. They do not, however, explain why we act irrationally under some circumstances but not others,[7] and what characterizes the circumstances in which we do.

Ecological rationality is one framework that has been put forward to explain such irrational behavior.[8] This view emphasizes that deci-

[2] Kahneman et al. (1982); Thaler (1992); Rosati & Stevens (2009)

[3] von Neumann & Morgenstern (2007)

[4] Shafir (1994); Waite (2001); Shafir et al. (2002); Bateson et al. (2002); Bateson et al. (2003); Latty & Beekman (2010); Parrish et al. (2015)

[5] Waite (2001)

[6] Bateson et al. (2002); Bateson et al. (2003)

[7] e.g., Schuck-Paim & Kacelnik (2002)

[8] e.g., Houston (1997); Houston & McNamara (1999); Houston et al. (2007a)

sion rules evolve in, and adapt to, complex natural environments. By this view, an individual that acts optimally (i.e., acts as if to maximize fitness) should violate rational choice principles in some environments. Here, we refer to such violations as *economically* irrational behavior that arises from inconsistent preferences. However, such behavior may be *ecologically* (or *biologically*) rational when the decision rule fits the structure of the environment.[9] For example, animals may fail to behave in a way that maximizes their short-term gains, acting as if conditions were worse than they actually are. If the environment fluctuates between mild and harsh conditions, then a bias toward behavior that is adaptive in harsh conditions may be favored (even if it is suboptimal in mild conditions) because doing well in poor environments often has a bigger impact on fitness outcomes in the long run.[10] Thus, economically irrational behavior is not necessarily suboptimal but can instead reflect optimal choices under specific sets of conditions. This approach has been useful to demonstrate that cognitive biases such as pessimism and contrast effects can be adaptive,[11] particularly in uncertain environments that vary over space and time (spatiotemporal heterogeneity), but tend to be more similar the closer they are in space and time (positive autocorrelation). This is because the options that an animal currently faces can provide information about likely future conditions in the environment, which can – and indeed should – affect the best strategy for the current choice. Indeed, both violations of transitivity and independence of irrelevant alternatives can be optimal under such conditions.[12]

Of course, this concept of rationality only becomes meaningful with respect to the specific properties of an environment. That is, the same decision rule may be ecologically rational in one setting but ecologically irrational in another. Models of ecological rationality can explain why animals with such evolved decision rules may violate economic principles in situations in which doing so is suboptimal (e.g., in deliberately simplified lab tasks), but they also allow us to construe environments in which animals *should* violate economic principles to behave optimally. Testing this explicitly is important in order to establish whether and how animals' proximate decision rules match those that are optimal at an ultimate level.

The purpose of this study was to empirically test the predictions of one such model by McNamara and colleagues (2014) on three primate species: humans, rhesus macaques (Old World monkeys), and capuchin monkeys (New World monkeys). These species are suitable to such a task because their decision-making strategies are likely to be adapted to environments with spatiotemporal, autocorrelated fluctuations (e.g., as a result of the monkeys' reliance on fruit[13]). Moreover, primates have demonstrated various degrees of statistical

[9] Kacelnik (2006)

[10] McNamara et al. (2011)

[11] The Modelling Animal Decisions Group et al. (2014)

[12] e.g., Houston et al. (2007b); Trimmer (2013)

[13] Oates (1986); MacLean et al. (2014)

learning and successfully solve cognitive tasks in which the outcomes of their decisions are determined probabilistically.[14]

In their model, McNamara et al. (2014) demonstrate that an individual that maximizes its long-term gain can exhibit violations of rational choice in a dynamic environment in which options are not always available. Instead, each of three foraging options varies over time and will probabilistically disappear if it is currently available, or reappear if it is currently unavailable. An individual is assumed to take some time to process the option that it chose and cannot make another decision during this time (but options continue to appear and disappear). That is, there is an opportunity cost associated with an individual's choice because better options may appear or unchosen options may disappear while it handles the current option. McNamara et al. identified specific cases in which the optimal decision rules either favor economically rational or irrational choices. We chose this model because it predicts violations of economic rationality regardless of state-dependence and therefore lent itself to implementation in a computerized task with our populations. (An animal's state, such as whether it is hungry or satiated, is known to influence decision-making,[15] but is often difficult to measure, manipulate, or control.[16])

For this study, we derived two different environments based on McNamara et al.'s (2014) model to assess whether primates would actually behave so as to maximize their gain and flexibly follow or break the rational choice principles. Subjects were tested with distinct option parameters that call for them to either violate (test condition) or follow (control condition) the transitivity principle to maximize long-term gain. (Note that the model also predicts violations of independence of irrelevant alternatives; however, because this principle is violated whenever transitivity is violated[17], we focus on transitivity alone.) If individuals act optimally, they should flexibly violate or adhere to the principle of transitivity in a way that is consistent with the statistical structure of the task setting. Thus, we predicted that subjects would adopt intransitive decision rules in the test condition, but transitive decision rules in the control condition.

We did not originally employ a modelling approach ourselves. However, to provide an initial test of a hypothesis that emerged from our results (see below), we subsequently extended McNamara et al.'s (2014) model to include a temporal discounting parameter to systematically assess how decision rules may change when future rewards are discounted (a well-known cognitive constraint on decision-making[18]). Specifically, we expected transitive decision rules to become optimal over a wider range of parameters when future rewards are valued less, relative to immediate rewards.

[14] e.g., Conway & Christiansen (2001); Parrish et al. (2014); Proctor et al. (2014)

[15] e.g., Caraco (1981)

[16] but see Schuck-Paim et al. (2004)

[17] McNamara et al. (2014); see also Fig. 4

[18] Stevens (2014)

2   Methods

2.1   Subjects

We recruited 30 undergraduate participants (26 female, 4 male, age: $M \pm SD = 21.33 \pm 4.66$, range: 18-40 years) from Georgia State University through an online system and gave them course credit for participation in the experiment. We also tested 16 capuchins monkeys (*Cebus [Sapajus] apella*, 9 female, 7 male, age: $M \pm SD = 14.06 \pm 5.37$, range: 7-28 years) and 7 rhesus macaques (*Macaca mulatta*, all male, age: $M \pm SD = 19.00 \pm 6.95$, range: 12-32 years) at the Language Research Center of Georgia State University. Our human and rhesus samples are biased towards one sex (in the case of humans because predominantly women signed up for this experiment, a bias that is true across our study pool, and in the case of rhesus because all of our rhesus are male), however because this model is based on foraging considerations there is no *a priori* reason to anticipate sex differences in either species.

Capuchin monkeys were socially housed in mixed-sex groups that they had been a part of for at least a decade. Social groups were housed in large indoor/outdoor enclosures and subjects separated voluntarily into attached testing boxes for cognitive and behavioral studies. Rhesus monkeys were individually housed with continuous auditory and visual access to other monkeys and, when possible, regular social periods with compatible partners. Water was available to the monkeys *ad libitum*, including during testing, and we never food deprived subjects for the purposes of testing. We provided all testing food in addition to their daily diet of vegetables, fruit, and primate chow.

2.2   General Procedure

Mathematical framework   Following McNamara et al. (2014), we derived the optimal decision rules using Markov decision processes (MDPs), which are commonly used to model sequential decision-making in situations in which outcomes are not under the full control of the decision maker; that is, when the state of the world is determined in part by chance.[19] The basic idea is that an individual can choose the action it takes in a given moment but does not know how the action will affect the environment. MDPs do not describe how individuals actually behave in these stochastic dynamic environments, but they do provide procedures to determine how an optimal decision maker *should* behave in order to maximize expected utility. The MDP employed here characterizes time in discrete units called time steps. Their actual duration is inconsequential for the model, but for

[19] Puterman (1994)

the computerized task, we arbitrarily set it to four seconds, a value chosen to result in overall trial and session durations common to computerized cognitive tasks.

Design   We tested each subject in two conditions with distinct statistical structures (Table 1) and optimal decision rules. Each of three options A, B, and C was associated with a reward, a time out, a probability of disappearing, and a probability of reappearing in the next time step. Not all options were always present. Rather, at the end of each time step, options that were currently available became unavailable (i.e., disappeared) with probability μ. Conversely, options that were unavailable became available again (i.e., reappeared) with probability λ. Options disappeared and reappeared independently from each other and independently from which option was chosen in a given trial (that is, for example, options that were chosen did not automatically disappear). The handling time of an option, equivalent here to a time out, reflected an opportunity cost because options could come and go during the time out that followed the previous choice but subjects could not make another choice. Therefore, choosing an option necessarily meant forgoing unknown future options.

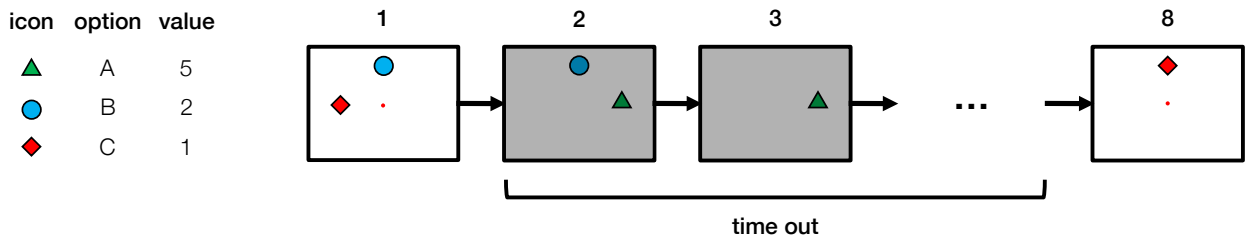| Option	Reward	μ	Control		Test	
			Timeout	λ	Timeout	λ
A	5	0.5	6	0.5	6	0.01
B	2	0.5	6	0.5	2	0.5
C	1	0.5	6	0.5	1	0.01

Table 1: Option parameters for control and test condition. The test condition is derived from Table 3 of McNamara et al. (2014). μ: probability that a currently available option will disappear in the next time step. λ: probability that a currently unavailable option will reappear in the next time step. Time outs are given as the number of four-second time steps.

The test condition adopted a case from McNamara et al. (2014). The options varied in their reward, their probability of reappearance, and their handling times (Table 1). An individual that acts optimally in this environment should violate transitivity. Option A should be chosen over Option B because neither option is likely to be available once handling is complete, and A yields the higher reward. Similarly, B should be chosen over C as it yields the bigger reward and, regardless of the choice, there may not be any options available afterward because both may have disappeared by the time handling is complete. Finally, and in violation of transitivity, C should be chosen over A. This is because 1) C has a higher rate of gain (1 reward per time step; rate of gain is calculated as the number of rewards an option yields divided by its handling time) than A (5/6 = 0.8 rewards per time step), and 2) even if C is not still available after handling, B (which also yields 2/2 = 1 reward per time step) is likely to reappear. Note that this circular strategy, in particular why C should be chosen

over A, is extremely counterintuitive, but is the decision rule that maximizes rewards in the mathematical model.

In the control condition, on the other hand, the three foraging options were identical except for their reward value (Table 1). Specifically, each option had the same 0.5 probability of reappearing and disappearing and a handling time of 6 time steps. This ensured that, unlike in the test condition, the options had the same opportunity cost. In this condition, an individual that maximizes its long-term gains should adhere to transitivity and choose options based only on their values: A should be chosen over B, B should be chosen over C, and A should be chosen over C. This control environment, in which the decision rules are both ecologically and economically rational, allowed us to establish that subjects do indeed adhere to transitivity when it maximizes gain.

Trial progression   The test and control condition used two distinct sets of icons of different shapes and colors to represent the three options A, B, and C on a computer screen. The assignment of icons to options was counterbalanced across subjects. Each time step, options were placed randomly in one of four possible positions on a white background (top, bottom, left, and right) to discourage side biases. Subjects continually made decisions by choosing between the visible options with a cursor (a solid red circle) that appeared in the middle of a color display at the beginning of each trial (Fig. 1).

icon	option	value
▲	A	5
●	B	2
◆	C	1

Figure 1: Schematic trial progression. In time step 1, the subject selects option B or C with the on-screen cursor (red dot) and gets rewarded. During the time out (here for 6 time steps), the subject cannot make another choice, but other choices may come and go: In time step 2, option C has disappeared but A has reappeared; in time step 3, option B has disappeared. This represents the opportunity cost associated with a decision. Afterward, in time step 8, the subject makes another decision.

Subjects could choose among all visible options during trials that lasted a maximum of four seconds (1 time step). Non-choice during a trial was considered a decision to do nothing for a unit of time. When one of the options was selected, rewards for that option (Table 1) were dispensed at a one-second rate, each accompanied by an auditory cue (a 'ding' sound). The decision was followed by a time out that represented the handling time of the selected option. This time out consisted of a number of four-second time steps depending on the option (Table 1). Subjects received their rewards during this time out, but were not be able to make another choice until after

the current handling time had ended (Fig. 1). (Note that the handling time occurs after rather than before subjects received the food reward. They are intentionally not part of the food acquisition but instead reflect the opportunity cost of a decision. This is because other options may disappear or reappear during the time out but are inaccessible to the subject during the delay.)

Probe trials   In the test condition, Options A and C had low probabilities of reappearing (Table 1, $\lambda = 0.01$) and were each only visible approximately every 100 time steps. As a result, critical trials, those in which at least two of the options A, B, and C were available at the same time, occurred very rarely by chance alone. Subjects were therefore presented with occasional probe trials at the end of their last session of the test condition, after they had considerable exposure to the task (see below), in order to better assess decision-making with respect to transitivity. In these sessions, probe trials for the critical choice sets appeared throughout the remaining session in a randomized order. Probe trials did not affect the choice set of the next trial; instead, the next choice set was determined by the reappearance and disappearance probabilities applied to the choice set available before the probe trial. For example, if the choice set before a probe trial was {A, B}, then for the choice set after the probe trial, A and B would each disappear with probability 0.5 and C would have the option to reappear with probability 0.01, regardless of the probe trial's choice set and which of those options was chosen.

Subjects' choices during probe trials were not reinforced because we wanted their reward histories to be unaffected by probe trials and only influenced by all of the other trials, which did adhere to the statistical structure described in Table 1. However, presenting probe trials did change the frequency with which options occurred and co-occurred in choice sets, especially those that included the rarest option C. For this reason, we presented them at the end of subjects' last test session after extensive exposure to the stochastic structure of the task and only presented a small number of probe trials to minimize their potential influence on subjects' choices (as most trials preceded the occurrence of probe trials). Nonetheless, this still allowed us to gain insight into which option subjects would choose in these critical choice sets if they did appear.

2.3   *Species-specific Procedures*

To accommodate the species-specific testing environments and to ensure subjects' motivation, the general procedure was adapted with respect to the hardware, rewards used, and testing schedule.

Humans   Humans made their choices by moving the cursor with the arrow keys on a keyboard. They received points that added to their total score displayed on-screen. After giving informed consent, participants received minimal instructions stating that they would make decisions by using the arrow keys on the keyboard. Participants participated in either the test or the control condition in one 90-minute session (1,350 time steps; between-subjects design). Twelve probe trials (3 each of {A, B}, {B, C}, {A, C}, and {A, B, C}) for the participants in the test condition were administered at random times after 45 minutes of exposure to the task.

Monkeys   Both capuchin and rhesus monkeys indicated their choices by moving the cursor by manipulating a joystick. They received banana-flavored food pellets as rewards (Fig. 2; for detailed information about the testing system, see ref. [20]).

[20] Evans et al. (2008)

Subjects participated in multiple sessions (*Mdn*: 8, *IQR*: 7-9 sessions) needed to complete 18,000 time steps (20 hours) in either condition, plus one additional session per condition (to present probe trials; see below). A session presented either the test or the control condition and monkeys alternated between the two conditions from session to session (within-subjects design). Sessions took place on different days, with a median of 2 days (*IQR*: 1-3 days) between test days. These monkeys are used to switching between tasks or conditions from day to day, trial sequences between the conditions differ markedly, and we used different icon sets to further distinguish the two conditions.

Figure 2: Experimental test setup for the monkeys. Subjects voluntarily separated from their home enclosure to attached test boxes, where they controlled an on-screen cursor with a joystick and received banana-flavored pellets as rewards.

Given that humans only experienced one session of the test condition, which included probe trials, we also presented monkeys with 12 unrewarded probe trials in their first session (3 each of {A, B}, {B,

C}, {A, C}, and {A, B, C}), after 60 minutes of exposure. We did this to better assess and compare decision rules used after limited early exposure to the task. In the last session of the test condition, after they had completed at least 18,000 time steps, we gave the monkeys 12 additional unrewarded probe trials (6 each of {A, C} and {A, B, C} – the choice sets least likely to occur by chance).

2.4   Data Analysis

Classifying decision rules   To evaluate the choices the subject made in response to a given choice set, we mapped each of the eight possible choice sets ({}, {A}, {B}, {C}, {A, B}, {A, C}, {B, C}, and {A, B, C}) to the option that was most frequently chosen ("preferred") by the subject. (As decisions tend not to be perfectly consistent, transitivity is often defined probabilistically.[21] Note that this criterion did not require a statistically significant difference in choices.) We then classified every subject's decision rule as transitive or intransitive and optimal or suboptimal. For instance, if an individual preferred A to B, preferred B to C, and preferred A to C, their decision rule would be transitive. It would be optimal in the control condition but suboptimal in the test condition. We used chi-square tests of independence to evaluate whether the number of observed decision rules in these categories differed significantly in the three species. We computed *p*-values using Monte Carlo tests[22] with 10,000 replicates.

[21] see Tversky (1969)

[22] Hope (1968)

Evaluating decision rules   We then evaluated each individual's decision rule (its eight choice-set-to-choice mappings, or policy, in a Markov decision process) against the optimal decision rule using fixed policy evaluation (see supplementary material). We compared the values of each subject's decision rule, which represent the expected utilities for each choice set, to the optimal values using Bayesian estimation (BEST)[23] to obtain credible values for the effect size and the difference in values between the optimal and actual decision rule for each subject. This procedure is similar to traditional *t*-tests but is able to accept the null value (given enough precision). Note that the values for a given decision rule could not, by definition, exceed the values of the optimal decision rule. The deviations from the optimal values were therefore always positive. Similar values between the optimal and actual decision rule (i.e., a difference near zero) suggest that a given decision rule tended to be optimal or led to little long-term loss, while average deviations larger than zero indicate that the decision rule was worse with respect to maximizing long-term gain than the optimal decision rule. Notably, while there is a single optimal decision rule that maximizes expected utility, subop-

[23] Kruschke (2013)

timal decision rules can nevertheless lead to long-term gains that are close to optimal. A non-zero but negligible difference between subjects' decision rules and the optimal decision rule may be reflective of decision heuristics.[24]

[24] Gigerenzer & Goldstein (1996)

3   Results

3.1   Observed Decision Rules

In their 90-minute sessions, humans made $M \pm SD = 299.60 \pm 28.49$ decisions in the control and $M \pm SD = 344.47 \pm 23.39$ decisions in the test condition. In the first 90 minutes, monkeys made $M \pm SD = 211.65 \pm 77.02$ decisions in the control and $M \pm SD = 241.83 \pm 94.54$ decisions in the test condition, with $M \pm SD = 2524.96 \pm 769.59$ total in the control and $M \pm SD = 3333.13 \pm 1035.31$ total in the test condition.

The frequency of subjects' observed decision rules (Table 2) varied significantly across species in the test condition, $\chi^2(4) = 9.34, p = 0.048$, but not in the control condition, $\chi^2(2) = 4.99, p = 0.134$. Almost all subjects in each of the three species found the optimal, transitive strategy in the control condition. In the test condition, capuchin monkeys were more likely to find the optimal, circular decision rule than either rhesus macaques or humans, standardized residual: $z = 2.00, p = 0.046$.

Species	Control		Test		
	Optimal Transitive	Suboptimal Other	Optimal Circular	Suboptimal Transitive	Suboptimal Other
Capuchin	16 (100%)	0 (0%)	5 (31%)	10 (63%)	1 (6%)
Rhesus	7 (100%)	0 (0%)	0 (0%)	5 (71%)	2 (29%)
Humans	12 (80%)	3 (20%)	0 (0%)	13 (87%)	2 (13%)

Table 2: Decision rules observed in the control and test condition. Capuchin and rhesus monkeys each participated in both conditions (within-subjects), whereas humans participated in either the control or the test condition (between-subjects).

3.2   Evaluation of Decision Rules

The results of the fixed policy evaluation (Table S1) suggests that those subjects who followed the suboptimal, transitive decision rule in the test condition (choosing $A > B > C$) did not incur significant losses (all 95% credible intervals for the effect size included zero). That is, the values for the transitive decision rule were not credibly different from the optimal values derived from the circular decision rule (choosing $A > B$, $B > C$, and $C > A$). However, the intervals are quite wide, indicating that a range of values are credible for the effect

size, and thus the precise extent to which the transitive and optimal values are similar is uncertain.

Only three other suboptimal decision rules emerged in the test condition (Table S1) and, strikingly, all led to much lower values compared to the optimal decision rule (all 95% credible intervals excluded zero, with a minimum credible effect size of 0.50 – typically considered a medium effect). Interestingly, all of these alternate decision rules are themselves transitive in that they form linear preference rankings: "choose A, avoid B" ($A > C > B$), "choose B, avoid A" ($B > C > A$), and "choose B, avoid C" ($B > A > C$). However, only the transitive decision rule that is optimal in the control condition ($A > B > C$) reflects a linear ordering by reward size ($5 > 2 > 1$).

The three humans who did not find the optimal transitive decision rule in the control condition followed a decision rule of choosing A over both B and C, and avoiding B (i.e., $A > B, C > B$, and $A > C$), which resulted in significantly lower long-term values than those obtained from the optimal strategy (Table S1, 95% credible intervals for effect size excluded zero, with a minimum credible effect size of 0.55). Interestingly, although they did show preferences between B and C, these participants often avoided making a choice at all when the biggest option (Option A) was absent. Human participants showed this preference on the population level as well, preferentially making choices when Option A was present, but tending to skip decisions that only presented Options B or C (Fig. S1).

3.3   Choices Over Time

The capuchin and rhesus monkeys had considerably more exposure to the task and therefore more time to experience the statistical environments. In the control condition, most monkeys quickly converged on the optimal transitive strategy, reliably choosing A over C (5 over 1) after approximately 780 time steps, A over B (5 over 2) by time step 4,300, and B over C (2 over 1) by time step 8,000 (Fig. S2, top row). Thus, length of acquisition nicely maps to the relative difference in value between the options. In the test condition, in which critical trials occurred much less frequently by chance alone than they do in the control condition (Fig. S1), most monkeys preferred A over B by time step 10,000 and B over C by time step 11,600 (Fig. S2, bottom row).

Remember that, in the test condition, the choice set {A,C} occurred very rarely because both options were unlikely to reappear ($\lambda = 0.01$, Table 1) when they were unavailable. Therefore, we only considered two time points (Fig. 3): the cumulative proportion of C choices after the first set of probe trials and after the second set of probe trials at the end of testing. Of the 7 rhesus monkeys, 2 initially chose in

line with the optimal, circular policy (C over A), while 4 did not (choosing A over C), and 1 chose A and C with equal frequency. However, most rhesus (6 out of 7) shifted away from the optimal decision rule (i.e., chose A at a higher proportion than C than they did at the beginning) and one showed no change. Of the 16 capuchin monkeys, 6 initially favored C over A and 10 favored A over C. Over time, the 10 capuchins who exclusively chose one over the other option continued to do so, while 3 chose more optimally and 3 chose less optimally (Fig. 3).
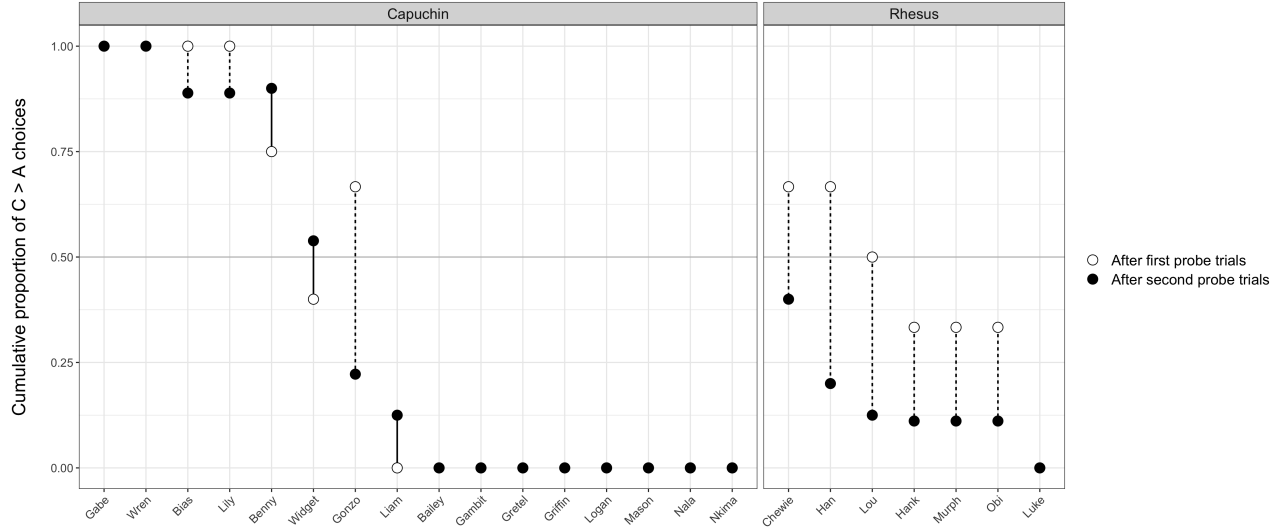

Figure 3: Cumulative proportion of choices over time for choice set {A, C} in the test condition after the first (open circles) and second set of probe trials (solid circles). Lines indicate whether a subject's proportion of optimal choices, C over A, increased (solid lines) or decreased (dashed lines).

4   Modeling Temporal Discounting

We had no *a priori* assumptions about how subjects would develop their decision rules, or about the underlying processes that led to particular selections. We adopted an undiscounted MDP (see supplementary material) to derive the optimal decision rule in the two conditions, and decisions were based both on the immediate reward of an action as well as on an estimated value of the next state. However, because there was no temporal discounting parameter, decisions were modelled to consider events that happen farther and farther in the future without discounting their value. This assumption sacrifices some ecological validity, as humans and animals alike are known to devalue future rewards, typically preferring smaller, immediate rewards over larger, delayed rewards.[25] Uncertainties inherent in the environment have, in fact, been hypothesized to select for such tendencies to discount the future and may serve to maximize gain.[26]

[25] Stevens (2014)

[26] Stephens et al. (2004); Stevens & Stephens (2010)
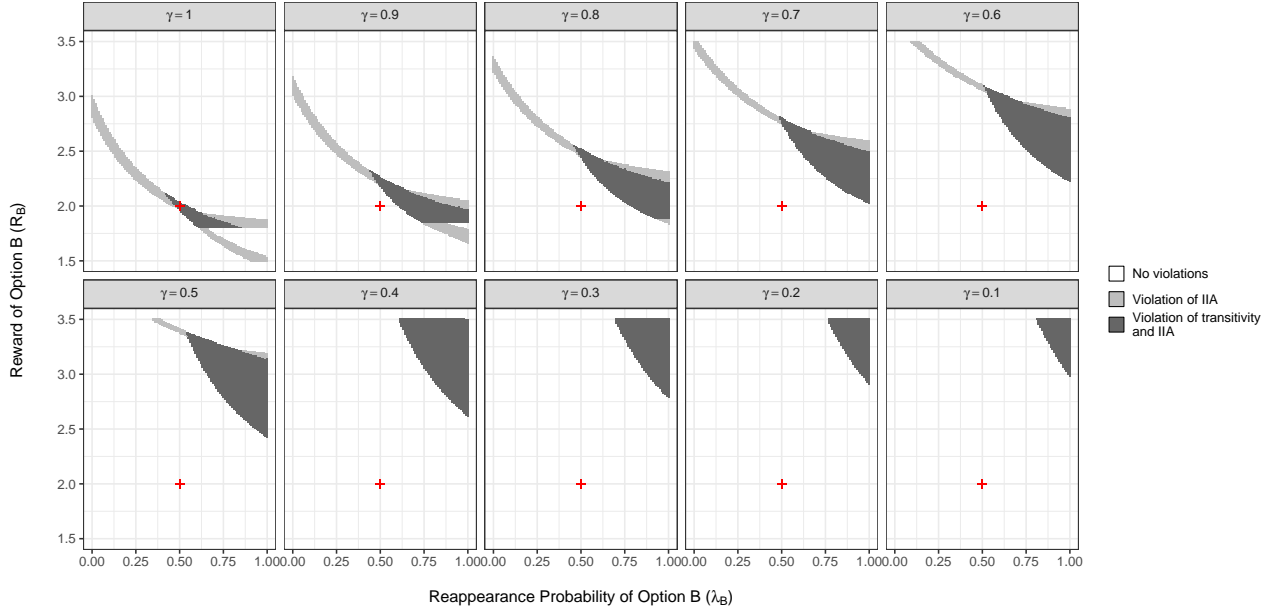
Therefore, the stochastic environments implemented here may have implicitly favored decision rules that maximize short-term gain, i.e. the transitive decision rule $A > B > C$.

To explore this possibility, we extended McNamara et al.'s model by introducing a temporal discounting parameter common to MDPs (Puterman 1994). This discount factor, γ, can vary between 0 and 1, and represents the relative weight of future and immediate rewards. In the original MDP, the γ is implicitly set to 1, i.e., future rewards are not discounted. We systematically varied the discount factor to assess its effect on the optimal decision rule for the specific case from McNamara et al. (2014) we used in the test condition (Table 1) and for a range of parameter values near it. Note that the discount factor could represent variations between as well as within species.

Figure 4 shows the regions of parameter combinations for which the optimal decision rule is to violate transitivity. When the future is not discounted ($\gamma = 1$; the case described by McNamara et al. (2014)), then the parameters from our test condition fall within this region. However, as the relative importance of future rewards compared to immediate rewards decreases, the distance from this region increases. In fact, with this set of parameters, the optimal decision rule switches between $\gamma = 0.9798$ (violate transitivity) and $\gamma = 0.9797$ (adhere to transitivity), indicating that even a small degree of temporal discounting can favor transitive decision rules in this case.

Figure 4: Parameter space with respect to the discount parameter (γ) and the reward (R_B) and probability of reappearance (λ_B) of Option B. Shaded regions indicate parameter combinations in which the optimal decision rule violates the rational choice principle of independence of irrelevant alternatives (IIA) (light grey) and transitivity (dark grey). $\gamma = 1$ indicates the undiscounted case adopted from McNamara et al. (2014) and used in this study. The crosshair marks the reward and reappearance probability of option B from the test condition; all other parameters were held constant at the values we used in that condition (Table 1).

Interestingly, even when the discount factor is low (i.e., the future is given little weight), for example when $\gamma = 0.1$, there are situations (parameter combinations) that call for violations of transitivity. This is the case when 1) the value of Option B increases and 2) Option B is very likely to reappear in the next time step when it is currently not available. The combination of these factors increases the expected reward from the future, therefore compensating for the discount.

5   Discussion

Successful decision rules readily emerged in the dynamic environments presented here, which implemented common stochastic characteristics of the real world. In the control condition, virtually all humans, rhesus macaques, and capuchin monkeys followed the optimal, transitive decision rule, which maximized both short- and long-term gains. In the test condition, only 5 of the 16 capuchins and none of the rhesus or humans found the optimal decision rule of violating transitivity, which maximized long-term but not short-term gains. Instead, most subjects again chose transitively with respect to reward size, which maximized short-term but not long-term gains. Notably, this transitive decision rule was the only suboptimal decision rule that consistently led to negligible overall losses compared to the optimal decision rule (if it had been followed). Other suboptimal strategies did, in fact, result in decreased long-term gains.

These results initially appear counter to models of ecological rationality, which would predict behavioral decision rules to match the statistical structure of the environment, leading to intransitive behavior in the test condition. These two accounts can be reconciled when we account for organisms' known tendency to devalue the future. In this case, the model also predicts adherence to transitivity even at small levels of temporal discounting. Extending the model in this way by no means invalidates it. Indeed, some subjects did employ the optimal, intransitive strategy and even the discounted case calls for violation of economic principles for some parameter combinations. Future research is needed to identify and examine those test cases that match test subjects' discounting rate. However, economically irrational behavior is only predicted to occur in certain circumstances and therefore may not face strong selection pressure, e.g., when errors are too rare or not costly enough to affect fitness.

Of course, not all violations of rational choice should be interpreted as being adaptive. According to models of bounded rationality,[27] suboptimal behavior may be the by-product of simple decision-making heuristics that typically perform well but sometimes fail. Transitivity could serve as a computationally inexpensive

[27] Simon (1955); Gigerenzer & Goldstein (1996); Hutchinson & Gigerenzer (2005)

mental shortcut because it only relies on knowledge of the reward size and does not require more sophisticated information processing to integrate factors such as an option's handling time and different probabilities of disappearing and reappearing. This study provides evidence that heuristics, such as choosing transitively, can indeed lead to near-optimal performance even when they are suboptimal. As individuals can only act within their cognitive restraints, it is arguably adaptive to preserve mental resources when possible, especially when there is no significant benefit to more complex decision mechanisms. Importantly, just knowing that a decision rule would be adaptive in a given environment (and therefore favored by evolutionary forces), does not tell us anything about the cognitive mechanisms (and their constraints) that an individual would use to actually make the decision at hand. Incorporating known cognitive limitations (from a proximate perspective) into evolutionary models based on fitness maximization (from an ultimate perspective), as we have done here, can help uncover the mechanisms involved in animals' decision-making and tease apart the situations in which economically irrational behavior is the result of adaptive decision rules and when it is not.

Behavioral decision rules may be prone to errors, and therefore be probabilistic rather than deterministic when used by an individual. This may similarly have influenced the decision patterns observed in this study, given that strategies can be unexpected with regard to transitivity when they are error prone but costly errors are rare.[28] Speaking to the latter point, in the test condition, the choice set {A, C}, which was critical to the classification of a decision rule as optimal (circular) or suboptimal (transitive), occurred very rarely by chance alone. Therefore, individual differences in cognitive abilities may explain some of the variability that capuchins, rhesus monkeys, and humans showed in the decision rules they adopted in both conditions. Better memory, better statistical learning abilities, or a greater ability to delay gratification may allow individuals to better maximize long-term over short-term gains and find the optimal decision rule in the test condition. Conversely, memory or learning constraints and lower ability to delay gratification may constrain individuals to suboptimal decision rules.

However, there is no evidence that capuchin monkeys, the only species who found the optimal strategy of violating transitivity in the test condition, generally exceed the memory, statistical learning, or self-control capabilities of humans and rhesus macaques; indeed, they often perform worse in tasks with near-identical methodology.[29] However, contrary to what is often the case, in this study, the monkeys actually made fewer decisions (i.e., worked more slowly)

[28] Houston (1997)

[29] Conway & Christiansen (2001); Tavares & Tomaz (2002); Beran & Parrish (2012); Beran (2015)

than the humans in the same timeframe. One possibility is that it may be easier to inhibit prepotent responses in a computer task, in which rewards are represented by icons, compared to tasks in which tangible rewards are visible and directly accessible. Of course, species other than primates may also have adapted to environments with spatiotemporal heterogeneity and positive autocorrelation and make adaptive decisions in other probabilistic contexts. Testing other species, particularly ones that differ more in their temporal discounting rates and cognitive architecture from humans,[30] would be invaluable in establishing how universal these decision rules may be.

[30] see e.g., Stevens & Hauser (2004)

Our model extension highlights that differences in how an organism *perceives* an option's value and statistical properties should affect the decision rule they exhibit. One other possibility is therefore that, in the test condition, Option B (the most frequently accessible option) was more salient to these capuchins, increasing the perceived value or perceived likelihood of it reappearing, which can favor intransitive decision rules even when future rewards are heavily discounted.

For the humans in the control condition, on the other hand, the most salient option appears to have been the biggest, most highly valued option (Option A). Indeed, on the population level, humans preferentially made decisions when the choice set included this option and tended to skip trials that only presented Options B or C. Note that non-choices (doing nothing for four seconds) could reflect a deliberate choice to wait for a better reward, but could just as well reflect lack of motivation or distraction. Given our interest in assessing decision rules with respect to transitivity, we focused on decisions in which they explicitly chose one of the three rather than to do nothing for a unit of time. And while our data does not allow us to distinguish between these two possibilities, it poses an interesting question for future research.

Humans were the only population that showed any suboptimal decision rules in the control condition. All three of these humans were female; however, given that the sample was predominantly female, we do not see this as evidence of sex differences at the population level. Perhaps more notably, there were also a few methodological differences between the experimental procedure for the monkeys and humans that may have impacted their responses differentially. Unlike the capuchin and rhesus monkeys, human participants received intangible rewards (virtual points displayed on-screen) rather than tangible rewards (such as food or money). The type of reward may have influenced their motivation. However, previous research found no difference in performance between real and facsimile money in a task that involved reward maximization under probabilistic uncertainty.[31] Moreover, most humans did manage to settle

[31] Bowman & Turnbull (2003)

on decision rules that reached or approached maximum long-term gains.

For practical reasons, human participants experienced the task for only a fraction of the time that the capuchin and rhesus monkeys did. This may have limited their ability to abstract the statistical properties of the environment and gain information about the options, which could favor the adoption of transitivity as a short-term maximizing heuristic. However, rhesus monkeys shifted away from (rather than towards) the optimal, circular decision rule over time and capuchin monkeys stuck to their initial strategies. One possibility is that transitivity is habitual, that is, organisms are likely to converge on it. Thus, while future research is needed, increased exposure alone may not necessarily make optimal decision rules more likely to be achieved.

Finally, the costs of suboptimal behavior in this study were relatively low, both in terms of the amount of reward and the number of missed opportunities. One way to address this issue would be to study scenarios in which the stakes are higher and errors are more costly. Specifically, this might be achieved by manipulating the reward sizes, increasing the handling times (and thereby opportunity costs), or by implementing an explicit choice to do nothing that requires a physical action rather than inactivity for a period of time. Further, decision rules are often not perfectly consistent (e.g., due to errors). In this study, we considered an option preferred if it was simply chosen more frequently; however, differences in the extent of these preferences could be modeled explicitly. More broadly, future work should evaluate the consequences of (economically) irrational behavior in the species' actual environments.

5.1   Conclusions

Taken together, the current study suggests that if the gap between short-term and long-term gains is not sufficiently large, or if future discounting is too substantial, additional mental resources may not be warranted (on a proximate level) and may therefore not change through natural selection (on an ultimate level). Heuristics such as transitivity have been argued to have evolved precisely to preserve cognitive resources without sacrificing good decision-making in most cases. This study provides empirical evidence that such cognitive shortcuts can indeed be optimal across a wide range of situations, can apply even more broadly when considering organisms' mental constraints (such as temporal discounting), and can lead to almost identical overall gains even in those situations in which other strategies are demonstrably better. However, violations of rational choice

principles can also be optimal even when such cognitive constraints are taken into account, and this finding provides ample opportunity for empirical tests in the future.

Ethical standard. This study was purely behavioral, non-invasive, and strictly adhered to the legal requirements of the United States of America. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. All applicable international, national, and institutional guidelines for the care and use of animals were followed. All procedures were approved by the Georgia State University Institutional Animal Care and Use Committee (IACUC capuchins: A13022, rhesus: A13020 and A16030) and the Institutional Review Board (IRB H14580). Georgia State University is fully accredited by the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC).

Supplementary material. The raw data underlying this article can be found at the Harvard Dataverse.[32] Supplementary data associated with this article can be found in the online version.[33]

[32] Watzek & Brosnan (2018b)
[33] Watzek & Brosnan (2018a)

References

Bateson, M., Healy, S. D., & Hurly, T. A. (2002). Irrational choices in hummingbird foraging behaviour. *Anim Behav*, *63*, 587–596. doi:10.1006/anbe.2001.1925.

Bateson, M., Healy, S. D., & Hurly, T. A. (2003). Context-dependent foraging decisions in rufous hummingbirds. *P Roy Soc B-Biol Sci*, *270*(1521), 1271–1276. doi:10.1098/rspb.2003.2365.

Beran, M. J. (2015). The comparative science of "self-control": What are we talking about? *Front Psychol*, *6*, 51. doi:10.3389/fpsyg.2015.00051/full.

Beran, M. J., & Parrish, A. E. (2012). Sequential responding and planning in capuchin monkeys (*Cebus apella*). *Anim Cogn*, *15*(6), 1085–1094. doi:10.1007/s10071-012-0532-8.

Bowman, C. H., & Turnbull, O. H. (2003). Real versus facsimile reinforcers on the Iowa Gambling Task. *Brain Cognition*, *53*(2), 207–210. doi:10.1016/S0278-2626(03)00111-8.

Caraco, T. (1981). Energy budgets, risk and foraging preferences in dark-eyed juncos (*Junco hyemalis*). *Behav Ecol Sociobiol*, *8*(3), 213–217.

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

Evans, T. A., Beran, M. J., Chan, B., Klein, E. D., & Menzel, C. R. (2008). An efficient computerized testing method for the capuchin monkey (*Cebus apella*): Adaptation of the LRC-CTS to a socially housed nonhuman primate species. *Behav Res Methods*, *40*(2), 590–596. doi:10.3758/BRM.40.2.590.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychol Rev*, *103*(4), 650–669.

Hope, A. (1968). A simplified Monte Carlo significance test procedure. *J R Stat Soc*, *30*(3), 528–598. doi:10.2307/2984263.

Houston, A. I. (1997). Natural selection and context-dependent values. *P Roy Soc B-Biol Sci*, *264*(1387), 1539–1541. doi:10.1098/rspb.1997.0213.

Houston, A. I., & McNamara, J. M. (1999). *Models of adaptive behaviour*. Cambridge: Cambridge University Press.

Houston, A. I., McNamara, J. M., & Steer, M. D. (2007a). Do we expect natural selection to produce rational behaviour? *Philos Trans R Soc B*, *362*(1485), 1531–1543. doi:10.1098/rstb.2007.2051.

Houston, A. I., McNamara, J. M., & Steer, M. D. (2007b). Violations of transitivity under fitness maximization. *Biol Lett*, *3*(4), 365–367. doi:10.1098/rsbl.2007.0111.

Hutchinson, J. M. C., & Gigerenzer, G. (2005). Simple heuristics and rules of thumb: Where psychologists and behavioural biologists might meet. *Behav Process*, *69*(2), 97–124. doi:10.1016/j.beproc.2005.02.019.

Kacelnik, A. (2006). Meanings of rationality. In S. Hurley, & M. Nudds (Eds.) *Rational animals?*, (pp. 87–106). New York: Oxford University Press.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York, NY: Cambridge University Press.

Kruschke, J. K. (2013). Bayesian estimation supersedes the *t* test. *J Exp Psychol Gen*, *142*(2), 573–603. doi:10.1037/a0029146.

Latty, T., & Beekman, M. (2010). Irrational decision-making in an amoeboid organism: Transitivity and context-dependent

preferences. *P Roy Soc B-Biol Sci*, *278*(1703), 307–312. doi:10.1098/rspb.2010.1045.

MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., et al. (2014). The evolution of self-control. *P Natl Acad Sci USA*. doi:10.1073/pnas.1323533111.

McNamara, J. M., Trimmer, P. C., Eriksson, A., Marshall, J. A. R., & Houston, A. I. (2011). Environmental variability can select for optimism or pessimism. *Ecol Lett*, *14*(1), 58–62. doi:10.1111/j.1461-0248.2010.01556.x.

McNamara, J. M., Trimmer, P. C., & Houston, A. I. (2014). Natural selection can favour 'irrational' behaviour. *Biol Lett*, *10*(1), 20130935–20130935. doi:10.1098/rsbl.2013.0935.

The Modelling Animal Decisions Group, Fawcett, T. W., Fallenstein, B., Higginson, A. D., Houston, A. I., Mallpress, D. E. W., et al. (2014). The evolution of decision rules in complex environments. *Trends Cogn Sci*, *18*(3), 153–161. doi:10.1016/j.tics.2013.12.012.

von Neumann, J., & Morgenstern, O. (2007). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Oates, J. F. (1986). Food distribution and foraging behavior. In B. B. Smuts, D. L. Cheney, & R. M. Seyfarth (Eds.) *Primate Societies*, (pp. 197–209). Chicago: University of Chicago Press.

Parrish, A. E., Brosnan, S. F., Beran, B. J., & Würsig, M. J. (2014). Differential responding by rhesus monkeys (*Macaca mulatta*) and humans (*Homo sapiens*) to variable outcomes in the Assurance Game. *Anim Behav Cogn*, *1*(3), 215. doi:10.12966/abc.08.01.2014.

Parrish, A. E., Evans, T. A., & Beran, M. J. (2015). Rhesus macaques (*Macaca mulatta*) exhibit the decoy effect in a perceptual discrimination task. *Atten Percept Psychophys*, *77*(5), 1715–1725. doi:10.3758/s13414-015-0885-6.

Proctor, D., Williamson, R. A., Latzman, R. D., de Waal, F. B. M., & Brosnan, S. F. (2014). Gambling primates: Reactions to a modified Iowa Gambling Task in humans, chimpanzees and capuchin monkeys. *Anim Cogn*. doi:10.1007/s10071-014-0730-7.

Puterman, M. L. (1994). *Markov decision processes: Discrete stochastic dynamic programming*. Hoboken, NJ: John Wiley & Sons.

Rosati, A. G., & Stevens, J. R. (2009). Rational decisions: The adaptive nature of context-dependent choice. In S. Watanabe, A. P. Blaisdell, L. Huber, & A. Young (Eds.) *Rational animals, irrational humans*. Tokyo: Keio University Press.

Schuck-Paim, C., & Kacelnik, A. (2002). Rationality in risk-sensitive foraging choices by starlings. *Anim Behav*, *64*(6), 869–879. doi:10.1006/anbe.2003.2003.

Schuck-Paim, C., Pompilio, L., & Kacelnik, A. (2004). State-dependent decisions cause apparent violations of rationality in animal choice. *PLoS Biol*, *2*(12), e402–11. doi:10.1371/journal.pbio.0020402.

Shafir, S. (1994). Intransitivity of preferences in honey bees: Support for 'comparative' evaluation of foraging options. *Anim Behav*, *48*, 55–67.

Shafir, S., Waite, T. A., & Smith, B. H. (2002). Context-dependent violations of rational choice in honeybees (*Apis mellifera*) and gray jays (*Perisoreus canadensis)*. *Behav Ecol Sociobiol*, *51*(2), 180–187. doi:10.1007/s00265-001-0420-8.

Simon, H. A. (1955). A behavioral model of rational choice. *Q J Econ*, *69*(1), 99–118. doi:10.2307/1884852.

Stephens, D. W., Kerr, B., & Fernandez-Juricic, E. (2004). Impulsiveness without discounting: the ecological rationality hypothesis. *P Roy Soc B-Biol Sci*, *271*(1556), 2459–2465. doi:10.1098/rspb.2004.2871.

Stevens, J. R. (2014). Evolutionary pressures on primate intertemporal choice. *P Roy Soc B-Biol Sci*, *281*(1786), 20140499–20140499. doi:10.1098/rspb.2014.0499.

Stevens, J. R., & Hauser, M. D. (2004). Why be nice? Psychological constraints on the evolution of cooperation. *Trends Cogn Sci*, *8*(2), 60–65. doi:10.1016/j.tics.2003.12.003.

Stevens, J. R., & Stephens, D. W. (2010). The adaptive nature of impulsivity. In G. J. Madden, & W. K. Bickel (Eds.) *Impulsivity: The Behavioral and Neurological Science of Discounting*, (pp. 361–388). Washington, DC: American Psychological Association.

Tavares, M. C. H., & Tomaz, C. (2002). Working memory in capuchin monkeys (*Cebus apella)*. *Behav. Brain Res.*, *131*(1-2), 131–137. doi:10.1016/S0166-4328(01)00368-0.

Thaler, R. H. (1992). *The winner's curse: Anomalies and paradoxes of economic life*. Princeton, NJ: Princeton University Press.

Trimmer, P. C. (2013). Optimal behaviour can violate the principle of regularity. *P Roy Soc B-Biol Sci*, *280*(1763), 20130858–20130858. doi:10.1098/rspb.2013.0858.

Tversky, A. (1969). The intransitivity of preferences. *Psychol Rev*, *76*, 31–48. doi:10.1037/h0026750.

Waite, T. A. (2001). Intransitive preferences in hoarding gray jays (*Perisoreus canadensis*). *Behav Ecol Sociobiol*, *50*(2), 116–121. doi:10.1007/s002650100346.

Watzek, J., & Brosnan, S. F. (2018a). (Ir)rational choices of humans, rhesus macaques, and capuchin monkeys in dynamic stochastic environments. *Cognition*, *178*, 109–117. doi:10.1016/j.cognition.2018.05.019.

Watzek, J., & Brosnan, S. F. (2018b). Replication data for: (Ir)rational choices of humans, rhesus macaques, and capuchin monkeys in dynamic stochastic environments. *Harvard Dataverse, V1*. doi:10.7910/DVN/RWPGQT.