

Ratio and difference comparisons of expected reward in decision-making tasks

DARRELL A. WORTHY, W. TODD MADDOX, AND ARTHUR B. MARKMAN
University of Texas, Austin, Texas

Several models of choice compute the probability of selecting a given option by comparing the expected value (EV) of each option. However, a subtle but important difference between two common rules used to compute the action probability is often ignored. Specifically, one common rule type, the *exponential* rule, compares EVs via a *difference* operation, whereas another rule type, the *power* rule, uses a *ratio* operation. We tested the empirical validity of each rule type by having human participants perform a choice task in which either the difference or the ratio between the reward values was altered relative to a control condition. Results indicated that participants can compare expected rewards by either ratio or difference operations but that altering the ratio between EVs produces the most dramatic changes in behavior. We discuss implications for several related research fields.

In experience-based choice tasks, such as the *n*-armed bandit task, a decision maker selects an option in order to maximize benefit and minimize cost. In the *n*-armed bandit task, the participant makes repeated choices from *n* different options, with the objective of maximizing total reward by selecting the options with the highest payoffs (Sutton & Barto, 1998). On each trial, the decision maker must decide whether to exploit the option that has been giving a good reward or to explore other options in order to gain new information about the environment. This task is analogous to many real-life decisions, such as whether to dine at a new restaurant or to stick with an option that has been successful in the past. Several of the learning models developed for this task assume that decision makers compare the expected value (EV) of each option when determining the response to select on the next trial (Busemeyer & Stout, 2002; Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006; Sutton & Barto, 1998; Worthy, Maddox, & Markman, 2007). These models use a form of the biased choice rule (Luce, 1959, 1963; Shepard, 1957) to compute the probability of selecting each option. The probability of selecting option *a* is equal to the EV for option *a* divided by the sum of the EVs for all possible options.

Action selection rules are ubiquitous in models of choice, and so it is important to understand their behavioral predictions and to determine whether human behavior conforms to the predictions of one selection rule or another. Action selection rules have been incorporated into models of choice behavior in animals (often in the form of the matching law; Corrado, Sugrue, Seung, & Newsome, 2005; Herrnstein, 1961; Lau & Glimcher, 2005; Sugrue, Corrado, & Newsome, 2004), similarity-based models of category learning (e.g., Maddox & Ashby, 1993; Medin & Schaffer, 1978; Nosofsky, 1986; Reed, 1972; Rodri-

gues & Murre, 2007), and many connectionist models of action selection (e.g., Kruschke, 1992; Minsky & Papert, 1968; Rumelhart, McClelland, & the PDP Research Group, 1986).

In this article, we focused on a subtle difference in the implementation of action selection rules in computational models that is largely ignored in the literature. Some models compare the *ratio* between the values representing each alternative via a power function, whereas other models compare the *difference* between the values representing each alternative via an exponential function. Despite this distinction between the models, decision rules are typically incorporated into models without explicitly choosing one rule over the other on the basis of its psychological properties. Although many researchers may understand the differences between a power rule and an exponential rule, there have been few systematic examinations of the effects of using either action selection rule in modeling applications (however, see Rodrigues & Murre, 2007). Furthermore, there have been few empirical tests of theoretical predictions derived from using either rule.

In this article, we compared these decision rules theoretically and empirically. We started by deriving behavioral predictions for power and exponential rules. To anticipate, the predictions from a power rule are unaffected as long as the ratio between the reward values remains unchanged, whereas the predictions from an exponential rule are unaffected as long as the difference between the reward values remains unchanged. Next, we reviewed the small number of studies that have examined the incongruities between these choice rules. Finally, we empirically tested the theoretical predictions of these models in two decision-making studies that used a variant of the *n*-armed bandit problem.

D. A. Worthy, worthyda@mail.utexas.edu

Ratio-Preserving Power Rules and Difference-Preserving Exponential Rules

Equation 1 depicts a ratio-preserving power rule. The probability of selecting option a on trial t is determined by comparing the ratio between the EV of choice a and the EV of choice b on trial t :

$$P_{a,t} = \frac{EV_t(a)}{EV_t(a) + EV_t(b)}. \quad (1)$$

In the Appendix, we show algebraically that the probability of selecting option a in Equation 1 is dependent on the ratio between the EVs of options a and b and, thus, is unaffected by scalar multiplication of the EVs. We refer to rules of this kind as *power rules*. Equation 1 has been referred to as the *matching law* in the animal behavior literature (Herrnstein, 1961). Models of animal behavior often use a power rule to compare the animal's response probabilities with the actual reward probabilities of each option (Herrnstein, 1961; Lau & Glimcher, 2005; Sugrue et al., 2004; Williams, 1988). Several popular models of category learning also use power rules to compute the probability of each categorization response—for example, the generalized context model (Nosofsky, 1986; Nosofsky & Zaki, 2002), the deterministic exemplar model (Ashby & Maddox, 1993; Maddox & Ashby, 1993), and the continuous-dimension version of RULEX (Nosofsky & Palmeri, 1998).

Another action selection rule often used in computational models is the exponential rule. This rule is often called a *softmax rule* in models of choice (Bridle, 1990; Sutton & Barto, 1998). Here, the probability of selecting option a on trial t is given by comparing the difference between the EV of choice a and the EV of choice b on trial t . An example of a model using this type of exponential action selection rule is

$$P_{a,t} = \frac{e^{[EV_t(a)]}}{e^{[EV_t(a)]} + e^{[EV_t(b)]}}. \quad (2)$$

In the Appendix, we show algebraically that the probability of selecting option a in Equation 2 is dependent on the difference between the EVs of options a and b and, thus, is unaffected by scalar addition to the EVs. This type of rule is often employed in models of choice for tasks like the n -armed bandit problem (e.g., Bussemeyer & Stout, 2002; Daw et al., 2006; Sutton & Barto, 1998; Worthy et al., 2007; Yechiam, Bussemeyer, Stout, & Bechara, 2005) and in other domains, including human foraging behavior (Roberts & Goldstone, 2006) and category learning (Kruschke, 1992; Love, Medin, & Gureckis, 2004).

Power and exponential rules make different assumptions about how decisions are made. For example, consider Decisions 1 and 2 with different EVs (represented in points) for options a and b :

Decision 1: $EV_t(a) = 6$; $EV_t(b) = 4$,

Decision 2: $EV_t(a) = 60$; $EV_t(b) = 40$.

Here, the EVs in Decision 1 have been multiplied by 10 to create the EVs for Decision 2, so that the ratio between the EVs remains the same, but the difference between the EVs is much greater in Decision 2 than in Decision 1 (i.e.,

2 vs. 20 for Decisions 1 and 2, respectively). Thus, a power rule yields the same probability of selecting option a in both decisions (.60), whereas an exponential rule yields different probabilities for selecting option a for Decisions 1 (.88) and 2 (.99).

In contrast, consider a third decision:

Decision 3: $EV_t(a) = 86$; $EV_t(b) = 84$.

In this case, the EVs from Decision 1 have been increased by 80 points, which maintains the same difference between the two EVs as in Decision 1 but changes the ratio so that it is different from that between the two EVs from Decision 1. The power rule now gives a probability of selecting option a of .51 for Decision 3, which is much smaller than the probability of .60 given by the same equation for Decision 1. However, the exponential rule gives the same probability for Decisions 1 and 3 (.88).

Recently, Rodrigues and Murre (2007) examined the effects of using a power rule in place of an exponential rule in several popular category-learning models by conducting model fits using both choice rules. They first fit data from Erickson and Kruschke's (2002) experiment using four models: ALCOVE (Kruschke, 1992), ATRIUM (Erickson & Kruschke, 1998), and two extensions of ALCOVE—exemplar-specific ALCOVE (ES-ALCOVE; Sakamoto, Matsuka, & Love, 2004) and ALCOVEc (see Rodrigues & Murre, 2007)—using the exponential rule used in the original formulations of ALCOVE and ATRIUM. When the exponential rule was used, they found that ATRIUM provided a better fit to the data than did any of the three versions of ALCOVE. They then considered using a power rule in place of the exponential rule for each model.

Rodrigues and Murre (2007) observed that the two action selection rules generally produced identical results, except when extremely small values (e.g., .01 and .001) were placed in the choice rules as activity values for each category node. In line with the examples presented above when .01 is used as EV_a and .001 is used as EV_b , the exponential rule gives a very low probability for selecting option a (.502), because of the small difference between the EVs (.009). In contrast, the power rule gives a much higher probability of selecting option a (.909), because the ratio between the EVs is much greater than the difference.

When Rodrigues and Murre (2007) fit the data from Erickson and Kruschke's (2002) experiment with the same four models, but with a power rule instead of an exponential rule, they found that fits to the data from all three versions of ALCOVE were improved. In contrast, fits of the data using ATRIUM were worse. Fits of ES-ALCOVE and ALCOVEc were better than fits of ATRIUM when a power rule was used in place of the exponential rule. A comparison between fits of ATRIUM using the exponential rule and fits of ES-ALCOVE and ALCOVEc using the power rule showed that all models yielded nearly identical fits, with ES-ALCOVE performing the best.

Thus, the conclusions drawn from fits of these models to the data were drastically different—indeed opposite—depending on which action selection rule was used. To our knowledge, the study described above is the only systematic examination of the differences between expo-

nential and power rules in the literature. These rules are ubiquitous in models across a variety of domains, yet they are seemingly used interchangeably, without regard to the disparities between them.

Ratio and Difference Comparisons

The different assumptions made by the two action selection rules raise an important empirical question: Are people more likely to make decisions on the basis of the ratio or the difference between the EVs representing each option? This question has been examined in detail in the psychophysics literature. Stevens (1957) argued for a psychophysical law by which the subjective magnitude of a stimulus is related to the physical magnitude by a power function. This law was a reaction to the Weber–Fechner law, which can be derived as an exponential function relating subjective magnitude to physical magnitude. Birnbaum and colleagues (Birnbaum, 1978; Birnbaum, Anderson, & Hynan, 1989) showed that participants can use both ratio and difference operations when instructed to do so. The comparative judgments involved in these experiments were primarily of physical stimuli.

There is evidence that numerical magnitudes are compared in much the same way as physical stimuli, such as loudness or length (Hinrichs, Yurko, & Hu, 1981; Moyer & Landauer, 1967). Moyer and Landauer asked participants to determine which of two single-digit numbers was larger. Reaction times (RTs) were inversely related to the numerical differences between the stimulus digits. Importantly, a ratio function described the data better than one based on absolute difference. Hinrichs et al. extended this work to two-digit number comparisons. They asked participants to determine whether the two-digit number presented on each trial was larger or smaller than some standard (e.g., larger or smaller than 55). They found evidence that two-digit numbers are compared holistically (i.e., the tens and ones digit places are not compared separately). In addition, they found that both ratio and difference functions fit the data well. However, the ratio function incorrectly predicted an asymmetry around the standard, in which greater RTs were expected for comparisons of larger numbers.

In this article, we present data from two behavioral experiments in which participants received rewards analogous to the rewards given in Decisions 1, 2, and 3 each time they selected from one of two decks of cards. In each experiment, participants in the control condition performed an experience-based choice task in which they received 1–10 points for each draw. Participants in the ratio-preserving condition performed the same task, except with reward values that had been multiplied by 10, so that the participants received 10–100 points for each draw (in increments of 10). Participants in the difference-preserving condition performed the same task as the control condition participants, except with reward values that had been shifted by 80 points, so that the participants received 81–90 points on each trial.

Exponential and power rules make opposite theoretical predictions for participants in the difference-preserving condition, who received reward values between 81 and 90 points. In this condition, the differences between the EVs are the same as those in the control condition, but

the ratios between the EVs are much smaller than those in the control condition. In this case, a power rule predicts that participants in this condition will perform worse than those in the control condition, because the probabilities for selecting either deck will be nearly equal. In contrast, an exponential rule predicts no difference.

The two rules also make opposite predictions for participants in the ratio-preserving condition, who received reward values between 10 and 100 points on each trial. In this condition, the ratios between the EVs are the same as those in the control condition, but the differences between EVs are much greater. The exponential rule predicts better performance for participants in the ratio-preserving condition relative to the control condition, because the exponential rule leads to larger differences in the selection probabilities. However, the power rule predicts no difference between these two conditions, because the ratios between the EVs for each option are equal.

EXPERIMENT 1

In Experiment 1, participants were asked to draw from one of two decks of cards on each of 80 trials. The decks were constructed so that optimal responding required exploiting the deck that currently gave the highest reward.¹ Figure 1A shows the points rewarded for selecting each deck on each trial. Deck B gave the highest reward over the first 50 trials, and Deck A gave the highest reward over the final 30 trials. Thus, participants needed to draw from Deck B for the first 50 trials and then draw from Deck A for the final 30 trials in order to perform optimally on the task.

To compare the predictions of the two choice rules, we manipulated the reward values on the decks between participants as described above. To maximize the number of points earned in the task, participants had to distinguish the high-value deck from the low-value deck in order to exploit the option with the highest expected reward on each trial. The power rule predicts worse performance for the difference-preserving condition relative to the control and ratio-preserving conditions, whereas the exponential rule predicts superior performance for the ratio-preserving condition relative to the control and difference-preserving conditions.

Method

Participants. Thirty University of Texas at Austin students participated in the experiment for course credit or monetary compensation (\$6 base pay). Participants were told that their goal was to earn as many points as possible and that they could earn a \$4 monetary bonus if they exceeded a prespecified performance criterion.

Materials. Participants performed the experiment on a personal computer, using MATLAB software. A deck appeared on the bottom half of each side of the screen. After each draw, the selected card value was displayed. On the right side of the screen, the phrase “Points required for the bonus” was shown, followed by the points required in each condition. Below this, the phrase “Your points” was shown, followed by the number of points each participant had earned.

Ten participants each were assigned to the control, difference-preserving, and ratio-preserving conditions. The participants in the control condition received reward values between 1 and 10 points on each draw, the participants in the difference-preserving condition received reward values between 81 and 90 points on each draw, and the participants in the ratio-preserving condition received reward

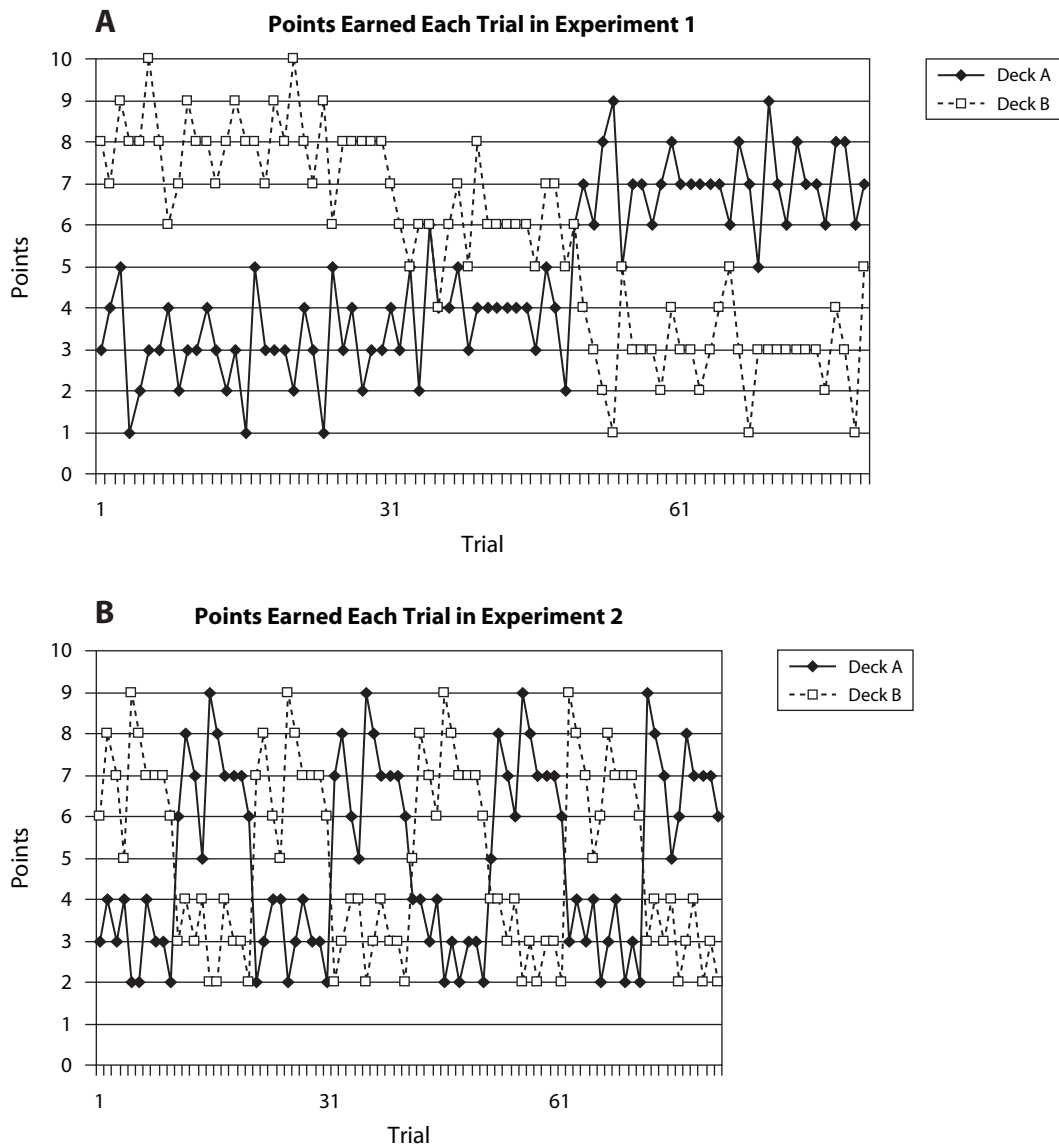


Figure 1. (A) Points given for selecting each deck on each trial in Experiment 1. (B) Points given for selecting each deck on each trial in Experiment 2.

values between 10 and 100 points on each draw. For the control condition, Deck A gave values that averaged 3 points during the first 30 trials, 4 points over the next 20 trials, and 7 points over the final 30 trials, whereas Deck B gave values that averaged 8 points over the first 30 trials, 6 points over the next 20 trials, and 3 points over the final 30 trials. The specific deck values were determined by choosing a random point value with the relevant mean and a standard deviation of .88, with all values being rounded to the nearest whole number. The reward values given on each trial were identical for the difference-preserving and ratio-preserving conditions, except that 80 points were added to each reward given for the difference-preserving condition, and each reward was multiplied by 10 points for the ratio-preserving condition. For the control condition, the maximum number of possible points that could be earned was 570, and the performance criterion was set at 550 points in order to ensure that a highly exploitative strategy was required in order to earn the monetary bonus. The performance criterion was adjusted accordingly for the difference-preserving condition (6,950 points) and the ratio-preserving condition (5,500 points).

Procedure. Participants were instructed to pick from one of two decks of cards. Information was given on the reward scale for each condition. Control condition participants were told that they would receive between 1 and 10 points, difference-preserving condition participants were told that they would receive between 81 and 90 points, and ratio-preserving condition participants were told that they would receive between 10 and 100 points. The participants were not given any information about the rewards given for selecting either deck, other than the range of possible rewards. There were 80 trials. After each draw, a card was turned over, the number of points received was shown, and the points were added to the total number of points, which was listed on the right side of the screen. The last card drawn from each deck remained visible until another card was drawn from that deck. After participants had completed the experiment, they were given the bonus if they had earned it.

Results

In order to compare the point totals across conditions, points earned by participants in the difference-

preserving condition were scaled by subtracting 80 points from each reward, and points earned by participants in the ratio-preserving condition were scaled by dividing each reward by 10. Figure 2A shows the average number of adjusted points earned in each condition. A one-way ANOVA revealed a main effect of condition [$F(2,27) = 4.27, p < .05, \eta^2 = .24$]. Pairwise comparisons revealed significantly better performance in the control condition ($M = 530.67$) than in the difference-preserving condition ($M = 497.34$) [$F(1,18) = 4.85, p < .05, \eta^2 = .21$]. Likewise, participants in the ratio-preserving condition ($M = 533.62$) earned significantly more points than did participants in the difference-preserving condition [$F(1,18) = 5.80, p < .05, \eta^2 = .24$]. However, there was no significant difference between the control and ratio-preserving conditions.

We next analyzed the proportion of optimal choices made by each participant on each trial. Recall that Deck B gave higher rewards than Deck A on the first 50 trials, and Deck A gave higher rewards than Deck B on the last 30 trials. Thus, an optimal choice was defined as selecting Deck B for the first 50 trials and Deck A for the final 30

trials. Figure 2B shows the average proportion of optimal choices made by participants in each condition. The pattern is the same as that of the total adjusted points earned data. A one-way ANOVA revealed a significant main effect of condition [$F(2,27) = 4.41, p < .05, \eta^2 = .24$]. Pairwise comparisons revealed that a significantly larger proportion of optimal choices was made in the control condition ($M = .87$) than in the difference-preserving condition ($M = .76$) [$F(1,18) = 4.56, p < .05, \eta^2 = .20$] and in the ratio-preserving condition ($M = .87$) than in the difference-preserving condition [$F(1,18) = 5.07, p < .05, \eta^2 = .22$]. Once again, there was no difference between the control and ratio-preserving conditions.

Discussion

These data support the predictions made by the power rule but not the exponential rule. When the reward values were shifted so that the difference between EVs remained constant, participants' performance was significantly worse than in the control condition. In contrast, when the reward values were multiplied, so that the ratio of the EVs remained constant, participants' performance was comparable to that in the control condition. These data suggest that human learners have difficulty exploiting the option with the highest EV in situations where the ratio between the EVs is low.

The differential performance between the control and difference-preserving conditions cannot be due merely to participants in the difference-preserving condition having to process larger reward values, because participants in the ratio-preserving condition also received numerically larger rewards than did participants in the control condition. Although those in the ratio-preserving condition received smaller rewards than did those in the difference-preserving condition, we did not find an inverse relationship between relative points earned and numerical magnitude of rewards.

We did not find strong support for the prediction of the exponential rule that participants in the ratio-preserving condition would be more likely than participants in the control condition to exploit the option with the highest EV on each trial and, thus, earn more points. However, participants in the ratio-preserving condition did earn slightly more points than did participants in the control condition, and they also made slightly more optimal choices.

To further test the predictions of each model, we designed a second experiment in which the optimal deck would switch more periodically. The motivation behind Experiment 2 was to replicate the findings of inferior performance for participants in the difference-preserving condition and to design a more difficult task that might better differentiate performance between the control and ratio-preserving conditions. In Experiment 1, the average proportion of optimal choices made by participants in the control and ratio-preserving conditions was very high (nearly 90%). Thus, it is possible that ceiling effects in these two conditions might have masked any true behavioral differences. By creating a less stationary environment where the optimal deck switched more often, overall performance should be further from optimal,

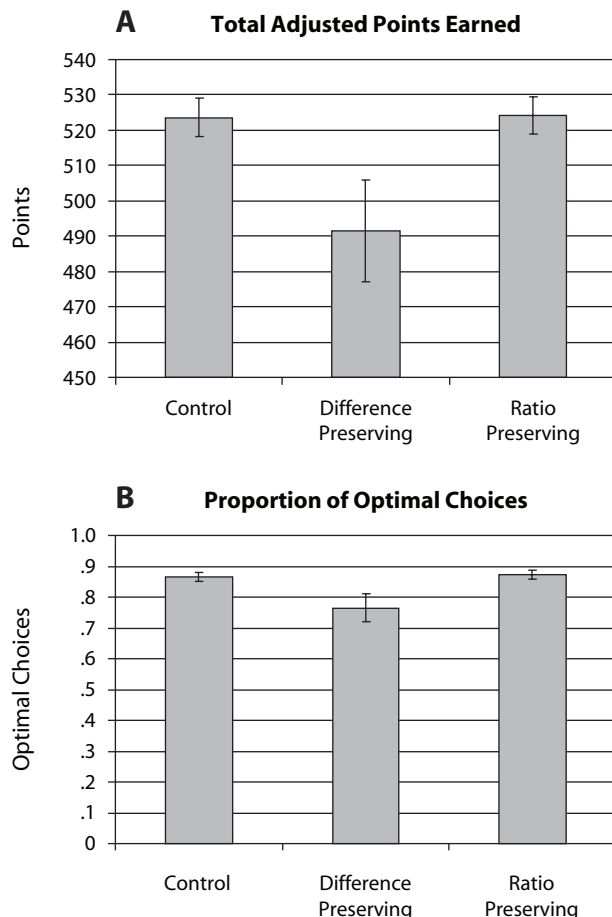


Figure 2. (A) Average total adjusted points earned by participants in each condition in Experiment 1. (B) Average proportion of optimal choices made by participants in each condition in Experiment 1.

leaving room for us to observe performance differences between the control and ratio-preserving conditions.

EXPERIMENT 2

In Experiment 2, we used a very different deck structure than we did in Experiment 1. At the start of the study in Experiment 2, one deck was optimal and the other was suboptimal. The decks switched status every 10 trials. When a given deck was optimal, it averaged seven points per card (in the control condition), whereas the suboptimal deck averaged three points per card. Figure 1B shows the points rewarded for selecting either deck on each trial. In this experiment, we examined whether the inferior performance of participants in the difference-preserving group is a robust phenomenon. We also tested the superior performance for participants in the ratio-preserving condition, as compared with participants in the control condition, that is predicted by the exponential rule. As discussed above, this more difficult task may allow us to see differences between the ratio-preserving condition and the control condition that may have been masked by a ceiling effect in Experiment 1.

Method

Participants. Thirty University of Texas at Austin students participated in the experiment for course credit or monetary compensation (\$6 base pay). Participants were told that their goal was to earn as many points as possible in order to earn a \$4 bonus.

Materials. Ten participants each were placed in the control, difference-preserving, and ratio-preserving conditions. For participants in the control condition, Deck A gave reward values that averaged seven points per draw ($SD = 0.88$), and Deck B gave reward values that averaged three points per draw ($SD = 0.88$) over the first 10 trials of the experiment. After 10 trials, the rewards given by the two decks reversed, so that during Trials 11–20, Deck B averaged seven points per draw and Deck A averaged three points per draw. The reward values reversed in this manner every 10 trials throughout the remainder of the experiment. We chose such a periodic design so that each deck would produce the same number of optimal draws.

As in Experiment 1, participants in the difference-preserving condition received the same reward values as participants in the control condition on each trial, except that 80 points were added to each reward value so that they received rewards ranging from 81 to 90 points on each trial. Similarly, participants in the ratio-preserving condition received the same rewards as those in the control condition, except that each reward value was multiplied by 10 points so that they received rewards ranging from 10 to 100 points in increments of 10. The maximum number of points a participant in the control condition could earn on the task was 560. The bonus was set at 525 points, which meant that participants had to be very vigilant in selecting the deck with the highest payoff. The bonus was adjusted accordingly for participants in the difference-preserving condition (6,925 points) and for participants in the ratio-preserving condition (5,250 points).

All other materials were the same as those used in Experiment 1.

Procedure. The procedure was identical to the procedure for Experiment 1.

Results

As in Experiment 1, we analyzed the total number of adjusted points earned by each participant in the task using the same procedure outlined in Experiment 1. Figure 3A shows the average number of adjusted points earned in each

condition. An ANOVA revealed significant differences between conditions [$F(2,27) = 8.18, p < .01, \eta^2 = .38$]. Pairwise comparisons revealed that, once again, participants in the difference-preserving condition ($M = 450.9$) earned significantly fewer adjusted points than did participants in the control condition ($M = 485.1$) [$F(1,18) = 5.15, p < .05, \eta^2 = .22$] or in the ratio-preserving condition ($M = 504.2$) [$F(1,18) = 12.59, p < .01, \eta^2 = .41$]. Interestingly, we also found that participants in the ratio-preserving condition earned significantly more adjusted total points than did participants in the control condition [$F(1,18) = 4.46, p < .05, \eta^2 = .20$].

We also analyzed the proportion of optimal choices in each condition. In this experiment, an optimal choice entailed selecting from the deck that gave the highest payoff. Figure 3B shows the average proportion of optimal choices in each condition. An ANOVA revealed a significant effect of condition [$F(2,27) = 8.37, p < .01, \eta^2 = .38$]. Pairwise comparisons showed that participants in the difference-preserving condition ($M = .643$) made significantly fewer optimal choices than did participants in the control condition ($M = .764$) [$F(1,18) = 5.83, p < .05, \eta^2 = .25$] or in the ratio-preserving condition ($M = .818$) [$F(1,18) = 12.23, p < .01, \eta^2 = .40$]. The difference in the numbers of optimal choices between the ratio-preserving condition and the control condition was marginally significant [$F(1,18) = 3.93, p < .10, \eta^2 = .18$].

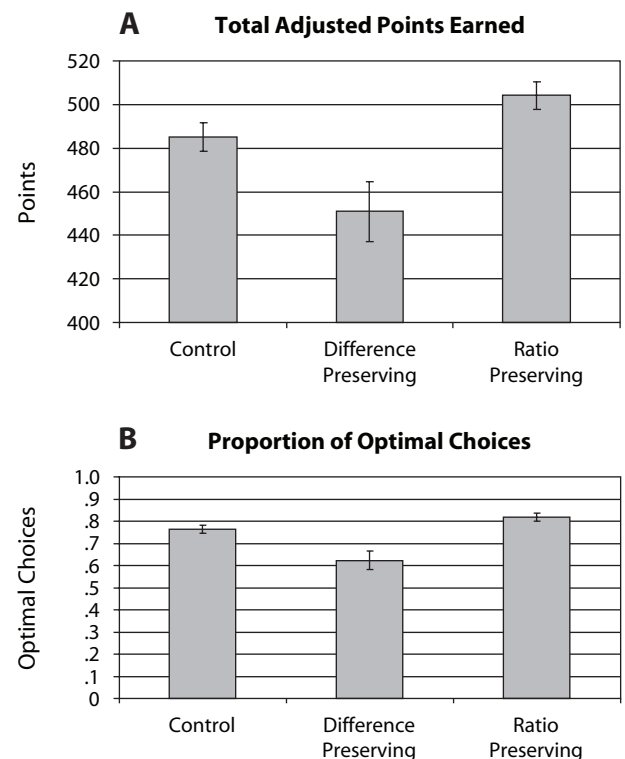


Figure 3. (A) Average total adjusted points earned by participants in each condition in Experiment 2. (B) Average proportion of optimal choices made by participants in each condition in Experiment 2.

Discussion

Consistent with the results of Experiment 1, participants in the difference-preserving condition were less able to exploit the option with the highest EV than were those in the control and ratio-preserving conditions. These participants earned fewer points and made fewer optimal choices than did participants in the control condition. These results confirm the prediction made by the power rule that as the ratio between the best and worst options decreases, the probability of exploiting the more rewarding option decreases. In contrast, this result contradicts the prediction of the exponential rule that the difference, and not the ratio, between the EV of each option governs responding.

In Experiment 2, we found that participants in the ratio-preserving condition earned more points than did participants in the control condition and made marginally more optimal choices throughout the task. These results support the prediction of the exponential rule that as the difference between the best and worst options increases, the probability of exploiting the more rewarding option increases. This pattern suggests that there may have been a ceiling effect in Experiment 1 that masked the difference between the ratio-preserving condition and the control condition.

These combined results suggest that shifting the scale of the reward values (as in the difference-preserving condition) may lead to rewards that are not as perceptually discriminable, whereas magnifying the scale of the reward values (as in the ratio-preserving condition) may lead to rewards that are more perceptually discriminable. Although we have so far found stronger evidence supporting the predictions of the power rule than supporting the predictions of the exponential rule, note that the exponential rule correctly predicted the advantage found for participants in the ratio-preserving condition in Experiment 2. Thus, both rules make some correct predictions about choice behavior, and human learners may be able to compare EVs of alternate options via either ratio or difference operations.

GENERAL DISCUSSION

Our goal in conducting these experiments was to test the empirical validity of behavioral predictions derived from the ratio and exponential action selection choice rules that are commonly used in a wide range of cognitive domains (e.g., Busemeyer & Stout, 2002; Daw et al., 2006; Kruschke, 1992; Love et al., 2004; Maddox & Ashby, 1993; Medin & Schaffer, 1978; Minsky & Papert, 1968; Nosofsky, 1986; Reed, 1972; Rodrigues & Murre, 2007; Rumelhart & McClelland, 1986; Sutton & Barto, 1998; Worthy et al., 2007; Yechiam et al., 2005). Even though these two types of rules are often used, there is typically no justification for selecting one rule over the other (for exceptions, see Corrado et al., 2005; Daw & Doya, 2006; Rodrigues & Murre, 2007). However, the differential output of the choice rules when EVs of varying ratio or differences are inserted is of importance to modelers and empirically minded scientists alike.

We found evidence that participants can use both ratio and difference operations to compare expected rewards in

experience-based decision-making tasks. In Experiments 1 and 2, we found that lowering the ratio between the rewards given for each deck decreased exploitation of the best option. In addition, Experiment 2 demonstrated that increasing the difference between the rewards increased exploitation. In the remainder of this section, we discuss the implications of our findings for several other fields.

Implications for Prospect Theory

In this article, we found that participants were less likely to exploit the best option as the ratio between the options decreased but the difference was held constant. This pattern of data is congruent with the negatively accelerated value function in prospect theory (Kahneman & Tversky, 1979). For example, the difference in value between a gain of \$5 and a gain of \$10 appears to be greater than that between a gain of \$85 and a gain of \$90. This fits in with predictions derived from the power rule, where the ratio between EVs determines choice. Our results, combined with evidence for prospect theory's value function, indicate that ratio operations may be the primary operation used by humans to make comparisons of subjective value.

However, the value function is based on one's current reference point, not on absolute magnitude. Participants who received rewards between 81 and 90 points failed to discriminate the better option as well as participants who received rewards between 1 and 10 points did. However, they should not have had difficulty if they were able to establish a reference point between 81 and 90 units before the task began. These results suggest that people may have a natural reference point around 0, and that experience is needed to develop a new reference point (i.e., instructions may not be enough).

The distinction must also be made between the present experience-based choice paradigm and the description-based choice paradigms often used to support prospect theory. In description-based choice tasks, participants are asked to make a choice between two options with known probabilities (e.g., choose between Option A, to gain \$4 with .8 probability or \$0 otherwise, and Option B, to gain \$3 with certainty). In contrast, in experience-based choice tasks, the decision maker must determine the value of each option by repeated sampling.

The results of several recent studies have shown differences in decision making depending on whether decisions were made on the basis of description or experience (e.g., Barron & Erev, 2003; Hertwig, Barron, Weber, & Erev, 2004; Weber, Shafir, & Blais, 2004). One notable difference is that in description-based choice tasks, participants often overweight rare events, whereas in experience-based tasks, participants underweight rare events (Barron & Erev, 2003; Hertwig et al., 2004). Future work should examine decision-making biases across both contexts in order to gain a more complete understanding of how humans make subjective value comparisons.

Ratios, Differences, and Numerical Comparison

Our results suggest that human participants are able to compare expected rewards by either ratio or difference operations but that altering the ratio between rewards leads to

the most drastic changes in behavior. These results may inform psychophysical theories of comparisons of subjective magnitudes and may help to extend these theories to comparisons of expected numerical rewards. Previous researchers examining numerical comparison found evidence that numbers are compared by both ratio (Moyer & Landauer, 1967) and difference (Hinrichs et al., 1981) operations.

Although we interpret our finding of inferior performance for the difference-preserving condition relative to the control condition as evidence that ratio operations are used to determine which choice to make, it is possible that the numbers used in the difference-preserving condition (81–90) are in some way more awkward, complex, or unfamiliar than those in the control (1–10) and ratio-preserving (10–100) conditions. There is evidence that numbers are processed holistically, rather than decomposed into units (Dehaene, Dupoux, & Mehler, 1990; Hinrichs et al., 1981), and there is no previous evidence that larger numbers are more difficult to compare. Additionally, even if numbers are decomposed into units simultaneously during processing, comparisons in all three conditions relied on only a single digit. The control and difference-preserving conditions relied on the ones digit, and the ratio-preserving condition relied on the tens digit for comparison. It is possible, however, that the 8 in the tens digit position distracted participants in the difference-preserving condition to such a degree that it led to their inferior performance. Future work should examine whether the differences found here are a product of the comparison operation used, as we suggest, or simply due to a deficiency in numerical processing.

It is important to examine the operations used to make value comparisons in other domains as well. We have pointed out relations of this work to other decision-making tasks and to comparative judgment tasks. However, there are noticeable differences between the present task, in which differences between the two options must be learned from experience, and numerical comparison tasks, in which two options are presented simultaneously and a comparative judgment is made. One area in which ratio and difference judgments could be tested is feedback-based category learning. This kind of category learning task is similar to the present experience-based choice paradigm in that information is gained through sampling, rather than by description. For example, consider two categories in which the members are lines that vary in length. It would be interesting to examine whether changes in the ratios or differences between the line lengths characteristic of members of each category affect category learning performance. Resolution of these questions will allow for a more complete understanding of the comparison operations that people use to make decisions based on experience.

Conclusion

Our results are informative for models of choice, because they suggest that using only a power rule or only an exponential rule to model human behavior may not adequately encompass the processes used to make a decision. Although the differences between the outputs of these two types of rule are often negligible, they are nevertheless

worthy of consideration when modeling behavior. Here, we have shown how two popular action selection rules make contrasting theoretical predictions about choice behavior and that the predictions of each rule diverge from people's performance under some conditions. Humans and learning models process rewards or input values of varying magnitudes quite differently, and these differences should not be simply ignored.

AUTHOR NOTE

This research was supported by AFOSR Grant FA9550-06-1-0204 and NIMH Grant MH077708 to W.T.M. and A.B.M., and by a supplement to NIMH Grant MH077708 to D.A.W. We thank Tyler Davis, Marc Tomlinson, Lisa Grimm, and Jeff Laux for valuable comments. We also thank Scott Lauritzen and Bo Zhu for help in collecting the data. Correspondence concerning this article should be addressed to W. T. Maddox or D. A. Worthy, Department of Psychology, University of Texas, 1 University Station A8000, Austin, TX, 78712 (e-mail: maddox@psy.utexas.edu or worthyda@mail.utexas.edu).

REFERENCES

- ASHBY, F. G., & MADDOX, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, *37*, 372-400.
- BARRON, G., & EREV, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making*, *16*, 215-233.
- BIRNBAUM, M. H. (1978). Differences and ratios in psychological measurement. In N. J. Castellan & F. Restle (Eds.), *Cognitive theory* (Vol. 3, pp. 33-74). Hillsdale, NJ: Erlbaum.
- BIRNBAUM, M. H., ANDERSON, C. J., & HYNAN, L. G. (1989). Two operations for "ratios" and "differences" of distances on the mental map. *Journal of Experimental Psychology: Human Perceptions & Performance*, *15*, 785-796.
- BRIDLE, J. S. (1990). Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimates of parameters. In D. S. Touretzky (Ed.), *Advances in neural information processing systems 2* (pp. 211-217). San Mateo, CA: Morgan Kaufmann.
- BUSEMEYER, J. R., & STOUT, J. C. (2002). A contribution of cognitive decision models to clinical assessment: Decomposing performance on the Bechara gambling task. *Psychological Assessment*, *14*, 253-262.
- CORRADO, G. S., SUGRUE, L. P., SEUNG, S. H., & NEWSOME, W. T. (2005). Linear-nonlinear-Poisson models of primate choice dynamics. *Journal of the Experimental Analysis of Behavior*, *84*, 581-617.
- DAW, N. D., & DOYA, K. (2006). The computational neurobiology of learning and reward. *Current Opinion in Neurobiology*, *16*, 199-204.
- DAW, N. D., O'DOHERTY, J. P., DAYAN, P., SEYMOUR, B., & DOLAN, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*, 876-879.
- DEHAENE, S., DUPOUX, E., & MEHLER, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception & Performance*, *16*, 626-641.
- ERICKSON, M. A., & KRUSCHKE, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, *127*, 107-140.
- ERICKSON, M. A., & KRUSCHKE, J. K. (2002). Rule-based extrapolation in perceptual categorization. *Psychonomic Bulletin & Review*, *9*, 160-168.
- HERRNSTEIN, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, *4*, 267-272.
- HERTWIG, R., BARRON, G., WEBER, E. U., & EREV, I. (2004). Decisions from experience and the effect of rare events in risky choices. *Psychological Science*, *15*, 534-539.
- HINRICHS, J. V., YURKO, D. S., & HU, J.-M. (1981). Two-digit number comparison: Use of place information. *Journal of Experimental Psychology: Human Perception & Performance*, *7*, 890-901.

- KAHNEMAN, D., & TVERSKY, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, **47**, 263-291.
- KRUSCHKE, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, **99**, 22-44.
- LAU, B., & GLIMCHER, P. W. (2005). Dynamic response-by-response models of matching behavior in rhesus monkeys. *Journal of the Experimental Analysis of Behavior*, **84**, 555-579.
- LOVE, B. C., MEDIN, D. L., & GURECKIS, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, **111**, 309-332.
- LUCE, R. D. (1959). *Individual choice behavior*. New York: Wiley.
- LUCE, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 103-189). New York: Wiley.
- MADDOX, W. T., & ASHBY, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, **53**, 49-70.
- MEDIN, D. L., & SCHAFFER, M. M. (1978). A context theory of classification learning. *Psychological Review*, **85**, 207-238.
- MINSKY, M., & PAPERT, S. (1968). *Perceptrons*. Cambridge, MA: MIT Press.
- MOYER, R. S., & LANDAUER, T. K. (1967). Time required for judgments of numerical inequality. *Nature*, **215**, 1519-1520.
- NOSOFSKY, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.
- NOSOFSKY, R. M., & PALMERI, T. J. (1998). A rule-plus-exception model for classifying objects in continuous-dimension spaces. *Psychonomic Bulletin & Review*, **5**, 345-369.
- NOSOFSKY, R. M., & ZAKI, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 924-940.
- REED, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, **3**, 482-487.
- ROBERTS, M. E., & GOLDSTONE, R. L. (2006). EPICURE: Spatial and knowledge limitations in group foraging. *Adaptive Behavior*, **14**, 291-313.
- RODRIGUES, P. M., & MURRE, J. M. (2007). Rules-plus-exception tasks: A problem for exemplar models? *Psychonomic Bulletin & Review*, **14**, 640-646.
- RUMELHART, D. E., MCCLELLAND, J. L., & THE PDP RESEARCH GROUP (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1. Foundations*. Cambridge, MA: MIT Press.
- SAKAMOTO, Y., MATSUKA, T., & LOVE, B. C. (2004). Dimension-wide vs. exemplar-specific attention in category learning and recognition. In M. Lovett, C. Schunn, C. Lebiere, & P. Munro (Eds.), *Proceedings of the 6th International Conference on Cognitive Modeling* (pp. 261-266). Mahwah, NJ: Erlbaum.
- SHEPARD, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, **22**, 325-345.
- STEVENS, S. S. (1957). On the psychophysical law. *Psychological Review*, **64**, 153-181.
- SUGRUE, L. P., CORRADO, G. S., & NEWSOME, W. T. (2004). Matching behavior and the representation of value in the parietal cortex. *Science*, **304**, 1782-1787.
- SUTTON, R. S., & BARTO, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.
- WEBER, E. U., SHAFIR, S., & BLAIS, A.-R. (2004). Predicting risk sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psychological Review*, **111**, 430-445.
- WILLIAMS, R. J. (1988). On the use of backpropagation in associative reinforcement learning. In *Proceedings of the IEEE International Conference on Neural Networks* (pp. 263-270). San Diego: IEEE.
- WORTHY, D. A., MADDOX, W. T., & MARKMAN, A. B. (2007). Regulatory fit effects in a choice task. *Psychonomic Bulletin & Review*, **14**, 1125-1132.
- YECHIAM, E., BUSEMEYER, J. R., STOUT, J. C., & BECHARA, A. (2005). Using cognitive models to map relations between neuropsychological disorders and human decision-making deficits. *Psychological Science*, **16**, 973-978.

NOTE

1. We define *exploitation* as selecting the option with the highest EV and *exploration* as selecting the option with a lower EV.

APPENDIX

Power Rule

It can be shown that for two-choice alternatives, the power rule reduces to the ratio between the two alternatives.

Let x represent the EV of choice a , y represent the EV of choice b , and r represent the ratio between x and y .

Let

$$r = \frac{x}{y}$$

and

$$y = \frac{x}{r}$$

The power rule can then be stated as

$$\begin{aligned} P_{a,t} &= \frac{x}{x + \frac{x}{r}} \\ &= \frac{x}{\left(1 + \frac{1}{r}\right)x} \\ &= \frac{r}{(1+r)}, \end{aligned}$$

so that x cancels out, and only the ratio of x and y remains.

APPENDIX (Continued)

Exponential Rule

Similarly, it can be shown that the exponential rule reduces to the difference between the two alternatives.

Let x represent the EV of choice a , y represent the EV of choice b , and d represent the difference between x and y .

Let

$$d = y - x$$

and

$$y = d + x.$$

The exponential rule can then be stated as

$$\begin{aligned} P_{a,t} &= \frac{e^x}{e^x + e^{(d+x)}} \\ &= \frac{e^x}{e^x + (e^x * e^d)} \\ &= \frac{e^x}{e^x(1 + e^d)} \\ &= \left(\frac{1}{1 + e^d} \right), \end{aligned}$$

so that x cancels out, and only the difference between x and y remains.

(Manuscript received May 9, 2008;
revision accepted for publication July 31, 2008.)