The Psychological Record
https://doi.org/10.1007/s40732-020-00451-w

ORIGINAL ARTICLE

A Proof of Concept Analysis of Decision-Making with Time-Series Data

David J. Cox[1] · Matthew T. Brodhead[2]

Accepted: 14 December 2020

Abstract
Behavior analysts make dozens of practice-related decisions every day. Past research has extensively examined practice-related decision making by medical and other healthcare professionals, but decision making by practicing behavior analysts has garnered little research attention. The purpose of this proof of concept study was to begin a translational research agenda toward understanding what variables influence behavior analyst decision making and precisely how those variables do so. Behavior analyst decision making often involves data and the most fundamental characteristics of data are those indicating data trustworthiness—validity, reliability, and accuracy. To isolate a single independent variable, we used a lengthening data path procedure to parametrically assess how reducing data accuracy changed decisions to continue or modify an intervention in 30 students of behavior-analytic masters or doctoral programs. When data accuracy was 100%, most participants waited 9–10 trials before intervening. When data accuracy was below 60%, most participants waited 4–6 trials before intervening. To begin exploring potential behavioral processes underlying the influence of data accuracy on practice-related decisions, we also examined how well one popular description of decision making described participant choice—probability discounting. Probability discounting described the pattern of choices well for 16 of the 30 participants, suggesting other analytic frameworks should be explored. Nevertheless, data accuracy systematically influenced the majority of participants choices to continue or modify an intervention, although the degree of influence differed on an individual basis.

Keywords visual analysis · choice · probability discounting · decision making

Introduction

To extend Herrnstein (1970), all practice-related decisions in applied behavior analysis (ABA) are choice. For every decision a behavior analyst makes, there are likely one or more responses the behavior analyst could make. For example, when faced with an escape-maintained problem behavior, a behavior analyst has at least seven different intervention options to choose from (Geiger, Carr, & LeBlanc, 2010). As another example, when faced with the addition of a new data point on a time-series plot, a behavior analyst has to decide to continue, modify, or end a functional analysis (Saini, Fisher, & Retzlaff, 2018) or a baseline condition (Vanselow, Thompson, & Karsina, 2011). Researchers have extensively examined practice-related decisions in medical and other clinical professions for decades (e.g., journals *Medical Decision Making* and *BMC Medical Informatics and Decision Making*). Practice-related decisions in ABA have not been as thoroughly researched.

Visual analysis of time-series data is particularly important to practice-related decisions in ABA because it is the primary method of data analysis taught by instructors of behavior-analytic courses (Wolfe & McCammon, in press) and for evaluating applied behavioral interventions (Dart & Radley, 2017). In addition, few ABA practice-related decisions are likely made separate from visual analysis. Therefore, visual analysis is a socially important area for studying practice-related decisions made by behavior analysts.

David J. Cox and Matthew T. Brodhead share the first author.

David J. Cox
david@bhcoe.org

Matthew T. Brodhead
mtb@msu.edu

1 Data & Analytics, Research & Development, Behavioral Health Center of Excellence, Los Angeles, CA 90046, USA

2 Department of Counseling, Educational Psychology, and Special Education, College of Education, Michigan State University, East Lansing, MI 48824, USA

Published online: 14 January 2021

Research on visual analysis has primarily focused on improving these skills in novice or untrained individuals. Visual analysis has been shown to improve through behavioral skills training (e.g., Maffei-Almodovar, Feliciano, Fienup, & Sturmey, 2017), equivalence-based instruction (e.g., Blair et al., 2019), visual aids (e.g., Fisher, Kelley, & Lomas, 2003), computer-based instruction (Wolfe & Slocum, 2015), decision-making algorithms (Kipfmiller et al., 2019), and systematic protocols (Wolfe, Barton, & Meadan, 2019). In general, research on visual analysis indicates that novice individuals differ from experts and that people can be taught to make decisions about treatment effects that agree with recognized experts in the visual analysis of behavioral data (e.g., Kahng et al., 2010; Ninci, Vannest, Willson, & Zhang, 2015; cf. Diller, Barry, & Gelino, 2016).

One disconnect between research on training visual analysis and practice-related decisions in everyday contexts are how data sets are displayed to the behavior analyst. Past research on practice-related decisions via visual analysis have primarily relied on static, complete data sets (e.g., computer-generated datasets that resemble common behavioral patterns). But many practice-related decisions involve ongoing visual analysis of dynamically changing data sets. For example, consider a time-series graph that has 15 data points, 5 of which represent a baseline condition, and 10 of which represent a treatment condition. Rather than making one decision using a graphical display involving all 15 data points, the behavior analyst may make up to 15 decisions—1 decision after each session where a new datum has been added to the data path.

Recent studies by Saini et al. (2018) and Vanselow et al. (2011) illustrate the importance of using ecologically valid methods of stimulus presentation to evaluate practice-related decision making. Saini et al. (2018) used structured decision-making criteria to identify the function of problem behavior for 100 standardized graphs displaying functional analysis (FA) data across three conditions. One condition determined behavioral function using the structured criteria and the complete set of FA data. A second condition determined behavioral function using the authors post-hoc visual inspection of the data (i.e., without the structured criteria) and the complete set of FA data. The third condition determined behavioral function using the structured criteria and a moving window of data akin to ongoing visual analysis of an updating data set. The researchers found that ongoing visual inspection identified the function of behavior in an average of 13 fewer sessions, but exact agreement of behavioral function across the three approaches was 79%.

In another example, Vanselow et al. (2011) examined how novice and expert behavior analysts made decisions to terminate baseline conditions. Participants were shown sets of baseline data that increased datum-by-datum and asked when they would terminate the baseline period. These researchers found that the length of baseline was similar across experts but more variable across novices. Furthermore, the researchers reported more variation across observers as variability in the data sets increased, and they reported less variation when information on the future intervention (independent variable) or target behavior (dependent variable) were provided to the participants.

The results of Saini et al. (2018) and Vanselow et al. (2011) suggest practice-related decisions using ongoing visual analysis are unlikely to always be equivalent to decisions examined in past research using complete data sets. In addition, the results from Vanselow et al. (2011) indicate that information additional to the visual display of the data can influence behavior analyst decision making. Given all practice-related decisions are instances of choice, an open question then becomes how variables known to influence choice from basic research affect data-based decision making.

Data Trustworthiness

There are likely to be hundreds of potential contextual variables that affect the ongoing visual analysis of data (e.g., one's own training in visual analysis, the length of time one has to decide). Past research has focused primarily on trend, level, and variability of the data path, as well as information about the upcoming intervention and target behavior. But data paths involve aggregations of data. A potentially more fundamental place to start might be the characteristics associated with the trustworthiness of data (Johnston & Pennypacker, 1993). Unless data are known to be valid, reliable, or accurate, it is unlikely that other variables such as trend, level, and variability (see Ledford & Gast, 2018) can be considered for analysis.

Collected behavioral data are considered valid when they are directly relevant to the target behavior of interest (Cooper, Heron & Heward, 2020). Decisions about data validity revolve around at least two areas. First, decisions are made to ensure the measured dimension of behavior (e.g., duration, latency, rate) is relevant to the concern or question about the behavior. Second, decisions are made to ensure that collected data are representative of responding within the conditions and times of day that are relevant to the concern or question about the behavior. It is currently unknown what variables influence decisions that improve data validity. However, these decisions likely take place before data begins to be collected and plotted visually. Thus, though data validity is a fundamental characteristic of trustworthy data, it is outside the scope of the present study.

Collected behavioral data is considered reliable when repeated instances of the same behavior lead to the same reported value (Johnston & Pennypacker, 1993). That is, when an observer records data on a measured dimension of behavior (e.g., duration, latency, rate), a similar response emitted at different points in time leads to a similar recorded quantitative value. Decisions about data reliability likely occur throughout service delivery when a more experienced behavior analyst observes service delivery of the direct provider (e.g., registered behavior technician) and the data they collect. It is currently unknown what variables influence decisions made by behavior analysts that affect the reliability of data collected by a direct provider.

Collected behavioral data are considered accurate when a recorded instance of a behavior corresponds to the actual performance of the observed individual (Kazdin, 2011). Accurate data assumes there is a ground truth to physical events that took place and that the observer's recording behavior was sensitive to any behavioral variability relevant to the behavior of interest. Decisions about data accuracy also are likely to occur throughout service delivery when a more experienced behavior analysts observes service delivery of the direct provider and the data they collect. Similar to validity and reliability, it is currently unknown what variables influence decisions made by a behavior analysts that affect the accuracy of data collected by a direct provider.

Though validity, reliability, and accuracy are fundamental characteristics of behavioral data, accurate data are the primary quality indicator of behavioral measurement (Cooper, Heron & Heward, 2007) and provide a real and honest indicator of program progress for evaluation (Hinz, McGee, Huitema, Dickinson, & Van Enk, 2014). Inaccurate data may result in false conclusions during data analysis (Lerman, Hovanetz, Strobel, & Tetreault, 2009) and raise consumer concerns about the legitimacy of behavior-analytic interventions. High data accuracy may also provide assurances to consumers of behavior analysis that what is being reported has actually occurred (Cooper at el., 2007). Past research on data accuracy has typically focused on environmental variables (e.g., rule statements, modality of data collection, observer presence) that influence the collection of accurate data (e.g., Lerman et al., 2009; O'Leary, Kent, & Kanowitz, 1975; Repp, Nieminen, Olinger, & Brusca, 1988). Less is known about how data accuracy influences practice-related decision making of the supervising behavior analyst. Thus, we focused on data accuracy as an initial exploration of how changing dimensions of data trustworthiness influence decision making.

Behavioral Processes that Control Decision Making

In this proof of concept study, we hypothesize that data accuracy will influence practice-related decision making. If it does, a follow-up question is: What behavioral processes account for the pattern of influence that data accuracy has on decision making? In laboratory research examining behavioral processes of choice and decision making, researchers typically isolate a single variable relative to a single behavioral process. For example, researchers might manipulate: the relative amount of reinforcer delivery across two alternatives to measure allocation of responding (i.e., matching law; Herrnstein, 1961), the delay between behavioral emission and contact with a reinforcer to measure allocation of responding (i.e., delay discounting; Rachlin & Green, 1972), the number of responses required to contact a reinforcer to measure the rate of responding and reinforcers earned (i.e., demand; Hursh & Silberberg, 2008), or the covariance between a stimulus signaling a reinforcer and the occurrence of a reinforcer to measure stimulus control over responding (i.e., signal detection; Baum, 2018; Rescorla, 1967, 1968).

Given the complex dynamics of everyday applied settings, it is highly probable that multiple behavioral processes interact to control practice-related decision making of behavior analysts. One approach to understanding the multiple control of practice-related decision making is to examine how well each behavioral process can account for decision making. Then, once known, researchers can systematically put the pieces back together to determine how different behavioral processes interact to multiply control decision making in nonlaboratory settings. Thus, as an initial exploration into the behavioral processes controlling practice-related decision making relative to data accuracy, we chose arguably the currently most commonly studied behavioral process in behavior analysis—discounting—to determine how well it might account for practice-related decision making relative to data accuracy.

Probability discounting Data accuracy can fall anywhere on a continuum from completely inaccurate to completely accurate. All instances of data collection may not reflect actual performance (i.e., 0% accuracy), all instances of data collection may perfectly reflect actual performance (i.e., 100% accuracy), or data may fall somewhere in between (e.g., 50% accuracy). In practice, what is considered actual client performance may be based on some form of criterion record (Mudford, Martin, Hui, & Taylor, 2009; i.e., the "true value"—Johnston & Pennypacker, 1993) such as an observed permanent product or data obtained through reliable and valid mechanical measurement system (e.g., Boykin & Nelson, 1981). When a true value is not available, data accuracy may be inferred by how well researchers have reduced potential threats to internal validity associated with measurement error (e.g., observer drift) or by measures of observer reliability (though high reliability does not imply high data accuracy; see Repp et al., 1988).

Probability discounting seems particularly relevant when asking how data accuracy influences behavior analyst decision making. Probability discounting refers to an observed reduction in the likelihood of emitting a response (Option A) as the likelihood (i.e., probability) decreases that the organism will contact some consequence following that response (Rachlin, Raineri, & Cross, 1991). Related to this, as the likelihood decreases that the organism contacts the consequence following Option A, the likelihood of emitting a second response (Option B) increases.

Consider an example where a participant is given a choice between:

Option A: 95% chance of getting $10, and
Option B: 100% chance of getting $5.

In this example, most participants will choose Option A on nearly 100% of opportunities to make a choice (e.g., Green, Myerson, & Ostaszewski, 1999; Rachlin et al., 1991). As the amount of Option B increases, eventually there will be an amount where the participant switches and chooses Option B (e.g., 95% chance of getting $10 vs. 100% chance of getting $8). Researchers typically calculate the amount that is halfway between the last amount that Option A was chosen and the first amount that Option B was chosen (e.g., Green et al., 1999; Rachlin et al., 1991). For example, if the last amount that Options A and B were chosen were $7 and $8, respectively, the halfway point would be $7.50. This halfway point is often referred to as an indifference point because it is the point where the participant is equally likely to choose between the uncertain Option A and the certain Option B—they are indifferent (i.e., do not have a strong preference for either option).

The influence of probability on choice is often described mathematically as hyperbolic-like (Fig 1). Most probability discounting tasks identify a series of indifference points (e.g., 5 or 7) across a range of probabilities of interest (e.g., 100%–5%). After the set of indifference points are obtained, a hyperboloid equation is often fit to the indifference points to quantitatively describe how changing the probability of contacting a consequence reduces the likelihood that an individual participant would emit a response to obtain it (Green et al., 1999; Rachlin et al., 1991). This hyperboloid equation can be written as:

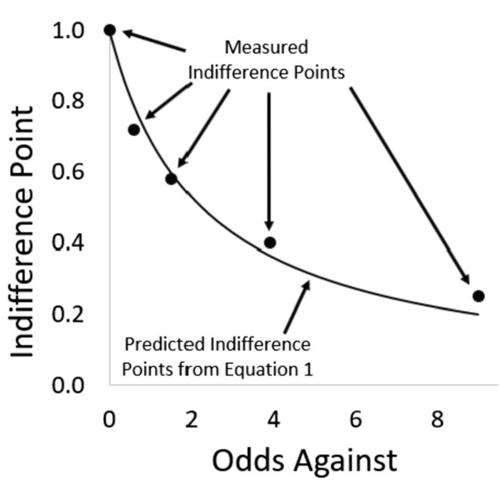

Fig 1. Example probability discounting graph. Markers represent the measured indifference points obtained from a choice task for an individual participant. The line represents the predicted indifference points from Eq. 1 once h is estimated using the individual participant's indifference points

$$I = \frac{A}{(1 + h\theta)^s}. \quad (1)$$

In this equation, I is the indifference point, A is the undiscounted amount of the larger consequence (e.g., $10, five M&Ms, 60 s access to iPad, or five tokens),   is the odds against any single response contacting the reinforcer (calculated as (1-p)/p where p is the probability of occurrence), h is a number that quantitatively describes how quickly the likelihood of emitting a response for that reinforcer will decrease based on the reducing probability of contacting the reinforcer, and s reflects nonlinear psychophysical scaling of amount and the odds against (see McKerchar, Green, & Myerson, 2010, for discussion on how to interpret the s parameter). Of note is that I equals A when the probability of the larger outcome is 100% (= 0). Thus, for outcomes with an undiscounted amount that is less clear within a decision context (e.g., improved health, client outcomes), A becomes the maximum number of responses that occur when the probability of the larger outcome is 100% (= 0).

Probability Discounting as a Framework for Evaluating Applied Behavior-Analytic Decision Making

Probability discounting is a methodological and descriptive-analytic framework that allows researchers to parametrically evaluate how choice changes as the probability reduces that a behavior will contact a consequence. Translational research conducted using a probability discounting framework has described how a number of variables influence socially significant behaviors (Critchfield, 2011; McKerchar & Renda, 2012). For example, the probability discounting framework helped elucidate how drug administration increased choice to engage in sexual behaviors involving different probabilities of contracting a sexually transmitted infection (Johnson, Herrmann, Sweeney, LeComte, & Johnson, 2017; Johnson, Sweeney, Herrmann, & Johnson, 2016). Probability discounting has helped describe how the probability of experiencing the benefits and side-effects of medication can affect medication adherence in individuals with chronic diseases (Bruce et al., 2016; Jarmolowicz et al., 2016, 2019). As a final example, probability discounting has been used to analyze the relationship between percent body fat and choices of food consumption (Rasmussen, Lawyer, & Reilly, 2010). Overall, a probability discounting framework provides a well-researched set of procedures to parametrically analyze how changing probability can influence socially significant behaviors.

Probability discounting may also provide a framework for understanding how behavior analysts make practice-related decisions. Visual analysis serves as one important area in which variables may be systematically controlled in order to identify how context (e.g., data accuracy) affects decision making in behavior-analytic practice. In visual analysis, data accuracy might be conceptualized as the probability that collected data reflects a client's true performance. If we assume positive client progress reinforces intervention decisions, then the probability that a client is actually making (or maintaining) progress may be described through data accuracy. If data accuracy influences intervention decisions consistent with a probability discounting framework, then we would expect choice between two intervention decisions to be systematically and hyperbolically influenced by data accuracy that ranges from 0% to 100%.

The purposes of this proof of concept experiment were therefore threefold. First, we were interested in studying decision making through visual analysis of behavioral data using a more ecologically valid method that reflects conditions common to applied settings. Thus, the first purpose of this experiment was to determine if a more ecologically valid, lengthening data-path method of stimulus presentation would result in patterns of visual analysis that are logically expected when data are 100% accurate. Logical patterns of choice with 100% accurate data would provide initial evidence for the construct validity of the lengthening data-path method of stimulus presentation.

Second, we were interested in isolating the influence of data accuracy on decision making. The trend, level, and variability of data paths will influence data-based decision making (e.g., Kipfmiller et al., 2019; Maffei-Almodovar et al., 2017; Sidman, 1960). Thus, the second purpose of this experiment was to determine if controlling these aspects of a lengthening data path, while systematically changing data accuracy, would result in systematic changes in decision making. Demonstrating that choices were likely controlled only by changing data accuracies would provide initial support that data accuracy influences decision making with lengthening data paths.

Lastly, we were interested in whether changing data accuracy could be conceptualized as changing uncertainty about whether a client is actually making progress. Thus, the third purpose of this experiment was to determine if the hyperboloid probability discounting equation would accurately describe each participant's data. In turn, this allowed us to evaluate the extent to which data accuracy influenced visual analysis in a manner consistent with previous research in the probability discounting literature.

Behavior analysts make multiple, daily, programmatic decisions for the many clients they serve. Explicit demonstration of precisely how basic determinants of choice influence practice-related decisions is relatively understudied. This study adds to a growing body of literature that extends basic determinants of choice to addresses socially relevant problems (see Mace & Critchfield, 2010; Poling, 2010). In particular, this study seeks to take an initial step toward bridging existing methodology in basic choice research (probability discounting) with variables that may affect decisions made by practicing behavior analysts during practice-related activities (visual analysis).

Method

Participants

We recruited 30 students enrolled in graduate-level coursework in behavior analysis from several universities throughout the midwestern and northeastern regions of the United States. Participants were pursuing terminal MA degrees or certificate programs that fulfilled coursework requirements for the Board Certified Behavior Analyst® credential. The median age of participants was 25 years (range: 22–39). Twenty-seven participants self-identified as female and three as male. Twenty-two participants self-identified as non-Hispanic or non-Latinx white, seven participants self-identified as Hispanic or Latinx white, and one participant self-identified as Hispanic or Latinx black or African American. Twenty-eight participants had previous experience with visual analysis of single-case data, and the average visual analysis experience for all participants was 1.67 years (range: 0–7 years).

Participants were recruited through advertisements distributed through email or course websites indicating that course credit could be earned by participating in a study about decision making in ABA. Participants were provided with a website link to the Qualtrics online platform specific to this study (Qualtrics, Provo, UT, 2019). Participants first completed an informed consent document, then the lengthening data-path tasks, and finally a demographic survey.

Data Paths and Graph Development

We created five unique data paths using the following autoregressive formula in Microsoft Excel (Kipfmiller et al., 2019; Wolfe & Slocum, 2015):

$$y = mx + b + \varepsilon. \tag{2}$$

Here, m is equal to the slope of the data path, b is equal to the y-intercept, and $\boldsymbol{\varepsilon}$ is a variability term calculated by the equation:

$$\varepsilon_i = (\alpha^* \varepsilon_{i-1}) + u. \tag{3}$$

For the variability term, α represents an autocorrelation value, ε_{i-1} is the variability from the previous observation in the time-series, and u is a random number between -20 and +20. Input values for the formula were as follows: $m = 4$, $b = 20$, and $\alpha = 1$. Each input generated 10 unique data points that trended upward to simulate the acquisition of a target behavior. The combination of slope and y-intercept were chosen to ensure that each data path never reached above 90% correct responding, which may lead to intervention decisions influenced by skill mastery rather than data accuracy.

The created data paths were then pilot tested with participants who were not recruited for and did not participate in the present study. The pilot study allowed us to measure the likelihood that intervention choices were controlled by the characteristics of the data path rather than the percentage of data accuracy. During pilot testing, we found that presenting a datum below the percentage of correct responding from the previous session (i.e., previous datum) influenced the intervention choices of many participants. As a result, each data path was modified so that each unique presentation of a datum represented only an increase in the percentage of correct responding from the previous session. When data were manually changed, the datum value became the average percentage correct value using the datum before and the datum after the value that we manually changed. For example, if the autoregressive model generated the consecutive values of 30, 25, and 50, the value 25 was manually changed to 40 (i.e., [30 + 50]/2 = 40). If the final value in the series was at or below the previous value, the final value was changed to the average between the previous datum value and 100.

Each data path was graphed in a time series format to provide a visual depiction of those data. Ten graphs were developed for each data path where each graph consisted of 1–10 data points to represent the dynamic accrual of data over 10 intervention sessions. All 10 graphs contained data paths with an increasing trend to simulate skill acquisition. The first graph contained the first data point in the series; the second graph contained the first and second data points in the series, and so on. No two graphs in each data series had the same number of data points. See Fig. 2 for examples, and for further information regarding labels and scales associated with each graph. In total, 50 unique graphs were developed for the study: 10 graphs for each of the five unique data paths. All plots can be found at: https://github.com/davidjcox333/Data-Accuracy-ABA-Decision-Making.

Procedure

Lengthening data-path task Participants completed five lengthening data-path tasks. The lengthening data-path tasks were presented in the survey program Qualtrics. After consenting to participate in the study, participants were provided only with the following instructions:

> For this task, you will be presented with a series of graphs that show the skill acquisition of a hypothetical client who is receiving behavioral intervention services. On each page, you will be asked to make a choice between continuing the hypothetical intervention or modifying the hypothetical intervention. All data and choices in this task are hypothetical, but please respond as though this were an actual case and you were making this decision for an actual client/student. It is also important to note that there are no right or wrong answers. We are interested in what choice you would make given only the available information on the screen.

Each lengthening data-path task consisted of 10 trials and began with a participant observing a graph with a single marker indicating the percentage of correct responses (y-axis) for the first hypothetical intervention session (x-axis; left panel, Fig. 2). Below the graph was a statement indicating how accurate the data were (e.g., "These data were taken with 10% accuracy"). Relative to probability discounting, the statement of data accuracy could be considered the probability that the data reflect true performance.

Each choice in the lengthening data-path tasks contained two possible responses: continue the intervention or modify the intervention (Fig. 2). The language contained in these responses did not change across trials. Based on our collective experiences in research and practice, the options to either continue or modify an intervention reflect common decisions made as a function of the ongoing review and analysis of data. In addition, we restricted the available choices to two options so we could measure a single decision-point where participants switched from not doing something (continuing the intervention) to doing something. In turn, this reduced the potential complexity of data interpretation for this initial, translational study in this domain. For each choice, the participant clicked a button to either "Continue intervention" or to "Intervene with intervention modifications" (grey buttons, Fig. 2). Following the participant's choice to continue or modify the intervention, the next trial was presented wherein one additional datum was added to the data path on the graph. The participant was then asked to again choose between continuing or modifying the intervention.

Data accuracy remained fixed and the above process continued until the participant was exposed to 10 trials (10 opportunities to make a choice) regardless of what option the participant selected on each trial. If at any time the participant selected "Intervene with intervention modifications," they were still exposed to the subsequent trials remaining in that

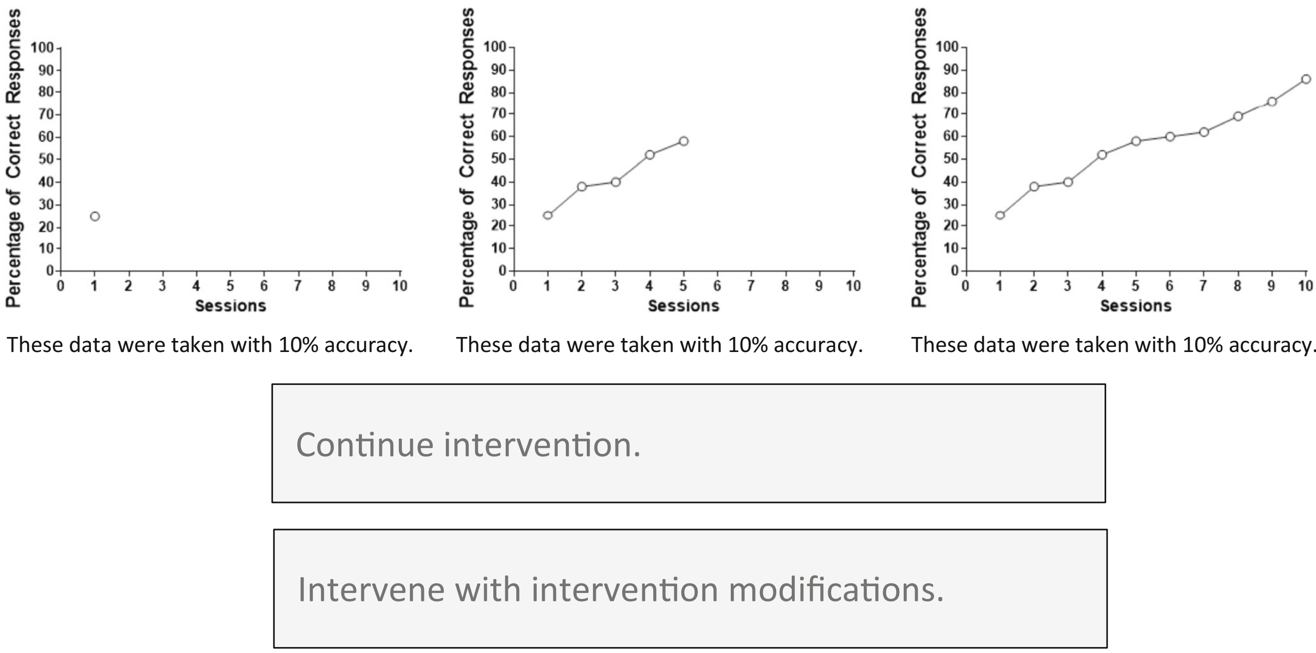

Fig. 2. Three screenshots of one discounting task participants may have observed in the experiment. Each task always began with a single data point (left panel) and participants chose whether to continue the intervention or to modify the intervention (buttons shown below the three screenshots). The number of data points on the plot increased by 1 following each choice (e.g., fifth choice of task shown in middle panel) until 10 sessions were reached (right panel). Below each plot was a statement regarding the accuracy of the data in the plot

task (with the exception of the 10th, or final, trial in the task). This was done to avoid creating a contingency where choosing to modify the intervention would lead to more rapid completion of the task.

Each participant repeated the lengthening data-path task for five different percentages of data accuracy (100%, 60%, 40%, 20%, and 10% accuracy). These probabilities were chosen based on the number of probabilities, range of probabilities, and step-size between probabilities that have shown to be successful in studying the hyperbolic nature of probability discounting in past research (e.g., Green et al., 1999; McKerchar, Green, & Myerson, 2010; McKerchar & Renda, 2012). For each of the five data paths, the data accuracy was shown below the lengthening data path as exemplified in Fig. 2.

Participants were randomly allocated to one of five different conditions. Each condition randomly combined one degree of data accuracy with one of the five unique data paths so that each participant made decisions with each unique data path. Randomization was conducted without replacement. This randomization was used to prevent overlap between each unique data path and their assigned accuracy across conditions. Each participant was presented a total of 50 trials: 10 trials within each of five unique data paths and their associated data accuracy. All experimental materials can be found online in Supplemental Information.

Data Analysis

One potential concern is that trend and level may influence decisions based on data paths that have an increasing or decreasing trend. That is, rather than data accuracy controlling choice throughout all 10 trials, the combination of trend and level may differentially control responding on earlier choice trials in the task (e.g., when only three data points are present) than in later choice trials in the task (e.g., when eight data points are present). Therefore, the first analysis of the data sought to identify whether the increasing trend and level of each data path confounded the influence of data accuracy on participant choice. To do this, we identified the first session at which each participant chose to intervene during the 10-trial block. Three possibilities seemed likely. First, if trend and level controlled decisions to continue or modify an intervention (and not data accuracy), we would expect participants to modify the intervention at approximately the same trial number across all tasks because we used the same autoregressive parameter values to design all five data paths. Second, if data accuracy controlled decisions to continue or modify an intervention (and not level or trend), we would expect participants to choose to modify the intervention at different trial numbers across all tasks. Third, if data accuracy, trend, and level interacted to control decisions to continue or modify an intervention, we would expect participants to choose to modify the intervention at different trial numbers across all tasks. Note that by examining binary choice at only the individual level, the second and third possibilities are difficult to differentiate from each other.

One way to differentiate the second and third possibilities above is to plot the proportion of participants who chose to intervene at each trial number for each lengthening data-path task. If data accuracy were the primary variable controlling decisions to continue or modify an intervention (and not trend or level), we would expect a consistent proportion of the participants to modify an intervention throughout all 10 trials because data accuracy remained unchanged. However, if trend and level interact with data accuracy to control decisions to continue or modify an intervention, we would expect a changing proportion of the participants to modify an intervention throughout all 10 trials as trend and level change. Thus, the first analysis of the data examined the slope of trend lines in the proportion of participants that chose to intervene at each choice trial. The slope was obtained using the "SLOPE" function in Microsoft Excel. The "SLOPE" function works by first generating a linear fit to the data via the least squares method followed by calculating the change in y divided by the change in x over the x-range of the observed data. This allowed us to determine if data accuracy was isolated as the variable that influenced participant choice.

We also conducted statistical analyses to test the independent and interactive effects that data accuracy and data path had on participant decisions to intervene. Because there were multiple independent variables (data accuracy and unique data path), we used multivariate regression to build models that predicted our dependent variable (trials to intervene). If data accuracy was the only variable influencing choice, we would expect a significant main effect of data accuracy, no significant main effect of the unique data path, and no significant interaction between data accuracy and unique data path. A multivariate logistic regression model was computed to predict the influence of data accuracy and unique data path on the proportion of participants that chose to intervene. A multivariate linear regression model was computed to predict the influence of data accuracy and unique data path on the number of trials participants waited to intervene Statistical analyses were conducted using the statsmodels package version 0.11.1 (Seabold & Perktold, 2010) for Python version 3.8.3.

The second analysis of the data sought to evaluate how data accuracy influenced each participant's individual choice to continue or intervene. We identified the first trial where each participant chose to "Intervene with intervention modifications" in each lengthening data-path task. We then plotted this trial number as a function of the percentage of data accuracy when expressed in odds against (as is common to probability discounting literature; for review, see McKerchar & Renda, 2012). Plotting the data in this way allowed us to visually analyze how a parametric reduction of data accuracy influenced the likelihood that each individual participant would "intervene."

For the final analysis of the data, we determined how well data were described by Equation 1. To do this, we used the Microsoft Excel Solver Add-In to fit Equation 1 to the set of the five identified switch points (i.e., the trial at which the participant chose to intervene) within each lengthening data-path task for each individual participant as well as to the group average switch points. To fit Equation 1, the value of A was set equal to 10 because 10 is the maximum number of trials a participant could be exposed to without choosing to intervene. Analyzing our findings in the context of Equation 1 allowed us to evaluate whether the probability discounting framework described individual participant choice, as well as to measure probability discounting at the group level as is common in probability discounting literature.

Results

Figure 3 depicts the proportion of participants who chose to intervene in each data accuracy condition. The upper-left panel shows the proportion of participants who chose to intervene at any point during the 10 trials of the associated data accuracy condition. Increasing data accuracy led to a systematic decrease in the proportion of participants who chose to intervene. In addition, there was no data accuracy condition where we observed a single, unanimous choice. By definition, a proportion of participants means that some participants never intervened and some did. The highest proportion of participants that intervened (83%) occurred when data accuracy was stated as 10%. The lowest proportion of participants that intervened (40%) occurred when data accuracy was stated as 100%. This visual analysis was supported statistically by the multivariate logistic regression analysis where, controlling for the unique data path, increasing data accuracy led to a reduction in the number of participants who chose to intervene (β = -2.25, $p < 0.001$).

It is possible that the unique characteristics of each data path influenced participant choices to intervene. The lower left-panel in Fig. 3 shows the proportion of participants who chose to intervene at any point during the 10 trials as a function of the unique data paths. The unique data paths are categorical so the ordering is irrelevant. The highest proportion of participants that intervened (70%) occurred with data path C and the lowest proportion (53%) occurred with data path B. Though this difference of 17% indicates a difference of five participants, statistically there was no main effect data path type (β = 0.07, p = 0.55), likely because of the similarity in the proportion of participants who intervened for the remaining three data paths and data path C. Lastly, accuracy and data path did not interact to influence whether participants chose to intervene (β = -0.53, p = 0.20).

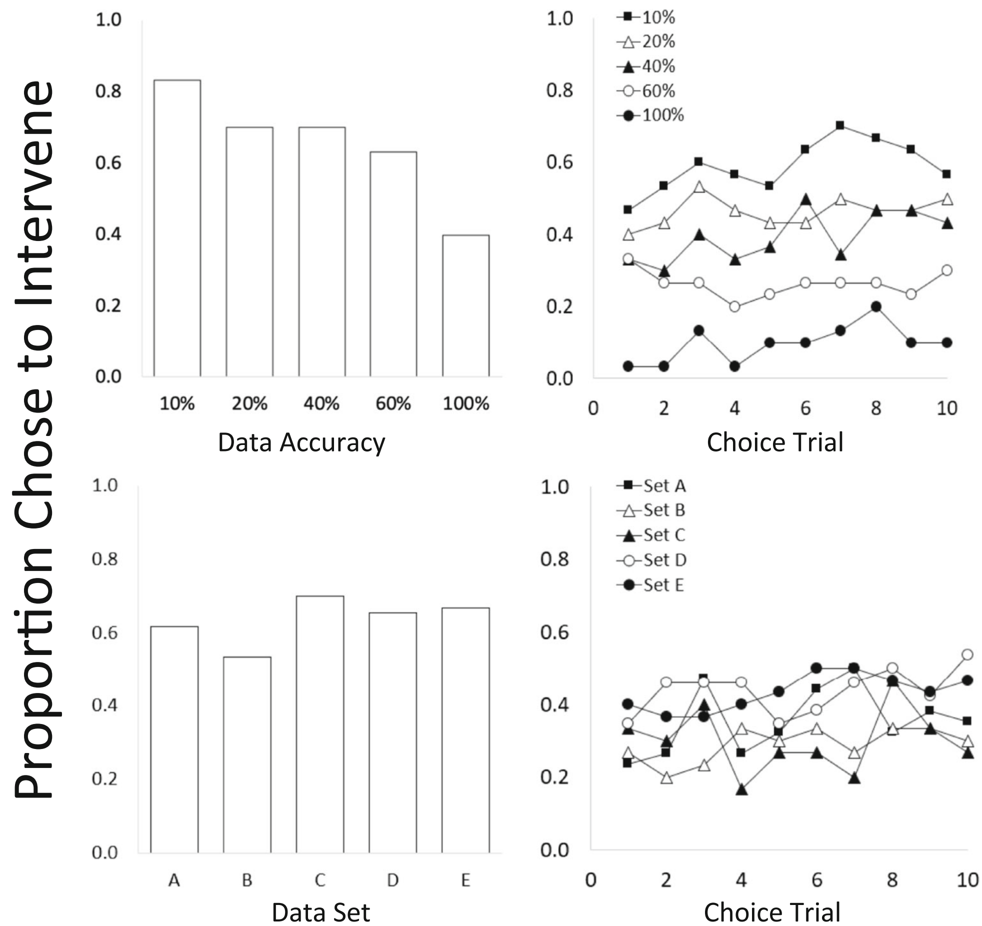

Fig. 3. Proportion of participants that chose to "intervene" based on data accuracy (top panels) or based on the data path (bottom panels). The left panels show the overall proportion of participants (y-axis) who chose to intervene at any point during the lengthening data-path task based on data accuracy (x-axis, top-left panel) or the lengthening data path set (bottom-left panel). The right panels show the proportion of participants who chose to intervene at each choice trial based on data accuracy (upper-right panel, different markers) or the lengthening data path set (bottom-right panel)

Nevertheless, it is possible that data accuracy interacted with trend and level to influence how quickly participants chose to intervene. Therefore, we also plotted the proportion of participants who chose to intervene as a function of each trial number in the upper right panel of Fig. 3. In the upper right panel of Fig 3, each data path represents a different data accuracy used for a lengthening data-path task (e.g., 10% data accuracy, 20% data accuracy). There are three primary observations about the data in Fig. 3. First, the proportion of participants who chose to intervene at each choice trial did not appear to be increasing or decreasing. With 30 participants, a systematic increase or decrease by 1 participant in the decision to continue or intervene (overall within the group) would equate to a slope of $1/30 = \pm 0.033$. The steepest slope observed was in the 40% data accuracy condition and equaled 0.016. Second, a systematic decrease in data accuracy across lengthening data-path tasks led to a systematic increase in the proportion of the participants that chose to intervene earlier in that task (at an earlier trial). Third, there was little overlap between most data paths in Fig. 3, especially with higher data accuracy. The condition with the lowest overlap was the 100% condition, which did not overlap with any other conditions in Fig. 3. The condition with the greatest overlap was the 40% accuracy condition which overlapped with four data points from the 20% accuracy condition and one data point from the 60% accuracy condition. The multivariate linear regression analyses support the visual analysis. Controlling for unique data path, increasing data accuracy led to an increase in the number of trials the participants waited to intervene ($\beta = 5.09$, $p < 0.001$).

The lower right panel in Fig. 3 shows the proportion of participants who chose to intervene as a function of each trial number where each data path represents a different unique data type. Similar to the upper right panel in Fig. 3, there were no significant trends observed as a function of the unique data paths. In addition, there was considerable overlap between the data paths for the majority of unique plots (i.e., Sets A–E). The one exception is the absence of overlap between the unique data paths presented to participants as Set B and Set E. Results of the multivariate linear regression analysis suggests that the unique data paths presented to participants did not influence the trial at which they chose to intervene ($\beta = -0.17$, $p = 0.46$) nor was there an interaction between data accuracy and unique data paths ($\beta = 1.26$, $p = 0.09$). Together, visual and statistical analyses using Fig. 3 suggest that participant choice in this

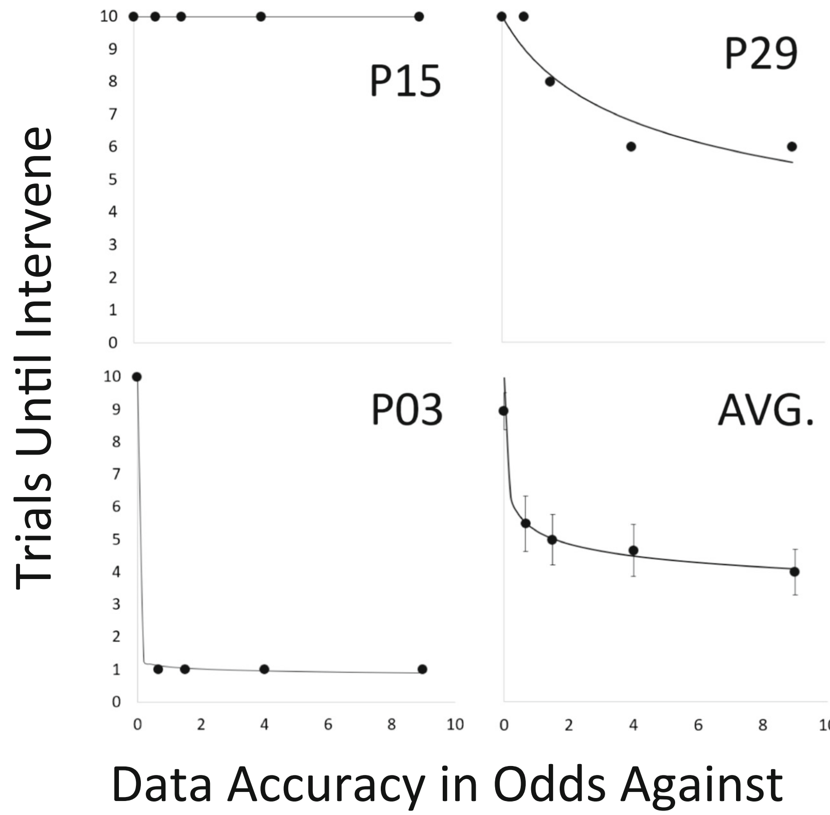

Fig. 4. Probability discounting plots showing the session at which participants first chose to intervene (*y*-axis) as a function of data accuracy expressed in odds against (*x*-axis; (1-*p*)/*p* where *p* = probability the data are accurate). The top left, top right, and lower left panels are from the participants who showed the lowest, median, and highest rates of discounting. The bottom right panel shows the average session number that all 30 participants chose to intervene. Solid lines represent the best fit of Eq. 1 to the observed data. Error bars represent SEM

experiment was likely controlled primarily by data accuracy and not the changing level and upward trend of the lengthening data paths used to present choice trials to the participants.

Fig. 4 provides three examples of how data accuracy influenced participant decisions to either continue or modify the intervention. To orient readers to Fig. 4, high data accuracy corresponds to a low odds against (e.g., data accuracy of 100% equals an odds against of 0). Plots with titles "P##" represent the individual participants showing the least (Fig. 4, upper left), median (Fig. 4, upper right), and highest (Fig. 4, lower left) influences of decreasing data accuracy on intervention decisions. The lower right plot in Fig. 4 shows the average number of trials until all 30 participants chose to intervene in each task. Overall, these plots depict the range of patterns of individual decision making for whether to continue or modify the intervention as a function of decreasing data accuracy.

To summarize the patterns of responding we observed, some participants chose never to intervene, regardless of data accuracy (e.g., P15; $n = 5$); some participants always intervened if the data were 60% accurate or lower (e.g., P03; $n = 7$); and some participants showed a systematic reduction in the trials until they intervened as a function of systematically reducing data accuracy (e.g., P29; $n = 18$). Overall, reducing data accuracy reduced the number of trials until participants chose to modify the intervention (lower right panel in Fig. 4), but the rate that data accuracy decreased trials to intervention modification differed between individual participants.

We also fit Eq. 1 to each individual participant's data and to data aggregated at the group level. The solid line in each plot in Fig. 4 shows the best fit of Eq. 1 to that set of indifference points (the trial at which participants chose to intervene). Five of the 30 participants (17%) did not change the trials until they intervened as a function of data accuracy (e.g., P15). For four of these five participants (those who continued the intervention throughout), the fit of Eq. 1 becomes uninterpretable.

Fig. 5 shows four plots that can be used to evaluate how well Eq. 1 described participant choices. The upper left panel shows R^2 values. R^2 is a general metric that is often a first look at how well a model describes a set of data. An R^2 value of 1.00 would indicate the model perfectly describes all the data. Thus, the closer that R^2 is to 1.00 when fitting Eq. 1 to an individual participant's data, the better the hyperbola-like equation describes that participant's changes in choice to intervene as a function of changes in data accuracy. The best fit of Eq. 1 led to a negative R^2 value for six participants (20%), which indicates that the average number of trials until the participant intervened was a better description of their pattern of choices than by using Eq. 1 to describe their choices. For the remaining 20 participants (63%), the minimum, median, and max observed R^2 values were 0.05, 0.73, and 0.99, respectively. At the group level, Eq. 1 resulted in an R^2 value of 0.93 (see Fig. 4, lower right panel).

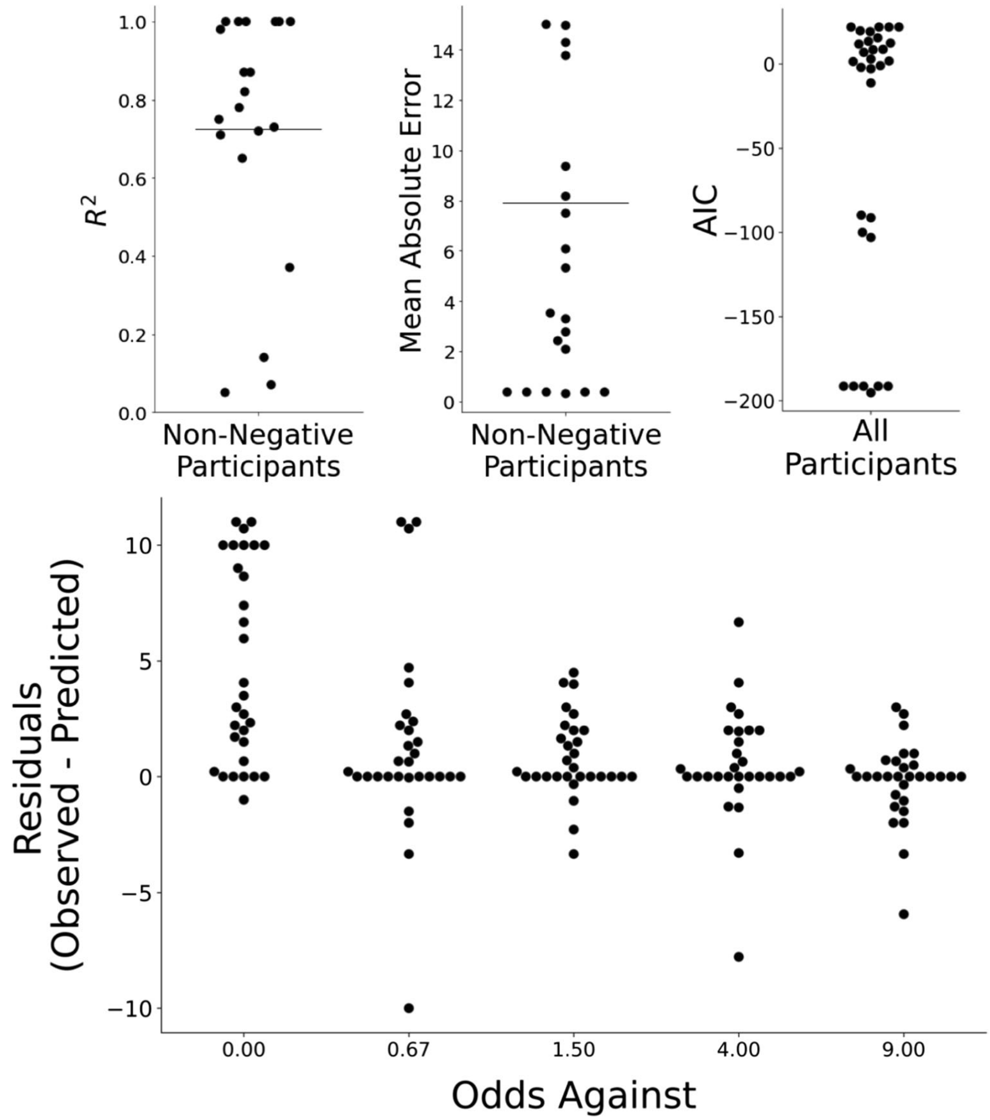

Fig. 5. R^2 (upper-left panel), mean absolute error (upper-middle panel), AIC (upper-right panel), and residual (lower panel) values for the fit of Eq. 1 to individual data. Solid lines in the left two panels represent the group median

The upper middle panel in Fig. 5 shows the mean absolute error when Eq. 1 was fit to individual participant data. Each marker represents the mean absolute error (MAE) from one participant for whom Eq. 1 did not result in a negative R^2 value. For this study, MAE represents the average difference between the trial the participant chose to intervene and the trial that Eq. 1 predicted the participant would intervene. The MAE was fewer than four trials for 11 of the 20 participants with nonnegative R^2 values and less than one for 6 participants.

The upper right panel in Fig. 5 shows the Akaike's Information Criterion (AIC) when Eq. 1 was fit to individual participant data. Each marker represents the AIC from a single participant. AIC provides an estimate for how much information is lost by using a model with free parameters compared to simply using the data itself. AIC is interpreted relatively such that lower AIC values indicate less information loss compared to higher AIC values (i.e., the model fits that data better when AIC is lower). Mirroring the results from the R^2 and MAE plots, AIC values suggested Eq. 1 described responding very well for six participants (those at the bottom of the panel), well for four participants (those in the middle), with the remaining fits being bunched between -11.31 and 21.98.

Finally, the lower panel in Fig. 5 shows the residuals at each odds against when Eq. 1 was fit to all individual participant data. Residuals are calculated by substracting the predicted trial the participant would intervene from the actual trial that the participant intervened. For unbiased models, residuals do not have an increasing or decreasing trend as the x-axis values change and the residuals should be randomly distributed and follow a Gaussian distribution centered on 0 about the y-axis. Overall, we observed reductions in the residuals from more positive to closer to zero moving from data accuracies of 100% (0 odds against) to 10% accurate (9 odds against), respectively. Related to this, the residuals did not approximate a normal distribution centered on 0 on the y-axis. Thus, at lower odds against accurate data, the residual plot indicates the model was bias in underpredicting how many trials until each participant chose to modify the intervention.

Discussion

This proof of concept experiment tested an ecologically valid methodology of data presentation to assess choice between two intervention decisions, to continue or modify an intervention, based on visual analysis of dynamically updating time-series data. The ecologically valid methodology presented graphs that incrementally increased by one datum following each choice trial. We used this methodology to examine how systematic manipulation of data accuracy would influence intervention decisions. Analysis of the results of this experiment, the implications for future translational research, and implications for practice, are discussed in turn.

Lengthening Data Path Methodology

Past research on visual analysis of time-series data in ABA has primarily focused on complete data sets involving within-subject test and control conditions (e.g., Fisher et al., 2003; Kahng et al., 2010; Ninci et al., 2015). The presentation of completed data sets is an excellent way to examine and teach individuals to recognize intervention effects and experimental control. Nevertheless, the majority of intervention decisions made by behavior analysts working in applied settings are unlikely to involve completed data sets. Rather, behavior analysts likely make decisions about an intervention based on incremental additions of one or more data points to a data path. Thus, one contribution of the present experiment is the extension of a more ecologically valid research methodology for presenting time-series data to behavior analysts in a manner likely to be encountered in applied settings (Saini et al., 2018; Vanselow et al., 2011).

When using any research methodology, it is important to ensure that the chosen method allows the researcher to isolate the independent variable of interest for a research question. For the present experiment, our independent variable was data accuracy. Visual analysis of behavioral data that is not at steady state is likely to be influenced by characteristics of the data path (e.g., trend, level, and variability; Bourret & Pietras, 2013). We controlled for the influence of variability on intervention decisions using an autoregressive equation and manual manipulation of individual datum through pilot testing. However, it was possible that the trend and level of each data path (and not data accuracy) controlled participant responding in the present experiment. To evaluate the potential presence of this confound we identified the first session at which each participant chose to intervene during a 10-trial block. Because we used the same autoregressive parameter values to design all five data paths, if trend and level were the primary variables controlling decisions to continue or modify an intervention (and not data accuracy), we would expect participants to choose to modify the intervention at approximately the same trial number across all tasks. We did not observe this pattern of responding. Instead, we found that the proportion of participants who chose to intervene remained relatively constant across all 10 trials and that systematically decreasing data accuracy systematically increased the proportion of participants who intervened. These data suggest this methodology allowed us to isolate the influence of data accuracy on visual analysis from the influence of trend, level, and variability of the data paths.

The method used in the present experiment does not suggest that characteristics of the data path (e.g., trend, level, and variability) do not affect choice when visually analyzing data that are not at steady state. Previous research has reported that variations in trend, level, and variability may differentially affect determinants of an experimental effect (e.g., DeProspero & Cohen, 1979; Diller et al., 2016; Kahng et al., 2010) or clinical outcomes (Kipfmiller et al., 2019). An important area for future research is to systematically manipulate trend, level, and variability of lengthening data paths to determine the independent and interactive influence of these characteristics on intervention decisions. Nevertheless, this study was designed to examine how one fundamental characteristic of data trustworthiness—data accuracy—influences decision making rather than as a comprehensive examination of all variables that might influence decision making in nonlaboratory settings.

The methodology reported in the present experiment could also be used to examine how other aspects of the intervention context influence intervention decisions (e.g., Saini et al., 2018; Vanselow et al., 2011). It seems likely that each intervention decision is controlled by multiple aspects of the intervention context. For example, choosing whether to modify an intervention, as well as how to modify an intervention, may depend on the skills and abilities of the direct care employee; the experience and background of the supervising behavior analyst; availability of resources required to implement different intervention modifications; the relationship between the target program and additional behavior targets within an entire ABA program; and client preferences and cultural background. Future translational research could examine how intervention decisions are influenced by each of the above in isolation as well as how they interact to influence intervention decisions.

Influence of Data Accuracy on Choice

The method used in the present proof of concept experiment allowed us to isolate the influence of data accuracy on intervention choice. We found that data accuracy systematically affected the majority of participant's choices during visual analysis. When data accuracy was high (e.g., 100%), most participants were more likely to continue the intervention throughout the task. However, when data accuracy was low (e.g., 10%), most participants were more likely to choose to modify the intervention after a few trials. It is important to note that although decreasing data accuracy led most participants to intervene after fewer trials, how quickly they chose to intervene varied across participants. Thus, the present experiment suggests that data accuracy influenced intervention decisions for most participants in our experiment, but the degree to which data accuracy influenced decisions likely differed between any two participants. Exactly how variable intervention decisions are in applied contexts is currently unknown. It is also unknown what past experiences led some participants to intervene almost immediately whereas the other participants given the exact same information chose never to intervene. Important areas for future research include identifying what experiences and current contingencies may lead to the high variability in practice-related decision making.

Our finding that participants may respond differently to decrements in data accuracy has important practical implications. Behavior analysts working in applied settings may have access to a criterion record to inform data accuracy (e.g., a permanent product). If a criterion record is not available, it is likely that a behavior analyst has access to calculated interrater reliability (IRR) or, where access to IRR is unlikely, the behavior analysts likely has unquantified past experience with each employee with whom they may infer data accuracy. Consider the following example where a supervisor conducts IRR checks with two employees. The IRR between the supervisor and one employee is 100% whereas the IRR between the supervisor and the second employee is 60%. Our tentative findings suggest that, all other variables held equal, one supervisor may respond to the IRR data in a similar fashion when making intervention decisions (e.g., P15 in Fig. 4). That is, they may emit the same clinical decision in the presence of 100% and 60% data accuracy. However, all other variables held equal, a second supervisor may respond to the IRR data differently when making intervention decisions (e.g., P03 in Fig. 4), and emit a clinical decision in the presence of 100% accuracy that is different than what they emit in the presence of 60% accuracy. Given the importance of consistency in applying contingencies for behavior change (e.g., Carroll, Kodak, & Fisher, 2013; Fryling, Wallace, Yassine, 2012; Pence & St. Peter, 2015; St. Peter, Byrd, Pence, & Foreman, 2016), this observed difference in how supervisors respond to different data accuracy could lead to observed differences in intervention outcomes across behavior analysts. Responding for 5 of the 30 participants did not appear to be under the control of the systematic manipulation of data accuracy. Four of those five participants chose to continue the intervention for all 10 trials during all five tasks, and one of the five participants chose to modify the intervention for all 10 trials during all five tasks. It is possible that the characteristics of the data path (gradual upward trend) may have resulted in these observed differences. That is, the four participants who always continued may have chosen not to do anything because the hypothetical client was making progress. It is also possible that past training in visual analysis received by these individual participants resulted in the observed pattern of responding. However, the five participants reported 0, 1, 1, 2, and 5 years of experience with visual analysis that comprised the least (0 years) and second-longest (5 years) duration of experience with visual analysis for the participants in this study. Thus, it is unlikely that simply years of experience played a role. Nevertheless, an important area for future research is to examine how different training experiences affect intervention decisions relative to data accuracy.

Finally, data accuracy is one variable out of many different variables that may influence practice-related decisions by behavior analysts. Other fundamental characteristics of data trustworthiness (e.g., reliability, validity) as well as additional contextual information (e.g., target behavior, upcoming intervention; Vanselow et al., 2011) are likely to influence the practice-related decisions that behavior analysts make. Practice-related decisions are likely controlled by a complex interaction of many different variables and this possibility limits the ability to directly translate the results of this study to applied contexts. Nevertheless, the experimental analysis of behavior has demonstrated success in advancing our understanding of human choice by isolating and empirically studying one independent variable at a time and integrating our understanding of different independent variables over time. This study is in line with that approach to understanding the behavioral processes underlying the multiple control of practice-related decision making.

The Probability Discounting Model

Many different behavioral processes might control practice-related decision making related to data accuracy. As an initial exploration into different behavioral processes, we fit the hyperbola-like probability discounting equation (Green et al., 1999) to individual and group patterns of choice. A common metric used by researchers to initially determine how well an equation describes a pattern of behavior is R^2. If R^2 values suggest a decent model fit, then researchers typically follow-up with more robust metrics for model testing (Dallery & Soto, 2013). R^2 is one way to quantify the difference between the observed level of the data and the level predicted by the model. R^2 values of 1.00 indicate the model perfectly describes the data and R^2 values of 0.0 indicate the model describes the data just as well as taking the mean of the data (i.e., the model does not provide much additional information). In probability discounting research, median R^2 values when fitting Eq. 1 to individual data is typically around or above 0.90 (e.g., $R^2 \approx 0.97$, Rachlin, Brown, & Cross, 2000; 0.94, Shead & Hodgins, 2009).

Equation 1 described the data relatively well for 16 of the 30 participants. For 10 of those 16 participants, Eq. 1 led to R^2 values above 0.80, MAE was less than four trials, and AIC was -90 or below. For 6 of those 10, participants the R^2 values were near 1.00, MAE was less than one trial, AIC values were less than -191, and the residuals were near zero. In total, these results suggest that for a small portion of the participants in Experiment 1 (between 6–10 participants), probability discounting provided a good description of how data accuracy influenced their choice to continue or modify an intervention.

A large body of research has examined how different aspects of the choice context can influence rates of discounting (e.g., DeHart & Odum, 2015; DeHart, Friedel, Frye, Galizio, & Odum, 2018; Naude, Kaplan, Reed, Henley, & DiGennaro Reed, 2018; Yi & Bickel, 2005). In addition, a large body of research has examined how choice with uncertain outcomes differs depending on whether the choices are described verbally or not (description-experience gap, e.g., Asgarova, Macaskill, & Hunt, 2020; Cox & Dallery, 2018; Wulff, Mergenthaler-Canseco, & Hertwig, 2018). Though we explicitly stated the accuracy of the data depicted in our tasks, the extent to which practitioners are explicitly aware of the accuracy of their data is unknown. Although some may view this as a limitation for translating the present study to practice-related settings, practitioners who are unaware of the accuracy of their own data seems more of a limitation of their practice—not this study. Data accuracy can be calculated by comparing observed behavior to some form of a criterion record (e.g., a permanent product) or, if a criterion record is not available, data accuracy may be inferred by reducing threats to internal validity associated with measurement error or through measures of observer reliability. The present experiment combined with past research on discounting and the description–experience gap suggests the way that data accuracy is described, or whether it is described at all, may influence intervention decisions. Each of these basic research areas provide a wealth of potential variables known to influence choice that could be examined in translational research on clinical decision making by behavior analysts.

The fits of Eq. 1 were poor for 14 of the 30 participants (47%). R^2 values were between 0.00 and 0.40 for four participants, were negative for six participants, and Eq. 1 was uninterpretable for four participants because data accuracy did not change their decision to continue the intervention. In addition, MAE increased beyond 10 indicating the model was off by a greater number of trials than the participants completed in the study; and the residuals were not distributed randomly, did not follow a Gaussian distribution, and were not always centered around 0 on the *y*-axis. In total, these results suggest that discounting is unlikely to be *the* behavioral process underlying practice-related decisions related to data accuracy for these 14 participants. An important topic for future research could be what variables lead data accuracy to control some behavior analyst intervention decisions hyperbolically, but not others (see McKerchar & Renda, 2012, for discussion of why hyperbolic patterns are important).

Alternative Models

Discounting is arguably the currently most commonly studied behavioral process in behavior analysis and thus we chose to examine the extent to which discounting plays a role in practice-related decisions involving data accuracy. But practice-related decisions are likely to be multiply controlled, and many additional or alternative behavioral processes might control practice-related decision making. For example, the influence of data accuracy on practice-related decision making could be studied from an information theoretic approach. Information theoretic approaches often focus on variables that influence an organism's ability to detect a signal as a function of the amount of noise or distractor stimuli increases (e.g., Gallistel, 2012; Ward, Gallistel, & Balsam, 2013). Here, reducing data accuracy might be analogous to increased noise and the signal might the "true progress" being made by the client.

The influence of data accuracy on practice-related decision making could also be approached from a probability matching perspective (e.g., Estes, 1976; Staddon, Hinson, & Kram, 1981). A probability matching approach would involve providing feedback to participants following each choice as to whether the choice they made improved data accuracy or influenced the trend in the data. With repeated trial presentation, one could determine whether the relative occurrence of positive outcomes for each response option (i.e., improved data accuracy or improved client progress) correspond to the relative allocation of responding to each response option.

Rule-governed behavior (Hayes, 1989) is a final example of alternative approaches to studying the influence of data accuracy on decision making. Given the idiosyncratic results of some participants, it is possible that participant behavior may have been at least partially under the control of verbal rules informed by prior experiences in applied settings or graduate coursework. Though we are not aware of an agreed-upon and explicit standard percentage of data accuracy in behavior-analytic research or practice, participants may have derived their own individual rules from standards commonly applied to IRR, where 80% IRR serves as an agreed upon baseline level of acceptability in single-case research. Future research could examine the isolated influence of rules on decision making relative to data accuracy and how those rules interact with contacted contingences.

Limitations

There were several limitations to the present experiment. One limitation was that we restricted participant choices to two options—continue the intervention or modify the intervention. These two choices likely do not represent the range of choices available during visual analysis (Kipfmiller et al., 2019; Maffei-Almodovar et al., 2017). Given the relative novelty of the lengthening data path procedure and our focus on studying the relevance of discounting to this decision-making context, we restricted choice to two options in order to extend previous methods from discounting research that have successfully identified indifference points for model fitting (e.g., Du, Green, & Myerson, 2002; Kirby & Marakovic, 1996). We also limited choice options to evaluate how systematic manipulation of data accuracy might differentially control participant responding. Given our findings that data accuracy differentially affected participant responding, future research may evaluate how the systematic manipulation of data accuracy influences choice using a variety of options (e.g., conduct an IRR check, provide feedback to data collector, discontinue the intervention). In addition, future research would benefit from a variety of methods of choice presentation given it is unlikely discounting is the behavioral process underlying these decisions for all participants.

A second limitation is the hypothetical nature of the choices in this proof of concept study. It is unclear whether participants would make the same choices were they presented with similar decisions and actual clients' data. However, it is also unclear it would be ethically permissible to manipulate data accuracy with an actual client to observe the effect on intervention decisions. One possible avenue for future research would be to take a descriptive approach and collect data on intervention decisions over time along with characteristics of the decision context (e.g., data accuracy, variability, trend, number of sessions, number of open programs). These data, although more complex and unable to demonstrate causal relationships, could provide important information for researchers interested in conducting laboratory studies on visual analysis of single-case intervention data.

Another limitation is that participant responding may have been under the control of a specified data accuracy criterion (e.g., 90%) or rule statement. That is, a participant may have exclusively chosen to continue the intervention for all 10 trials when the probability of accurate data was 100%, but intervene at the first session when the probability of accurate data was 60% or lower. This pattern of choice making was observed for 6 of 30 participants. However, the average number of trials until all participants intervened when data accuracy was 60% or lower was approximately four to six trials. Thus, the majority of participants did not intervene immediately with data accuracy at 60% or lower. This indicates that a specified criterion or rule statement likely did not exclusively control the point at which participants chose to intervene. Nevertheless, the extent to which individual participants were using a rule to guide their responding is unknown. An important avenue for future research would be to determine the extent to which intervention decisions are influenced by self-generated or experimenter provided rules about what to do when data accuracy is at varying levels.

A fourth limitation is that we used five data paths with a similar upward trend and low variability that do not resemble the range of visual data displays commonly observed in research and practice. We used these data path characteristics because pilot data indicated that participant choice was sensitive to data variability and higher levels of percentage of correct responding. In particular, during pilot testing, participants often chose to modify the intervention if variability led to large reductions of correct responding from one choice trial to the next. Likewise, during pilot testing, participants often chose to modify the intervention if the percentage of correct responding reached high levels (e.g., 90%). Both observations during pilot testing make sense. It seems reasonable to modify an intervention if the client shows no change or worsens performance. Likewise, it seems reasonable to modify an intervention (in this case, discontinue) if the client reaches a predetermined mastery criterion (e.g., a high percentage of correct responding). Because we sought to isolate the influence of data accuracy on choice, we controlled for variability and uppermost level when designing data paths. Nevertheless, compared to data in the current study, data in applied settings may involve greater variability, reducing trends, different dependent variables from percent correct responding, and varied graphical displays of experimental design (e.g., A-B-A-B, multi-element) or graph characteristics (e.g., a standard celeration chart). These additional variables provide many future avenues for research on behavior analyst decision making using the lengthening data path design.

On a similar note, the time-series graphs we used in this study depict only one type of visual arrangement and contextual features. We chose these graphs because they allowed us to isolate our independent variable of interest (a fundamental characteristic of data trustworthiness—accuracy) without having to account for additional variables added through including multiple intervention conditions (e.g., a baseline condition) and contextual features (e.g., intervention objective, target behavior). Also, in educational settings, assessment may serve as the baseline measure and may not often be depicted in skill acquisition graphs explicitly. Nevertheless, baseline data and other points of comparison (e.g., assessment data) are important for determining intervention effects using visual analysis. Future research could continue to extend the current methods to more complex visual displays of data to evaluate how decision making is influenced by an interaction among data accuracy and data from multiple intervention conditions.

A final limitation to the present study was the sample used. We used a convenience sample consisting of graduate students in behavior analysis with an average of 1.67 years of experience with visual analysis. The variety of demographic information we collected was limited to age, ethnicity, race, self-identified gender, whether the participant has past visual analysis experience (Yes/No), and their years of experience in behavior analysis. The idiosyncratic nature of our findings suggest that additional variables, not accounted for in this study, may at least partially affect participant responding. It is possible that people with more visual analysis experience of intervention data would have demonstrated different patterns of choice than the current sample. Of particular interest might be experiments that obtain more open-ended responses as to what decision each person would make as each data point is added to a lengthening data path. In addition, data could be collected on what additional information about the client they may gather before they decide and graph characteristics they attend to when making decisions. Such research may lead to future systematic explorations of more complex decision-making environments in ABA.

Lastly, we examined only one characteristic of trustworthy data and provided little additional contextual information that are likely to interact to multiply control decision making in nonlaboratory contexts. As with other areas of professional decision making in health-care contexts, a plethora of opportunities exist to systematically explore how different variables influence practice-related decision making in isolation, as well as how these variables combine to multiply control decision making in complex and dynamic environments.

Conclusion

Past research on visual analysis has primarily focused on decisions with complete data sets. Less research has focused on how data accuracy influences intervention decisions with data paths that lengthen from plotting additional data—which are commonly encountered in research and practice. We sought to fill this gap by parametrically analyzing how reductions in data accuracy influenced the point when participants modified an intervention as we added data to a lengthening data path. For 25 of 30 participants, systematic reductions in data accuracy led to systematic reductions in the number of trials until they modified a hypothetical intervention. It is unclear the probability discounting equation is the best way to quantitatively describe intervention decisions relative to data accuracy. Nevertheless, the lengthening data-path method used in the current study provides an ecologically valid procedure that allows researchers to study behavior analyst intervention decision making with data presentations common to research and practice.

Compliance with Ethical Standards The researchers do not have any potential conflicts of interest to report. All research procedures were performed in accordance with the 1964 Helsinki declaration, its later amendments or comparable ethical standards, and informed consent was obtained from all individual participants included in the study.

Data Accessibility Raw data can be accessed at the following repository: https://github.com/davidjcox333/Data-Accuracy-ABA-Decision-Making

References

Asgarova, R., Macaskill, A. C., & Hunt, M. J. (2020). Gain-loss asymmetry in experiential probability discounting. *The Psychological Record, 70*, 359–371. https://doi.org/10.1007/s40732-020-00379-1.

Baum, W. M. (2018). Three laws of behavior: Allocation, induction, and covariance. *Behavior Analysis: Research & Practice, 18*, 239–251. https://doi.org/10.1037/bar0000104.

Blair, B. J., Tarbox, J., Albright, L., MacDonald, J. M., Shawler, L. A., Russo, S. R., & Dorsey, M. F. (2019). Using equivalence-based instruction to teach visual analysis of graphs. *Behavioral Interventions, 34*, 405–418. https://doi.org/10.1002/bin.1669.

Bourret, J. C., & Pietras, C. J. (2013). Visual analysis in single-case research. In G. J. Madden, W. V. Dube, & K. A. Lattal (Eds.), *APA handbook of behavior analysis, Vol. 1: Methods and principles* (pp. 199–218). Washington, DC: American Psychological Association.

Boykin, R. A., & Nelson, R. O. (1981). The effects of instructions and calculation procedures on observers' accuracy, agreement, and calculation correctness. *Journal of Applied Behavior Analysis, 14*, 479–489.

Bruce, J. M., Bruce, A. S., Catley, D., Lynch, S., Goggin, K., & Jarmolowicz, D. P. (2016). Being kind to your future self: Probability discounting of health decision-making. *Annals of Behavioral Medicine, 50*(2), 297–309. https://doi.org/10.1007/s12160-015-9754-8.

Carroll, R. A., Kodak, T., & Fisher, W. W. (2013). An evaluation of programmed treatment-integrity errors during discrete-trial instruction. *Journal of Applied Behavior Analysis, 46*, 379–394. https://doi.org/10.1002/jaba.49.

Cooper, J.O., Heron, T.E., & Heward, W.L. (2007). Applied Behavior Analysis: Second Edition. New York, NY: Pearson.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). Applied Behavior Analysis: Third Edition. New York, NY: Pearson.

Cox, D. J., & Dallery, J. (2018). Verbal behavior and risky choice in humans: Exploring the boundaries of the description-experience gap. *Behavioural Processes, 157*, 301–308. https://doi.org/10.1016/j.beproc.2018.09.002.

Critchfield, T. S. (2011). Translational contributions of the experimental analysis of behavior. *The Behavior Analyst, 34*, 3–17.

Dallery, J., & Soto, P. L. (2013). Quantitative description of environment-behavior relations. In G. J. Madden, W. V. Dube, & K. A. Lattal (Eds.), *APA handbook of behavior analysis, Vol. 1: Methods and principles* (pp. 219–249). Washington, DC: American Psychological Association.

Dart, E. H., & Radley, K. C. (2017). The impact of ordinate scaling on the visual analysis of single-case data. *Journal of School Psychology, 63*, 105–118. https://doi.org/10.1016/j.jsp.2017.03.008.

DeHart, W. B., Friedel, J. E., Frye, C. C. J., Galizio, A., & Odum, A. L. (2018). The effects of outcome unit framing on delay discounting. *Journal of the Experimental Analysis of Behavior, 110*, 412–429. https://doi.org/10.1002/jeab.469.

DeHart, W. B., & Odum, A. L. (2015). The effects of the framing of time on delay discounting. *Journal of the Experimental Analysis of Behavior, 103*, 10–21. https://doi.org/10.1002/jeab.125.

DeProspero, A., & Cohen, S. (1979). Inconsistent visual analyses of intrasubject data. *Journal of Applied Behavior Analysis, 12*, 573–579. https://doi.org/10.1901/jaba.1979.12-573.

Diller, J. W., Barry, R. J., & Gelino, B. W. (2016). Visual analysis of data in a multielement design. *Journal of Applied Behavior Analysis, 49*, 980–985. https://doi.org/10.1002/jaba.325.

Du, W., Green, L., & Myerson, J. (2002). Cross-cultural comparisons of discounting delayed and probabilistic rewards. *Psychological Record, 52*, 479–492.

Estes, W. K. (1976). The cognitive side of probability learning. *Psychological Review, 83*, 37–64. https://doi.org/10.1037/0033-295X.83.1.37.

Fisher, W. W., Kelley, M. E., & Lomas, J. E. (2003). Visual aids and structured criteria for improving visual inspection and interpretation of single-case designs. *Journal of Applied Behavior Analysis, 36*, 387–406. https://doi.org/10.1901/jaba.2003.36-387.

Fryling, M. J., Wallace, M. D., & Yassine, J. N. (2012). Impact of treatment integrity on intervention effectiveness. *Journal of Applied Behavior Analysis, 45*, 449–453. https://doi.org/10.1901/jaba.2012.45-449.

Gallistel, C. R. (2012). Extinction from a rationalist perspective. *Behavioural Processes, 90*(1), 66–80. https://doi.org/10.1016/j.beproc.2012.02.008.

Geiger, K. B., Carr, J. E., & LeBlanc, L. A. (2010). Function-based treatments for escape-maintained problem behavior: A treatment-selection model for practicing behavior analysts. *Behavior Analysis in Practice, 3*(1), 22–32. https://doi.org/10.1007/BF03391755.

Green, L., Myerson, J., & Ostaszewski, P. (1999). Amount of reward has opposite effects on the discounting of delayed and probabilistic outcomes. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 25*, 418–427.

Hayes, S. C. (1989). *Rule-governed behavior: Cognition, contingencies & instructional control* (reprint ed.). Reno, NV: Context Press. https://psycnet.apa.org/record/1989-98580-000.

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. https://doi.org/10.1901/jeab.1961.4-267.

Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior, 13*, 243–266. https://doi.org/10.1901/jeab.1970.13-243.

Hinz, K. L., McGee, H. M., Huitema, B. E., Dickinson, A. M., & Van Enk, R. A. (2014). Observer accuracy and behavior analysis: data collection procedures on hand hygiene compliance in a neurovascular unit. *American Journal of Infection Control, 42*, 1067–1073.

Hursh, S. R., & Silberberg, A. (2008). Economic demand and essential value. *Psychological Review, 115*, 186–198.

Jarmolowicz, D. P., Reed, D. D., Bruce, A. S., & Bruce, J. (2019). On the behavioral economics of medication: A research story. *Behavioural Processes, 165*, 66–77. https://doi.org/10.1016/j.beproc.2019.05.019.

Jarmolowicz, D. P., Reed, D. D., Bruce, A. S., Catley, D., Lynch, S., Goggin, K., … Bruce, J. M. (2016). Using EP50 to forecast treatment adherence in individuals with multiple sclerosis. *Behavioural Processes, 132,* 94–99. https://doi.org/10.1016/j.beproc.2016.09.003

Johnson, M. W., Herrmann, E. S., Sweeney, M. M., LeComte, R. S., & Johnson, P. S. (2017). Cocaine administration dose-dependently increases sexual desire and decreases condom use likelihood: The role of delay and probability discounting in connecting cocaine with HIV. *Psychopharmacology, 234*(4), 599–612. https://doi.org/10.1007/s00213-016-4493-5.

Johnson, P. S., Sweeney, M. M., Herrmann, E. S., & Johnson, M. W. (2016). Alcohol increases delay and probability discounting of condom-protected sex: A novel vector for alcohol-related HIV transmission. *Alcoholism: Clinical & Experimental Research, 40*(6), 1339–1350. https://doi.org/10.1111/acer.13079.

Johnston, J. M., & Pennypacker, H. S. (1993). *Strategies and tactics of behavioral research* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kahng, S. W., Chung, K. M., Gutshall, K., Pitts, S. C., Kao, J., & Girolami, K. (2010). Consistent visual analyses of intrasubject data. *Journal of Applied Behavior Analysis, 43*, 35–45. https://doi.org/10.1901/jaba.2010.43-35.

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). New York, NY: Oxford University Press.

Kipfmiller, K. J., Brodhead, M. T., Wolfe, K., LaLonde, K., Sipila, E. S., Bak, S. M. Y., & Fisher, M. H. (2019). Training front-line employees to conduct visual analysis using a clinical decision-making model. *Journal of Behavioral Education.* https://doi.org/10.1007/s10864-018-09318-1.

Kirby, K. N., & Marakovic, N. N. (1996). Delay-discounting probabilistic rewards: Rates decrease as amounts increase. *Psychonomic Bulletin & Review, 3*, 100–104.

Ledford, J. R., & Gast, D. L. (2018). *Single case research methodology* (3rd ed.). New York, NY: Routledge.

Lerman, D. C., Hovanetz, A., Strobel, M., & Tetreault, A. (2009). Accuracy of teacher-collected descriptive analysis data: a comparison of narrative and structured reading formats. *Journal of Behavioral Education, 18*, 157–172.

Mace, F. C., & Critchfield, T. S. (2010). Translational research in behavior analysis: Historical traditions and imperative for the future. *Journal of the Experimental Analysis of Behavior, 93*, 293–312. https://doi.org/10.1901/jeab.2010.93-293.

Maffei-Almodovar, L., Feliciano, G., Fienup, D. M., & Sturmey, P. (2017). The use of behavioral skills training to teach graph analysis to community based teachers. *Behavior Analysis in Practice, 10*, 355–362. https://doi.org/10.1007/s40617-017-0199-3.

McKerchar, T. L., Green, L., & Myerson, J. (2010). On the scaling interpretation of exponents in hyperboloid models of delay and probability discounting. *Behavioural Processes, 84*, 440–444. https://doi.org/10.1016/j.beproc.2010.01.003.

McKerchar, T. L., & Renda, R. C. (2012). Delay and probability discounting in humans: An overview. *The Psychological Record, 62*, 817–834.

Mudford, O. C., Martin, N. M., Hui, J. K. Y., & Taylor, S. A. (2009). Assessing observer accuracy in continuous recording of rate and duration: three algorithms compared. *Journal of Applied Behavior Analysis, 42*, 527–539. https://doi.org/10.1901/jaba.2009.42-527.

Naude, G. P., Kaplan, B. A., Reed, D. D., Henley, A. J., & DiGennaro Reed, F. D. (2018). Temporal framing and the hidden-zero effect: Rate-dependent outcomes on delay discounting. *Journal of the Experimental Analysis of Behavior, 109*, 506–519. https://doi.org/10.1002/jeab.328.

Ninci, J., Vannest, K. J., Willson, V., & Zhang, N. (2015). Interrater agreement between visual analysts of single-case data: A meta-analysis. *Behavior Modification, 39*, 510–541. https://doi.org/10.1177/0145445515581327.

O'Leary, K. D., Kent, R. N., & Kanowitz, J. (1975). Shaping data collection congruent with experimental hypotheses. *Journal of Applied*

Behavior Analysis, 8, 43–51. https://doi.org/10.1901/jaba.1975.8-43.

Pence, S. T., & St. Peter, C. C. (2015). Evaluation of treatment integrity errors on mand acquisition. *Journal of Applied Behavior Analysis, 48*, 575–589. https://doi.org/10.1002/jaba.238.

Poling, A. (2010). Looking to the future: Will behavior analysis survive and prosper? *The Behavior Analyst, 33*, 7–17. https://doi.org/10.1007/bf03392200.

Qualtrics. (2019). Qualtrics (software; version June, 2019).. Provo, UT: Author. Retrieved from https://www.qualtrics.com

Rachlin, H., Brown, J., & Cross, D. (2000). Discounting in judgments of delay and probability. *Journal of Behavioral Decision Making, 13*, 145–159.

Rachlin, H., & Green, L. (1972). Commitment, choice and self-control. *Journal of the Experimental Analysis of Behavior, 17*, 15–22. https://doi.org/10.1901/jeab.1972.17-15.

Rachlin, H., Raineri, A., & Cross, D. (1991). Subjective probability and delay. *Journal of the Experimental Analysis of Behavior, 55*, 233–244.

Rasmussen, E. B., Lawyer, S. R., & Reilly, W. (2010). Percent body fat is related to delay and probability discounting for food in humans. *Behavioural Processes, 83*, 23–30. https://doi.org/10.1016/j.beproc.2009.09.001.

Repp, A. C., Nieminen, G. S., Olinger, E., & Brusca, R. (1988). Direct observation: Factors affecting the accuracy of observers. *Exceptional Children, 55*, 29–36. https://doi.org/10.1177/001440298805500103.

Rescorla, R. A. (1967). Pavlovian conditioning and its proper control procedures. *Psychological Review, 74*, 71–80. https://doi.org/10.1037/h0024109.

Rescorla, R. A. (1968). Pavlovian conditioning. It's not what you think it is. *American Psychologist, 43*, 151–160. https://doi.org/10.1037/003-066X.43.3.151.

Saini, V., Fisher, W. W., & Retzlaff, B. J. (2018). Predictive validity and efficiency of ongoing visual-inspection criteria for interpreting functional analyses. *Journal of Applied Behavior Analysis, 51*, 303–320. https://doi.org/10.1002/jaba.450.

Seabold, S., & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. *Proceedings of the 9th Python in Science Conference*. Retrieved from https://www.statsmodels.org/stable/index.html.

Shead, N. W., & Hodgins, D. C. (2009). Probability discounting of gains and losses: Implications for risk attitudes and impulsivity. *Journal of the Experimental Analysis of Behavior, 92*, 1–16. https://doi.org/10.1901/jeab.2009.92-1.

Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. New York, NY: Basic Books.

St. Peter, C. C., Byrd, J. D., Pence, S. T., & Foreman, A. P. (2016). Effects of treatment-integrity failures on a response-cost procedure. *Journal of Applied Behavior Analysis, 49*, 308–328. https://doi.org/10.1002/jaba.291.

Staddon, J. E. R., Hinson, J. M., & Kram, R. (1981). Optimal choice. *Journal of the Experimental Analysis of Behavior, 35*, 397–412.

Vanselow, N. R., Thompson, R., & Karsina, A. (2011). Data-based decision making: The impact of data variability, training, and context. *Journal of Applied Behavior Analysis, 44*, 767–780. https://doi.org/10.1901/jaba.2011.44-767.

Ward, R. D., Gallistel, C. R., & Balsam, P. D. (2013). It's the information. *Behavioural Processes, 95*, 3–7. https://doi.org/10.1016/j.beproc.2013.01.005.

Wolfe, K., Barton, E., & Meadan, H. (2019). Systematic protocols for the visual analysis of single-case research data. *Behavior Analysis in Practice, 12*, 491–502. https://doi.org/10.1007/s40617-019-00336-7.

Wolfe, K., & McCammon, M. N. (in press). The analysis of single-case research data: current instructional practices. *Journal of Behavioral Education*. Advance online publication. https://doi.org/10.1007/s10864-020-09403-4.

Wolfe, K., & Slocum, T. A. (2015). A comparison of two approaches to training visual analysis of AB graphs. *Journal of Applied Behavior Analysis, 48*, 472–477. https://doi.org/10.1002/jaba.212.

Wulff, D. U., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin, 144*(2), 140–176. https://doi.org/10.1037/bul0000115.

Yi, R., & Bickel, W. K. (2005). Representation of odds against in terms of frequencies reduces probability discounting. *The Psychological Record, 55*, 577–593.