Journal of the
Experimental Analysis of Behavior

A CROWDSOURCED NICKEL-AND-DIME APPROACH TO ANALOG OBM RESEARCH: A BEHAVIORAL ECONOMIC FRAMEWORK FOR UNDERSTANDING WORKFORCE ATTRITION

Amy J. Henley, Florence D. DiGennaro Reed, Derek D. Reed, and Brent A. Kaplan

UNIVERSITY OF KANSAS

Incentives are a popular method to achieve desired employee performance; however, research on optimal incentive magnitude is lacking. Behavioral economic demand curves model persistence of responding in the face of increasing cost and may be suitable to examine the reinforcing value of incentives on work performance. The present use-inspired basic study integrated an experiential human operant task within a crowdsourcing platform to evaluate the applicability of behavioral economics for quantifying changes in workforce attrition. Participants included 88 Amazon Mechanical Turk Workers who earned either a $0.05 or $0.10 incentive for completing a progressively increasing response requirement. Analyses revealed statistically significant differences in breakpoint between the two groups. Additionally, a novel translation of the Kaplan-Meier survival-curve analyses for use within a demand curve framework allowed for examination of elasticity of workforce attrition. Results indicate greater inelastic attrition in the $0.05 group. We discuss the benefits of a behavioral economic approach to modeling employee behavior, how the metrics obtained from the elasticity of workforce attrition analyses (e.g., P_{max}) may be used to set goals for employee behavior while balancing organizational costs, and how economy type may have influenced observed outcomes.

Key words: performance management, demand, behavioral economics, incentives, Amazon Mechanical Turk, crowdsourcing, human

---

Organizational behavior management (OBM) is a subdiscipline of applied behavior analysis that utilizes the principles of behavior to effect change with people and groups in business (Wilder, Austin, & Casella, 2009). Incentives are a popular method to improve employee behavior; as such, incentives have been the focus of much research. Incentives are rewards individuals earn upon reaching a performance criterion and include money, time off from work, and other tangible or intangible items (e.g., Henley, DiGennaro Reed, Kaplan, & Reed, 2016; Luiselli et al., 2009). Both laboratory (e.g., Slowiak, Dickinson, & Huitema, 2011) and applied (e.g., Miller, Carlson, & Sigurdsson, 2014) studies have reliably shown improvements in performance when participants earn incentives than when they receive hourly pay. In some cases, performance improved between 15% and 300% and net profits totaled between $56,000 and $400,000 a year (Bucklin & Dickinson, 2001). However, the most effective incentive magnitude is unclear, which makes it challenging for employers to extrapolate findings to their particular settings. Smaller incentives may be insufficient to change behavior, but larger incentives may not be sustainable for the organization.

Unfortunately, there is a paucity of research involving parametric analyses of incentive magnitude over and above hourly or base pay and research on *related* aspects of incentives has produced inconsistent results (see Bucklin & Dickinson, 2001 for a review). For example, previous studies have failed to find differences in performance when examining the percentage of incentive pay as a proportion of base pay. This finding has led researchers to conclude the *availability* of an incentive is more important than the percentage of incentive pay relative to base pay (e.g., Dickinson & Gillette, 1994; Matthews & Dickinson, 2000). However, other studies have shown higher percentages of incentives yielding higher performance levels when

Amy J. Henley, Department of Applied Behavioral Science, University of Kansas; Florence D. DiGennaro Reed, Department of Applied Behavioral Science, University of Kansas; Derek D. Reed, Department of Applied Behavioral Science, University of Kansas; Brent A. Kaplan, Department of Applied Behavioral Science, University of Kansas.

Correspondence concerning this article should be addressed to Florence D. DiGennaro Reed, Department of Applied Behavioral Science, University of Kansas, 4001 Dole Human Development Center, 1000 Sunnyside Avenue, Lawrence, KS 66045-7555. E-mail: fdreed@ku.edu.

doi: 10.1002/jeab.220

researchers increased the ecological validity of the simulated workplace (Oah & Lee, 2011). Conflicting findings across studies underscore the need for additional research on the most effective incentive arrangement. Potentially, by drawing from other research areas, OBM researchers may identify alternative methods to address these issues.

Operant behavioral economics—hereafter referred to in short form as *behavioral economics*—may be a useful framework for evaluating employee behavior and, in particular, the effects of incentive systems (Jarmolowicz, Reed, DiGennaro Reed, & Bickel, 2016; Roma, Hursh, & Hudja, 2016; Sigurdsson, Taylor, & Wirth, 2013; Wine, Gilroy, & Hantula, 2012). Behavioral economics integrates principles of behavioral science and microeconomics to understand decision-making and behavior (Hursh, 1980). Demand curve analyses, a prominent analytic technique in behavioral economics, quantify price sensitivity in reinforcer consumption as a function of increasing cost. Several relevant characteristics of demand curves are worth noting. The inelastic portion of the curve is characterized by a less than one-unit decrement in consumption for each unit price increment, indicating that consumption remains relatively stable as price increases. The elastic portion of the curve is characterized by a greater than one-unit change in consumption with a one-unit increase in price (i.e., consumption substantially decreases as price increases). The cost associated with unit elasticity (i.e., where the curve shifts from inelastic to elastic) is termed $P_{\max}$. The contemporary model of demand developed by Hursh and Silberberg (2008) derives α (alpha) to quantify the rate of change in elasticity across the entire range of prices. Smaller α values indicate relative inelasticity; that is, relatively less price-sensitive responding.

When applied to an organization, demand curve analyses quantify the persistence of work behavior at increasing requirements to determine the value of a reinforcer—put simply, the analysis quantifies *motivation* to complete work tasks across increasing work requirements. In one of the first studies to apply demand curve analyses to employee behavior, Henley et al. (2016) used a hypothetical task (distributing flyers on campus) to evaluate within-subject demand for monetary compensation delivered at two delays with undergraduate participants. Participants completing this task reported the likelihood they would pass out a given number of flyers across an increasing work requirement. Participants were provided instructions indicating they would receive payment immediately after working (i.e., short delay) or after a 4-week delay. Henley et al. found higher demand in the short delay condition for all behavioral economic indices related to demand elasticity, which was statistically significantly different from demand when payments were delayed by 4 weeks. The differences in demand were equal to a 45% increase in the amount of work completed in the short delay condition.

Henley et al. (2016) demonstrated that demand for workplace incentives may be sensitive to hypothetical contextual variables (in this case, delay of payment), replicating previous research showing the strength of reinforcement is modulated by the delay to reinforcement (e.g., Grace, Schwendiman, & Nevin, 1998; Hursh & Fantino, 1973; Woolverton & Anderson, 2006). Moreover, the demand curves were visually similar to previous research and the data were well accounted for by the exponential model of demand (Hursh & Silberberg, 2008). These preliminary findings suggest demand curve analyses may reveal unique and important workplace contingency information for use in an incentive system to address employee behavior.

Although these results are promising, participants did not experience the outcomes described in Henley et al. (2016) and the researchers did not measure the hypothetical work behavior—passing out flyers. Instead, participants reported an *estimation* of how likely they would be to complete each work requirement. The study also has limited external validity because the participants were not paid employees. Identifying an analog preparation to address these limitations using an alternative method to better simulate organizational settings is particularly important given the challenges experienced by researchers attempting to conduct incentive research in organizations.[1]

[1]OBM researchers commonly use simulated work environments due to constraints faced by organizations (e.g., financial pressures that conflict with conducting research, restrictions posed by union contracts). However, undergraduate participants may not be representative of paid employees and analog settings lack real-world organizational contingencies despite researchers' best efforts to approximate applied settings (Johnson, Casella, McGee, & Lee, 2014).

Amazon Mechanical Turk (mTurk; www.mturk.com) is an online crowdsourcing service where requesters from public, private, and research sectors post tasks for users—appropriately termed *Workers* within the mTurk community—to complete in exchange for money. Accordingly, mTurk may be a novel solution for OBM researchers because it functions as a workplace. At any given point, there are over 500,000 Workers in the mTurk labor market with demographic variability more diverse than university-based samples (Paolacci & Chandler, 2014). The tasks are typically brief computer-mediated jobs referred to as "human intelligence tasks" (HITs) (Buhrmester, Kwang, & Gosling, 2011). When used for research purposes, researchers serve as requesters and post a HIT with their study to the online mTurk platform on which Workers may opt to complete the HIT (for a review of how mTurk may be used in behavioral research, see Mason & Suri, 2012, as well as Paolacci, Chandler, & Ipeirotis, 2010).

Given the aforementioned advantages to crowdsourcing research participation, mTurk is becoming an increasingly popular site for social and behavioral research because it provides access to thousands of Workers around the world resulting in diverse participant samples that can be collected relatively inexpensively (for pennies, rather than dollars) and in a short period of time (Bohannon, 2011; Ipeirotis, 2010). Recent research in the fields of psychology and behavior analysis has demonstrated the reliability of mTurk as a research platform (e.g., Buhrmester et al., 2011; Johnson, Herrmann, & Johnson, 2015). Most pertinent to OBM researchers, however, is the finding that Workers are sensitive to HIT pay structures (Horton & Chilton, 2010), rendering this population even more appealing to behavioral economic evaluations of workplace requirements and compensation arrangements.

The present study had two aims. First, we sought to expand the current body of literature integrating behavioral economics and OBM and to provide further evidence of the utility of behavioral economic analyses for understanding organizational behavior. Our second aim was to use behavioral economic analyses to evaluate the effects of incentive magnitude on behavior of mTurk Workers. To accomplish these aims, we employed a novel procedure for conducting human operant research in which we integrated an experiential demand preparation within an online survey format. After completing a brief survey for a flat payment amount, participants had the opportunity to complete a work task for a fixed incentive amount. Over successive trials, the response requirement to earn the incentive increased. Half of the participants earned a small incentive per trial completed, and the other half earned a larger incentive. Our analyses included both traditional (progressive ratio [PR] breakpoint) and contemporary (exponential demand curve) behavioral economic approaches. For the exponential demand curve analyses, we employed a novel approach to understanding elasticity wherein we translated survival curve logic and generated survival curves of workforce attrition. This analysis subsequently informed an interpretation of "elasticity of workforce attrition" within a demand framework.

Method

Instrumentation and Participants

The survey was created using Qualtrics® Research Suite (see http://www.qualtrics.com). We posted two HITs on the mTurk website containing a survey link. Each HIT looked identical to Workers; however, a different survey link was used for each HIT. The two surveys were identical except for the incentive magnitude participants would have the opportunity to earn (described in more detail below). To access the HIT we required mTurk Workers to meet the following three criteria: (1) reside in the United States, (2) have completed at least 1,000 approved HITs, and (3) have a 95% or above approval rate (participant criteria adopted from Roma et al., 2016). After participants selected the study link, the survey presented an information statement. The Human Subjects Committee at the University of Kansas approved all study procedures. At the end of the survey, participants received a unique access code displayed on the screen. Workers who entered this code into the HIT on mTurk within 3 hr as specified by the HIT instructions received a $0.15 payment and served as participants. One hundred and ninety-four mTurk Workers submitted an access code to the HIT. On average, Workers submitted the access code within 20 min of completing the study.

We omitted results from the second attempt for participants who completed the study for both incentive magnitudes (n = 6). Based on the HIT selected, participants completed one of two surveys that differed in incentive magnitude; we subsequently matched pairs of participants between the two groups on gender, income bracket, and education level. Creating matched pairs of participants yielded 88 participants (n = 44 per group). Sixty-one percent of the sample was male, with a median age of 35 (range, 20 to 67 yrs). A large majority of participants were Caucasian and had at least some college coursework (or more). Although approximately 30% of the sample earned less than $30,000 annually, participants reported a wide range of incomes up to greater than $100,000. Table 1 summarizes the demographic variables in more detail.

Table 1

Participant demographic characteristics.

Demographic	$0.05 ($n$ = 44)	$0.10 ($n$ = 44)
Age (years)		
Range	20 - 67	21 - 63
Mean	37.80	37.43
Gender		
Female	17	17
Male	27	27
Ethnicity		
Asian	3	3
Black/African American	2	2
Hispanic/Latino	0	1
Native American	1	0
White/Caucasian	38	38
Education		
Less than high school	1	1
High school/GED	7	7
Some college	9	9
2-year college degree	2	4
4-year college degree	17	14
Master's degree	7	6
Doctoral degree	0	1
Professional degree (JD, MD)	1	2
Income		
Less than 30,000	13	13
30,000 – 39,999	6	6
40,000 – 49,999	5	5
50,000 – 59,999	6	6
60,000 – 69,999	3	3
70,000 – 79,999	5	5
80,000 – 89,999	3	3
90,000 – 99,999	1	1
100,000 or more	2	2

Flat Fee Survey

Participants earned a base pay of $0.15 for completing a survey and a practice trial, which required an average of 5.33 min to complete. The survey included demographic items such as age, gender, state of residence, education level, income bracket, and ethnicity. The survey also included 51 items assessing choice for the purposes of another study. The practice trial described how to complete the simulated work task and provided one practice opportunity. The practice trial was included to ensure all participants had experience with the work task before having an opportunity to earn incentives in exchange for work.

Incentivized Work Task

The work task contained slider questions that displayed a random number between -100 and 100 presented on the left end of the slider scale (depicted in Fig. 1). The slider bar default start position was set to zero (center of scale) for all questions; as a result, we excluded target values of zero. Using their mouse, participants moved a bar (from the start point at zero) along the scale until the number on the right side of the slider matched the target number on the left. Participants could also move the bar using a touch screen on a mobile device. If participants provided an incorrect response (e.g., moved the slider to 30 when the target number was 41) the survey displayed an error message reading, "Incorrect. Please try again." The work task included a progressive ratio schedule with 13 ratio requirements (1, 2, 4, 8, 12, 16, 25, 32, 48, 64, 96, 128, and 256). The completion of each requirement resulted in a fixed monetary incentive paid at the end of the study.

At the beginning of each ratio requirement, the computer program informed participants of the incentive amount and the number of slider questions required to earn the incentive. Participants then responded to the question,

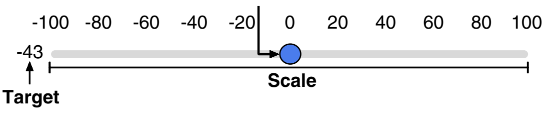

Fig. 1. Schematic drawing of the simulated work task (i.e., slider question). As participants moved the slider along the scale, a number representing the value at which the slider resides was displayed on the right end of the scale, opposite the target value.

"Would you like to continue?" If a participant selected *yes*, the program presented the corresponding block of slider questions. If a participant selected *no*, the survey ended. To ensure the only effort manipulation involved the increasing ratio requirements, we adjusted the target values for each ratio requirement so they equaled zero when summed. The values used in the practice trial and the first PR requirement summed to a value of zero when combined. Half of the participants ($n = 44$) had the opportunity to earn \$0.05 per ratio requirement completed, and the other half had the opportunity to earn \$0.10 for each ratio requirement completed.

Data Analysis

For all analyses, we used unit price as compared to the ratio value to compare the two incentive magnitudes. Unit price is a cost-benefit ratio calculated as the ratio requirement divided by the incentive amount, in this case either \$0.05 or \$0.10. We used unit price values instead of PR values to equate the benefit component of the ratio and establish a common price (i.e., number of responses required to earn \$0.01) given the differences in incentive size between the two groups (Hursh, Raslear, Bauman, & Black, 1989).

Breakpoint. Breakpoint was defined as the highest unit price completed by the participant. We used a Mann-Whitney *U*-test to compare differences in breakpoint between the two groups. A nonparametric test was used due to non-normally-distributed data as indicated by the D'Agostino and Pearson omnibus normality test.

Workforce attrition. Next, we sought to evaluate differences in elasticity between the two incentive magnitudes. Because our data were dichotomous (i.e., participants did or did not complete each ratio requirement) and participants experienced the ratio requirements only once, we could not generate individual demand curves. These elements prevent the quantification of the number of reinforcers consumed over repeated exposures to each ratio, a feature that many researchers consider a distinguishing aspect of demand. Greenwald and Hursh (2006) provide a method for transforming dichotomous responses into a group level demand curve (referred to as market percent consumption); however, a method for constructing a demand curve with dichotomous responses in which the organism experienced the ratio requirements just once has not yet been developed. In light of the differences noted, any such method might be best conceptualized as an analog to demand.

Dichotomous choices without repetition are comparable to data used in survival curve analyses. Survival curves have been successfully used in business as a method for quantifying workforce attrition (e.g., Lane & Andrew, 1955; Moncrief, Hoverstad, & Lucas, 1989) and may be applicable to the current data. Thus, we sought to translate survival curve logic to interpret measures of elasticity of workforce attrition in which attrition is defined as the ratio requirement at which individual Workers discontinued responding and exited the HIT. Within this framework, attrition occurring at higher ratios indicates a greater quantity of work completed and, therefore, reflects a measure of Worker performance.

In our translation, workforce survival (i.e., the percentage of workforce remaining; Q) at each PR step along the x-axis was calculated according to an adaptation of the Kaplan-Meier estimate of survival (Kaplan & Meier, 1958), such that:

$$Q = \hat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right) \tag{1}$$

where the survival function $\left(\hat{S}(t)\right)$ is described by the number of participants who quit the HIT (d) over the duration (t) of the study up to point i in the observations (unit price), with n_i representing the number of participants continuing to respond just prior to t_i. Prototypical Kaplan-Meier estimates of survival are interpreted as the product of conditional probabilities from Equation 1. In our adaptation, Q is interpreted both as the percentage of the workforce remaining in the PR progression and the conditional probability of workforce survival.[2] We used a log-rank test to compare whether the survival curves were

[2]Note that in typical survival curve analyses, some participants result in censored observations (e.g., dropout or attrition not due to the event of interest), but in the current study all participants were followed through to the end of their highest ratio requirement completed. Thus, workforce survival at each ratio requirement is the number of participants completing the given ratio requirement divided by 44 (the total number of participants present in each group at the onset of the study).

statistically significantly different from each other assuming the null hypothesis was true (i.e., no difference in survival between the two groups).

For the elasticity of workforce attrition analyses, a market percent consumption approach was adopted from Greenwald and Hursh (2006) to convert individual dichotomous choices at each ratio requirement into a group-level metric (i.e., the number of Workers finishing the given ratio requirement) that permits an analysis of elasticity of workforce attrition via survival curve logic (i.e., the percentage of the workforce remaining following attrition due to increases in ratio requirements). To accomplish this, we calculated the percentage of participants remaining following successive ratio requirements (i.e., workforce survival).

Unit price values were subsequently used in fitting the exponential model of demand to derive elasticity of workforce attrition (instead of consumption) curves and associated metrics (Hursh & Silberberg, 2008):

$$\log Q = \log Q_0 + k\left(e^{-\alpha(Q_0 \cdot C)} - 1\right) \tag{2}$$

where Q is equal to workforce survival, Q_0 equals the maximum percentage of workforce surviving when the requirement is zero or near-zero, in this case one slider question, k is the scaling constant equal to the number of log units of workforce survival, C equals cost, or the number of slider questions required to earn \$0.01 (i.e., unit price), and α is a fitted parameter quantifying the rate of change in elasticity, or how the dependent variable (in this case, workforce survival; typically, consumption of reinforcers) changes as a function of cost, across the entire range of ratio requirements. When fitting Equation 2 in the present study, we used a shared constant value of 2 for k, set Q_0 to maximum possible consumption (100 for both groups), and left α unconstrained.

The elasticity of workforce attrition analysis offers added analytic utility beyond PR breakpoint analyses and survival curves in its ability to estimate essential value (*EV*) of the incentive, as well as $P_{\max}$ (see below). Moreover, nonlinear regression using Equation 2 provides an estimate of Q across the entire range of possible unit prices imposed on the workforce as compared to only those values tested.

An extra sum-of-squares *F*-test was used to evaluate differences in the rate of change of elasticity of workforce attrition (i.e., α) across groups. The extra sum-of-squares *F*-test in Graphpad Prism 7.00 compares the fits of each group curve with the fit of a curve to a single aggregated dataset. In our case, this test determines whether a single curve describes both group datasets, or whether the curves of the two groups differ significantly warranting a unique curve for each group.

Two additional indices derived from Equation 2, *EV* and $P_{\max}$, were calculated (see Hursh, 2014, for a thorough review of these indices). Essential Value was calculated as:

$$EV = \frac{1}{100 \cdot \alpha \cdot k^{1.5}} \tag{3}$$

which quantifies the reinforcing value of the incentive using α and k from Equation 2. Similarly, point of unit elasticity, $P_{\max}$, uses α and k from Equation 2 to describe the price at which the curve transitions from inelastic to elastic, such that:

$$P_{\max} = \frac{m}{(Q_0 \cdot \alpha \cdot k^{1.5})}, \text{ where } m = 0.083k + 0.65. \tag{4}$$

The Exponential Model of Demand GraphPad Prism Template (www.ibrinc.org) was used with GraphPad Prism version 7.00 for Windows (GraphPad Software, La Jolla, California, USA; www.graphpad.com) to fit the demand equation to the data. A freely available Microsoft Excel-based tool (Kaplan & Reed, 2014) was used for calculating *EV* and $P_{\max}$.

Results

Figure 2 depicts the unit price breakpoint for each participant by incentive magnitude. When considering the number of slider questions completed per \$0.01 (i.e., unit price), breakpoint was lower for the \$0.10 group ($M = 2.22$, $SD = 2.43$) with a median unit price of 1.6 as compared to the \$0.05 group ($M = 3.48$, $SD = 3.39$) with a median of 2.4 slider questions per \$0.01. This difference was statistically significant ($U = 624$, $p = .0035$).

Figure 3 depicts the survival curve of workforce attrition. As the ratio requirement—and hence unit price—increased, the percentage of participants who discontinued responding

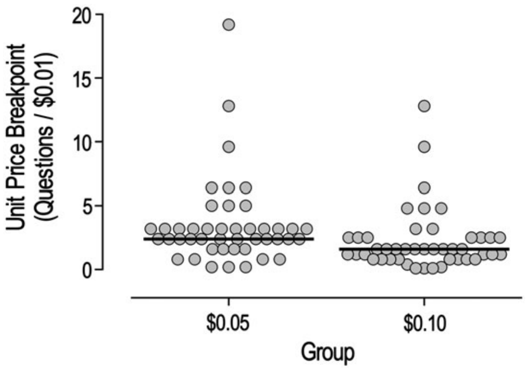

Fig. 2. Individual unit price breakpoints in the $0.05 and $0.10 groups. A ratio value of zero on the y-axis indicates participants did not complete any work tasks to earn incentives. Horizontal lines represent group medians.

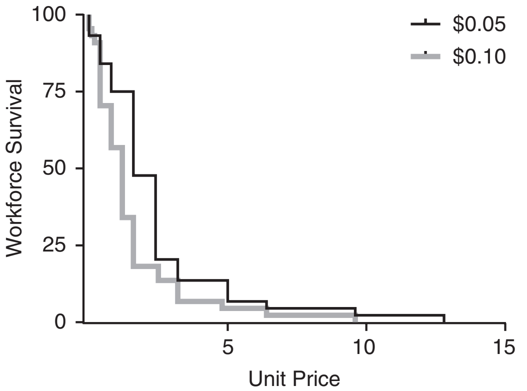

Fig. 3. Survival curve of workforce attrition for the $0.05 and $0.10 groups. The y-axis depicts the percentage of participants in each group who continued responding to the work task across increasing unit price values (x-axis).

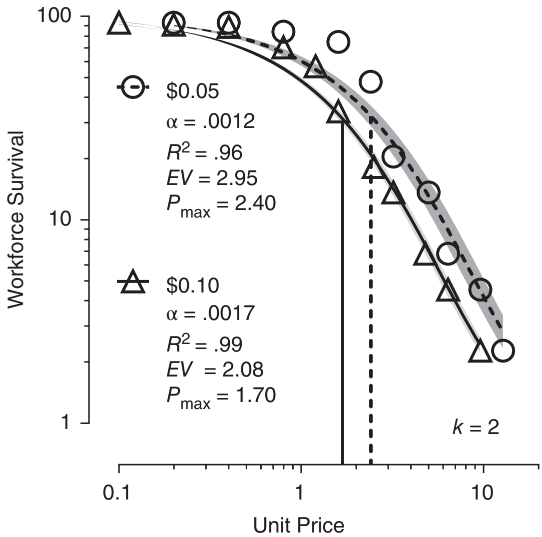

Fig. 4. Elasticity of workforce attrition analyses for the $0.05 and $0.10 groups with the 95% confidence intervals for α shaded in gray. The y-axis depicts the percentage of participants in each group who completed the ratio requirement. The x-axis depicts unit price. Vertical lines correspond to $P_{\max}$.

occurred at a higher rate in the $0.10 group. Results from the log-rank test indicated a statistically significant difference between the two groups, χ^2 (1) = 6.68, p = .0098.

Figure 4 depicts the elasticity of workforce attrition analyses for both groups using survival curve logic. Overall, workforce survival decreased systematically as a function of the ratio requirement. The data were well accounted for by the exponential model of demand with R^2 values of .96 for the $0.05 group and .99 for the $0.10 group. Elasticity of workforce attrition in the $0.05 group was more inelastic (α = .0012) than the $0.10 group ($\alpha$ = .0017). A sum-of-squares F-test indicated this difference was statistically significant (F [1, 19] = 24, p < .0001). Essential Value was 2.95 in the $0.05 group and 2.08 in the $0.10 group. $P_{\max}$, expressed as the number of slider questions per $0.01, was 2.40 and 1.70 for the $0.05 and $0.10 groups, respectively.

Discussion

The first aim of this study was to evaluate the applicability of a behavioral economic framework to employee behavior and extend the literature on behavioral economics and OBM. The second aim of the study was to evaluate the effects of incentive magnitude on Worker persistence using a novel human operant procedure within a crowdsourcing platform. To achieve these aims, we employed mTurk Workers to complete a simple task—in this case, moving a slider along a scale to match a number—similar to the kinds of tasks commonly used in mTurk HITs. Worker behavior was analyzed using the traditional (PR breakpoint) approach and the contemporary (exponential demand curve) behavioral economic approach, wherein we translated survival curve logic to interpret measures of elasticity of workforce attrition.

With respect to the first aim, the data returned from the mTurk workforce followed systematic patterns of responding expected from PR contingencies (Jarmolowicz & Lattal, 2010). That is, the number of Workers persisting decreased systematically as a function of the ratio requirement. The fits of the elasticity of workforce attrition curves were also consistent with other controlled experiments using actual contingencies (Greenwald & Hursh, 2006) and were well accounted for by the exponential equation with R^2 values of .96 and .99. Consistent findings across all three measures (breakpoint, survival, and elasticity) suggest a behavioral economic approach may be a viable framework to evaluate employee behavior.

Although we obtained similar findings across the three analyses, the quantitative metrics derived from the exponential equation may offer unique benefits in the workplace, possibly representing an improvement over prototypical PR breakpoint and survival curve analyses. For example, from an organizational perspective P_{max} estimates the upper boundary of work requirement inelasticity; work expectations beyond P_{max} would thereby generate elastic attrition wherein attrition significantly increases (i.e., the workforce no longer persists or "survives"). Calculating P_{max} could allow an organization to balance costs of an incentive program and simultaneously set goals for desired employee behavior that enhances the effectiveness of the incentive. Additionally, behavioral economists have critiqued PR breakpoint because it is a discontinuous measure and is dependent on the PR step sizes (see Hursh & Silberberg, 2008; also see Stafford & Branch, 1998), and thereby recommend the measure of elasticity to model consumption across the entire range of potential constraints via nonlinear regression. Such an approach may be advantageous to OBM researchers when workplace analyses limit available research resources and/or subject availability for experimental sessions. This addresses a noted criticism of the use of PR schedules in applied settings (e.g., Roane, 2008) and allows researchers to derive information about performance at every possible price as compared to only those tested.

Regarding the second aim, the median unit price breakpoint for the \$0.05 group was 2.4 slider questions per \$0.01, which was statistically significantly higher than 1.6 for the \$0.10 group. Relatedly, the elasticity of workforce attrition analyses indicated the use of a larger incentive resulted in a statistically significantly greater change in elasticity at the aggregate level. That is, Workers in the \$0.10 group more readily quit the task in the face of progressively increasing ratio requirements. This finding is similar to that found by Cassidy and Dallery (2012) who found rats' demand was greater (i.e., relatively less elastic) for one food pellet as compared to two food pellets when an open economy was used. Although we recognize a vast number of differences exist between their study and the current study, both are similar in terms of the presence of an open economy. The mTurk workplace is considered an open economy because Workers are able to earn money (presumably a reinforcer) from a variety of sources outside of a given HIT (akin to a laboratory preparation) or even outside the mTurk workplace itself; and they may choose to do so at any given time. It may be the case that the openness of mTurk's economy resulted in our similar findings to Cassidy and Dallery. Future research could evaluate the role of economy type on measures of elasticity to more accurately inform organizations' incentive structures.

Results of the current study, however, stand in contrast to previous research indicating α should be similar despite variations in magnitude of the commodity (Hursh & Silberberg, 2008). This difference may be due to the commodity under investigation or because the dependent variable was workforce survival, rather than the total number of reinforcers consumed at each price. Equivalent α values are theoretically expected across different magnitudes because the exponent of the exponential demand equation standardizes price as $Q_0 \times C$. However, demand curve analyses using a percentage or proportion obviate differences in the true cost necessary to defend baseline levels of demand; regardless of reinforcer magnitude, Q_0 is constrained to 100 or 1 for a percentage and proportion, respectively. Although the constraint on Q_0 is a potential limitation of using percentages and proportions, as demand investigations are adapted to answer novel questions with diverse commodities and behaviors of interest, this method may have merit and make conceptual sense. Traditionally, research has evaluated demand for consumable commodities (e.g., pellets of food, alcoholic beverages) wherein, despite individual differences, the quantity of a reinforcer consumed when free (i.e., Q_0) is reached as a function of biological constraints (i.e., satiation). Even when the commodity is—in the traditional or lay sense—a nonconsumable good such as gasoline, practical constraints may play a role in moderating Q_0 (e.g., availability of storage mechanisms or space and expiration). Conversely, it is difficult to identify a factor that may constrain consumption at a price of zero when the commodity of interest is money. Percentage and proportion are a relatively recent adaptation of consumption within demand analyses and are an important area of investigation for future research.

In terms of advancing the behavioral economic and OBM literature, the current, use-inspired basic study using reverse translation is one of the first of its kind to evaluate organizational concerns (e.g., delivery of incentives) using a behavioral economic framework and, to our knowledge, the first to integrate a human operant task (whereby participants engaged in the actual task and experienced real outcomes) within a crowdsourcing platform. Relatedly, the demographics obtained via the mTurk workforce were indeed more diverse than typical undergraduate convenience samples (cf. Paolacci & Chandler, 2014), possibly providing a better approximation of a typical American workforce than the usual undergraduate participants in a human operant laboratory and greater ecological validity. Collectively, these attributes of the study render mTurk a viable platform for OBM research that directly assesses actual workers engaged in actual work-for-hire (i.e., high ecological validity).

The current study also translated the Kaplan-Meier estimate of survival used in prototypical survival curve analyses to a behavioral economic analysis of elasticity of workforce attrition, which represents a novel contribution in its own right. That the aggregate data were well accounted for by Equation 2 not only provides further validation of the behavioral economic model of demand, but also suggests such analyses may be applicable for assessing reinforcer value and workplace attrition in organizations. The current methodology may allow researchers in OBM to better refine interventions prior to implementation, especially for situations in which workforce attrition is a major issue. Finally, the elasticity of workforce attrition analyses in this study are, to our knowledge, a novel metric for organizational assessment. Metrics such as α, EV, and $P_{\max}$ may prove useful in identifying the most efficacious incentives among a range of alternatives (i.e., ones that yield the largest EV and $P_{\max}$) and to determine incentive amounts that maximize an organization's return on investment.

Despite these strengths, several limitations are worth discussing. Although the payment and incentive amounts used in the current study are typical for mTurk, they are not representative of those used in traditional work environments, thereby limiting the degree to which the current findings may generalize to other work settings. The task—moving a slider bar—was relatively simple and researchers may obtain dissimilar findings with complex tasks. Additionally, Workers might have responded to the amount of time allotted to complete the task,[3] rather than the PR schedule, or Workers may have maximized earnings by quitting the PR task to begin another HIT. This arrangement is not typical of most work settings where employees are required to complete a task to earn pay-for-performance or remain present for an entire workday. It is also possible that the participant sample does not represent the general population in terms of sensitivity to changes in cost/benefit ratios. Potentially, mTurk Workers' histories of selecting and completing HITs that maximize return on their brief efforts may increase their sensitivity to changes in price, which could influence the accuracy of the demand curve model fits and/or the generalizability of the present findings. Finally, there are many work tasks in which the primary outcome or benchmark is not quantity (e.g., accuracy or quality), which was used as the dependent variable in the current study. The degree to which incentive amounts affect these other types of work tasks is unknown, but a benefit of the current experimental preparation allows for

[3]Before selecting a HIT, Workers are able to see several details regarding the HIT such as the requester's name, expiration time of and time allotted to complete the HIT, payment amount, and the number of HITs available.

examination of effects on other dimensions of work (e.g., accuracy).

References

Bohannon, J. (2011). Social science for pennies. *Science, 334* (6054), 307. doi: 10.1126/science.334.6054.307

Bucklin, B. R., & Dickinson, A. M. (2001). Individual monetary incentives. *Journal of Organizational Behavior Management, 21*(3), 45–137. doi: 10.1300/J075v21n03_03

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3–5. doi: 10.1177/1745691610393980

Cassidy, R. N., & Dallery, J. (2012). Effects of economy type and nicotine on the essential value of food in rats. *Journal of the Experimental Analysis of Behavior, 97*(2), 183–202. doi: 10.1901/jeab.2012.97-183

Dickinson, A. M., & Gillette, K. L. (1994). A comparison of the effects of two individual monetary incentive systems on productivity: Piece rate pay versus base pay plus incentives. *Journal of Organizational Behavior Management, 14*(1), 3–82. doi: 10.1300/J075v14n01_02

Grace, R. C., Schwendiman, J. W., & Nevin, J. A. (1998). Effects of unsignaled delay of reinforcement on preference and resistance to change. *Journal of the Experimental Analysis of Behavior, 69*(3), 247–261. doi: 10.1901/jeab.1998.69-247

Greenwald, M. K., & Hursh, S. R. (2006). Behavioral economic analysis of opioid consumption in heroin-dependent individuals: Effects of unit price and pre-session drug supply. *Drug and Alcohol Dependence, 85*(1), 35–48. doi: 10.1016/j.drugalcdep.2006.03.007

Henley, A. J., DiGennaro Reed, F. D., Kaplan, B. A., & Reed, D. D. (2016). Quantifying efficacy of workplace reinforcers: An application of behavioral economic demand to evaluate hypothetical work performance. *Translational Issues in Psychological Science, 2*(2), 174–183. doi:10.1037/tps0000068

Horton, J. J., & Chilton, L. B. (2010, June). The labor economics of paid crowdsourcing. In *Proceedings of the 11th ACM Conference on Electronic Commerce* (pp. 209–218). ACM. doi: 10.1145/1807342.1807376

Hursh, S. R. (1980). Economic concepts for the analysis of behavior. *Journal of the Experimental Analysis of Behavior, 34*, 219–238. doi: 10.1901/jeab.1980.34-219

Hursh, S. R. (2014). Behavioral economics and the analysis of consumption and choice. In F. K. McSweeney & E. S. Murphy (Eds.), *The Wiley Blackwell handbook of operant and classical conditioning* (pp. 275–305) Hoboken, NJ: John Wiley & Sons.

Hursh, S. R., & Fantino, E. (1973). Relative delay of reinforcement and choice. *Journal of the Experimental Analysis of Behavior, 19*(3), 437–450. doi: 10.1901/jeab.1973.19-437

Hursh, S. R., Raslear, T. G., Bauman, R., & Black, H. (1989). The quantitative analysis of economic behavior with laboratory animals. In K.G. Grunert, K. Günter, & F. Ölander (Eds.), *Understanding economic behaviour* (pp. 393–407). SpringerNetherlands. doi: 10.1007/978-94-009-2470-3_22

Hursh, S. R., & Silberberg, A. (2008). Economic demand and essential value. *Psychological Review, 115*(1), 186–198. doi: 10.1037/0033-295X.115.1.186

Ipeirotis, P. (2010). Demographics of Mechanical Turk. (CeDER Working Paper-10-01). New York University. Retrieved from: http://hdl.handle.net/2451/29585.

Jarmolowicz, D. P., & Lattal, K. A. (2010). On distinguishing progressively increasing response requirements for reinforcement. *The Behavior Analyst, 33*(1), 119–125.

Jarmolowicz, D. P., Reed, D. D., DiGennaro Reed, F. D., & Bickel, W. K. (2016). The behavioral and neuroeconomics of reinforcer pathologies: Implications for managerial and health decision making. *Managerial and Decision Economics, 37,* 274–293. doi: 10.1002/mde/2716

Johnson, D. A., Casella, S. E., McGee, H., & Lee, S. C. (2014). The use and validation of preintervention diagnostic tools in organizational behavior management. *Journal of Organizational Behavior Management, 34*(2), 104–121. doi: 10.1080/01608061.2014.914009

Johnson, P. S., Herrmann, E. S., & Johnson, M. W. (2015). Opportunity costs of reward delays and the discounting of hypothetical money and cigarettes. *Journal of the Experimental Analysis of Behavior, 103*(1), 87–107. doi: 10.1002/jeab.110

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation for incomplete observations. *Journal of the American Statistical Association, 53,* 457–481. doi: 10.1080/01621459.1958.10501452

Kaplan, B. A., & Reed, D. D. (2014). Essential value, $P_{\max}$, and $O_{\max}$ automated calculator [spreadsheet application]. Retrieved from: http://hdl.handle.net/1808/14934.

Lane, K. F., & Andrew, J. E. (1955). A method of labour turnover analysis. *Journal of the Royal Statistical Society. Series A (General), 118*(3), 296–323. doi: 10.2307/2342495

Luiselli, J. K., DiGennaro Reed, F. D., Christian, W. P., Markowski, A., Rue, H. C., St. Amand, C., & Ryan, C. J. (2009). Effects of an informational brochure, lottery-based financial incentive, and public posting on absenteeism of direct-care human services employees. *Behavior Modification, 33*(2), 175–181. doi: 10.1177/0145445508320624

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods, 44*(1), 1–23. doi: 10.3758/s13428-011-0124-6

Matthews, G. A., & Dickinson, A. M. (2000). Effects of alternative activities on time allocated to task performance under different percentages of incentive pay. *Journal of Organizational Behavior Management, 20*(1), 3–27. doi: 10.1300/J075v20n01_02

Miller, M. V., Carlson, J., & Sigurdsson, S. O. (2014). Improving treatment integrity in a human service setting using lottery-based incentives. *Journal of Organizational Behavior Management, 34*(1), 29–38. doi: 10.1080/01608061.2013.873381

Moncrief III, W. C., Hoverstad, R., & Lucas Jr, G. H. (1989). Survival analysis: A new approach to analyzing sales force retention. *Journal of Personal Selling & Sales Management, 9*(2), 19–30. doi: 10.1080/08853134.1989.10754521

Oah, S., & Lee, J. H. (2011). Effects of hourly, low-incentive, and high-incentive pay on simulated work productivity: Initial findings with a new laboratory method. *Journal of Organizational Behavior Management, 31*(1), 21–42. doi: 10.1080/01608061.2011.541820

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science, 23*(3), 184–188. doi: 10.1177/0963721414531598

Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5*(5), 411–419.

Roane, H. S. (2008). On the applied use of progressive-ratio schedules of reinforcement. *Journal of Applied Behavior Analysis, 41,* 155–161. doi: 10.1901/jaba.2008.41-155

Roma, P. G., Hursh, S. R., & Hudja, S. (2016). Hypothetical purchase task questionnaires for behavioral economic assessments of value and motivation. *Managerial and Decision Economics, 37,* 306–323. doi: 10.1002/mde.2718

Sigurdsson, S. O., Taylor, M. A., & Wirth, O. (2013). Discounting the value of safety: Effects of perceived risk and effort. *Journal of Safety Research, 46,* 127–134. doi: 10.1016/j.jsr.2013.04.006

Slowiak, J. M., Dickinson, A. M., & Huitema, B. E. (2011). Self-solicited feedback: Effects of hourly pay and individual monetary incentive pay. *Journal of Organizational Behavior Management, 31*(1), 3–20. doi: 10.1080/01608061.2011.541816

Stafford, D., & Branch, M. N. (1998). Effects of step size and break-point criterion on progressive-ratio performance. *Journal of the Experimental Analysis of Behavior, 70*(2), 123–138. doi: 10.1901/jeab.1998.70-123

Wilder, D. A., Austin, J., & Casella, S. (2009). Applying behavior analysis in organizations: Organizational behavior management. *Psychological Services, 6*(3), 202–211. doi: 10.1037/a0015393

Wine, B., Gilroy, S., & Hantula, D. A. (2012). Temporal (in)stability of employee preference for rewards. *Journal of Organizational Behavior Management, 32*(1), 58–64. doi: 10.1080/01608061.2012.646854

Woolverton, W. L., & Anderson, K. G. (2006). Effects of delay to reinforcement on the choice between cocaine and food in rhesus monkeys. *Psychopharmacology, 186*(1), 99–106. doi: 10.1007/s00213-006-0355-x

Received: January 31, 2016
Final Acceptance: August 10, 2016