

Methodological considerations in the study of delay discounting in intertemporal choice: A comparison of tasks and modes

CARTER L. SMITH

BioVid Corporation, Princeton, New Jersey

AND

DONALD A. HANTULA

Temple University, Philadelphia, Pennsylvania

An experimental comparison of two commonly used delay-discounting procedures (binary choice and fill in the blank) and modes of administration (paper and pencil and computer based) was conducted. Statistically significant main effects were found for task type—steeper discounting was observed in the binary-choice task—but not for mode of administration. As simple evidence of validity, hyperbolic curves consistently provided a better fit to the data than did exponential curves for both tasks. Further, magnitude effects were also observed across conditions. Correlational results varied largely as a function of the discounting index (either k or area under the curve) under consideration. Across the two tasks, discounting indices showed modest levels of reliability [$r(\text{AUC}) = .33$; $r(k) = .75$]. The findings pointed to refinements for both the methodology and criteria that are used to study delay discounting and raised questions about the commonly assumed relationship between discounting and the construct of impulsivity.

Delay discounting is the inflation of temporally proximal outcomes relative to distant outcomes in intertemporal-choice scenarios. It is of interest to researchers in many areas of basic and applied psychology, including addictions (e.g., Audrain-McGovern et al., 2004; Bickel et al., 2007), behavior analysis (e.g., Ainslie & Monterosso, 2003; Critchfield & Kollins, 2001), consumer behavior (e.g., Hantula & Bryant, 2005; Smith & Hantula, 2003), cross-cultural studies (e.g., Du, Green, & Myerson, 2002; Kirby et al., 2002; Ostaszewski, 1997; Ostaszewski, Green, & Myerson, 1998), developmental psychology (e.g., Forzano, Szuba, & Figurilli, 2003; Green, Myerson, & Ostaszewski, 1999), educational administration (e.g., Logue & Anderson, 2001), gambling (e.g., Dixon, Jacobs, & Sanders, 2006; Madden, Ewan, & Lagorio, 2007), health psychology (e.g., Chapman, 1996; Chapman et al., 2001), and organizational psychology (e.g., Saunders & Fogarty, 2001; Schoenfelder & Hantula, 2003).

The most widely accepted form of discounting is a hyperbolic equation that is derived from the matching law (Davison & McCarthy, 1988):

$$V = \frac{A}{1 + kD}, \quad (1)$$

where V is the time-discounted value of the reward, A is the subjective present value, D is the total delay to delivery, and k is a discounting coefficient (Ainslie, 1992;

Mazur, 1987). The constant, 1, is added in the denominator in order to ensure that the curve does not extend to infinity at very short intervals.

Although the concept of hyperbolic discounting developed in the animal learning and conditioning laboratory (Ainslie, 1974; Rachlin & Green, 1972), much recent work concerns human decision making. An important issue concerns the psychometric properties of the methods that are used to estimate discounting parameters. As the concept and its attendant measurement protocols move further into mainstream psychology, delay discounting is increasingly being treated, either implicitly or explicitly, as both a construct and a psychological instrument. Thus, the present article presents a systematic comparison of two commonly used delay-discounting tasks (binary choice vs. fill in the blank) and different modes of administration (computer vs. paper and pencil), as well as an examination of the psychometric properties of discounting tasks.

DISCOUNTING PROCEDURES

Binary-Choice Tasks

These tasks are adapted from a set of procedures that were proposed by Mazur (1987) for studying self-control in nonhuman animals, normally with consumable rewards. In human studies, rewards are generally hypothetical. Typically, these tasks feature hypothetical monetary choices

D. A. Hantula, hantula@temple.edu

(HMCs) that involve evaluations of temporal trade-offs between two rewards, one of which is available immediately—often termed the smaller-sooner alternative (SS)—whereas the other, the larger-later (LL) reward, is delayed by some time period, *t*. In a serial presentation of such choice pairs, participants are asked to select the preferred single reward from each pair. An example of such a pair is shown in the top portion of Table 1. As each new scenario is presented, it features an incremental increase in the value of the adjusting alternative (Reynolds, Karraker, Horn, & Richards, 2003). This process is repeated for a specified series of reward values, very much as in a psychophysical titration procedure (Raineri & Rachlin, 1993). The discount parameter, *k*, is calculated by taking the geometric mean of the two decision scenarios that include the point of preference reversal (assuming that preference reversal occurs). The entire cycle is then repeated for each of the *n* specified delay periods. Normally, a final *k* estimate is obtained by averaging the *k* values from each of a series of increasing delay periods.

Fill-in-the-Blank (FITB) Tasks

These tasks follow Chapman’s (1996) procedure and involve the same premise as that of binary-choice tasks. FITB tasks present participants with a hypothetical reward scenario in which the rewards will become available after various delay periods. Unlike the binary-choice tasks, however, this procedure asks the participant to indicate, for each delay period, the equivalent present value of some LL reward. For example, suppose that the LL alternative is \$1,000 to be delivered in 1 year. The participant would be asked to specify a smaller amount of money that would be as desirable as the LL amount if it were delivered immediately instead of after the proposed delay period. This process is then repeated with the same LL reward amount in any number of delay scenarios, as illustrated in the bottom portion of Table 1.

Comparison of Procedures

Neither procedure is without shortcomings. A number of logistical problems emerge in binary-choice tasks with humans that do not arise in the animal choice studies from which these procedures evolved. One central problem with such procedures is that the length of the discounting task can vary considerably. Longer binary-choice sequences can sometimes exceed 100 choices; for example, Madden, Bickel, and Jacobs (1999) used 189 pairwise choices, and Mitchell (1999) used 138 pairwise choices. In such cases, the problems of fatigue and inaccurate responding are more likely to emerge, because the serial decision scenarios are exceedingly repetitive. Also, because of the incremental presentation of the reward pairs, in which reward sequences remain the same across delay periods, par-

ticipants may be alerted to the pattern and begin to make their decisions before the stimuli are presented. There are several obvious benefits to the FITB approach. The first is expedience; estimates of *k* can be obtained in a fraction of the time that is needed to complete a binary-choice task. The second, related, benefit is the minimization of respondent fatigue effects. At the same time, the FITB procedures place far greater cognitive demands upon the respondent and require more detailed instructions.

MODE OF ADMINISTRATION

Mode-of-administration effects are a common consideration in psychological measurement, and the advent of computer- and Internet-based research protocols makes this issue all the more important (Church, 2001; Finegan & Allen, 1994; Mead & Drasgow, 1993; Pettit, 2002; Stanton, 1998). Systematic variations in responding can occur simply as a result of the differences between surveys administered via the Internet, over the telephone, by mail, or in person (see, e.g., J. F. Epstein, Barker, & Kroutil, 2001; Fouladi, McCarthy, & Moller, 2002). As methodological artifacts, mode differences are thought to stem principally from the effects of social-desirability bias (Lautenschlager & Flaherty, 1990; Richman, Kiesler, Weisband, & Drasgow, 1999). Not surprisingly, such differences are most dramatic when sensitive information is requested, such as in the areas of mental health (J. F. Epstein et al., 2001), alcohol abuse (Aquilino, 1994), sexual behavior (Gribble, Miller, Rogers, & Turner, 1999; Turner et al., 1998; Turner, Ku, Sonenstein, & Pleck, 1996), or health-related behaviors (Wilkerson & Nagao, 2002; Wilson, Genco, & Yager, 1985); individuals tend to reveal more personal information and disclose more about stigmatized behaviors in a computerized questionnaire than they do in mail, telephone, or in-person interviews. This issue may be of considerable significance in the present study, because most discounting tasks deal with monetary decisions, which are often associated with norms, and with other socially sensitive information. Given the different levels of anonymity across these two modes of administration, one might expect to observe steeper discounting in a computer-administered condition, because participants are less inclined to mask their possibly “impulsive” or “irrational” decision making.

PSYCHOMETRIC ELEMENTS OF DISCOUNTING TASKS

Reliability

Simpson and Vuchinich (2000) found that the 1-week test-retest reliability for HMC tasks was very strong (*r* =

Table 1
Sample Items From Delay-Discounting Tasks

Sample binary-choice items	
Would you prefer (a) \$100 today or (b) \$1000 in one month?	
Would you prefer (a) \$100 today or (b) \$1000 in 5 years?	
Sample fill-in-the-blank items	
\$ _____ delivered immediately would be just as good as \$1,000 delivered in 1 month.	
\$ _____ delivered immediately would be just as good as \$1,000 delivered in 5 years.	

.91). Johnson and Bickel (2002) reported reliability of $r = .83$ between two methods of establishing discount rates. Subsequently, in a multisession study, Baker, Johnson, and Bickel (2003) found correlations ranging from $r = .71$ to $r = .78$ for hypothetical monetary rewards, and from $r = .76$ to $r = .77$ for real monetary rewards. In the aggregate, these data offer strong preliminary evidence for the stability of k over short intervals. A longitudinal field study of Bolivian Amerindians conducted by Kirby et al. (2002) reported that correlations for both monetary and food rewards actually varied considerably between 3-month periods [$Rng(\text{monetary}) = -.09$ to $.33$; $Rng(\text{food}) = .03$ to $.46$]. Another approach to demonstrating reliability is the measurement of alternate forms. L. Epstein et al. (2003) compared the Kirby discounting questionnaire and Richards's AA discounting task, demonstrating evidence ($r = .82$) for alternate-form reliability. Similarly, Kowal, Yi, Erisman, and Bickel (2007) found correlations ranging from $r = .60$ to $r = .80$ with two different computer algorithms for presenting delay-discounting tasks.

Dependent Variables: k and AUC

In most studies of intertemporal choice, the primary criterion is an estimate of k (Equation 1); however, there are a number of problems with using k as the criterion. First, k is an appropriate index of discounting only in situations in which the obtained discount curve is accurately described by hyperbolic decay. Second, Myerson, Green, and Warusawitharana (2001) raised conceptual and statistical concerns about k . In principle, if a sensitivity parameter is added to the denominator of the discounting equation shown in Equation 1, both the bias (k) and sensitivity parameters might influence the curvature of an individual's decay function, but only a bias parameter will reflect differences in discounting, whereas changes in a sensitivity parameter would indicate that there are measurement differences. Furthermore, as Myerson et al. (2001) argued, analyzing k values using traditional inferential statistics can be inappropriate because of extreme violations of normality in distributions of k that have been documented in much of the existing literature. Nonparametric tests are less powerful than their parametric counterparts, and they cannot currently analyze data from multiway independent samples (e.g., comparing the amount and type of reward across different subject samples). Finally, k values are highly heterogeneous across different types of commodities, and k can theoretically range from zero to infinity, making comparisons across studies difficult.

Myerson et al. (2001) suggested *area under the curve* (AUC) as an alternative approach to delay-discounting data analysis, one that requires no a priori assumptions about the shape of the discount function or the number of free parameters used in modeling. AUC involves computing the area of the trapezoids that are created by plotting the coordinates of indifference points for each delay period. The equation for the area of the trapezoids is

$$x_2 - x_1 [(y_1 + y_2)/2], \quad (2)$$

where the ordinate represents the proportion of absolute (nondiscounted) reward value, and the abscissa represents

the proportion of maximum possible delay to reinforcement. These values are summed to obtain a total area that ranges from 0.0 to 1.0. Accordingly, the steeper the discounting function is, the smaller the AUC.

Myerson et al. (2001) argued that AUC has several merits that make it appropriate for use in discounting research. First, AUC is designed to handle multiple measurements across time, which allows for a more complete picture of how the value of delayed rewards changes across time (Pruessner, Kirschbaum, Meinlschmid, & Hellhammer, 2003). It is also particularly apt for the assessment of growth and decay between conditions (Matthews, Altman, Campbell, & Royston, 1990). Second, in contrast to many data sets that are derived from multiple observations, AUC is appropriate for parametric statistical analyses. This is because, unlike k values, AUC has an upper and a lower boundary, which has the benefit of reducing the likelihood of radically skewed distributions. AUC analyses allow for effective domain comparisons: k estimates obtained for independent domains cannot be compared, because k has no upper limit and is determined partially by the values that are used to obtain the estimate, whereas AUC, which has a limited range, allows for domain comparisons using a common scale.

To illustrate the differences between the use of k and the use of AUC data for the purposes of assessing discounting, consider the following hypothetical example. Imagine that there are 2 participants (Participant X and Participant Y) in a discounting study in which \$1,000 is the LL reward (e.g., 1 week, 1 month, 6 months, 1 year, 2 years, 3 years). Participant Y discounts much more steeply than does Participant X over short intervals; however, as the intervals increase, their respective rates of discounting reverse, with Participant X discounting relatively steeply. Figure 1 shows the hypothetical data for both participants and includes both k and AUC summary data. Visual inspection of the two indifference curves suggests that the 2 participants discount in a fairly similar fashion. If one were to rely upon only k for assessment of discounting, however, one would conclude that Participant Y uniformly discounts much more steeply than does Participant X. Participant X's k value of 1.204 translates into an annualized discount rate of 379.84%, whereas Participant Y's estimated k value translates into 14,108.77% per year. Discount rates of the magnitude associated with Participant X are not at all uncommon in HMC tasks (cf. Coffey, Gudleski, Saladin, & Brady, 2003). This dramatic difference obscures the relatively modest differences between the respondents' respective indifference curves, and it illustrates the problems that are associated with using averaged k values to estimate an individual's "true" k parameter. The AUC data, however, show that actual reported values yield identical discounting over the entire range of delay periods.

Construct Validity

At one level of conceptualization, the demonstration of construct validity in a delay-discounting task is a straightforward matter, because the construct is relatively simple. Unlike most psychological constructs, which attempt to bundle large and often disparate clusters of behavior, hy-

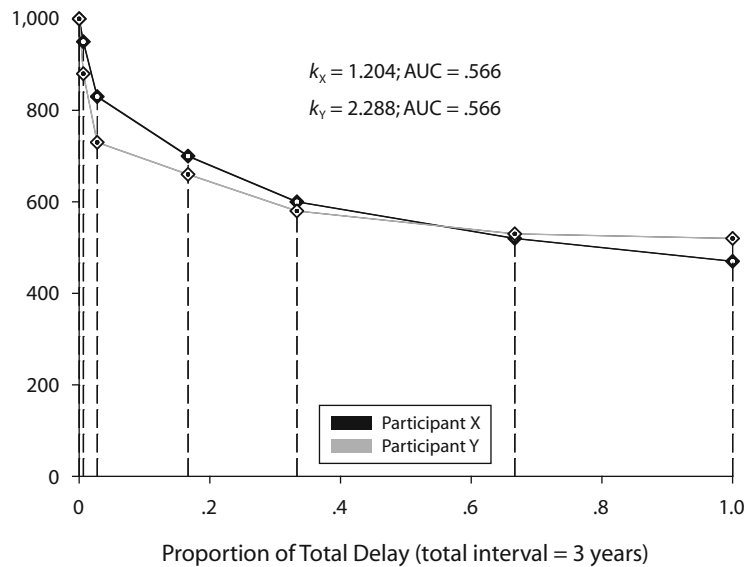


Figure 1. Hypothetical examples of discount curves with *k* and AUC criteria.

perbolic discounting involves only a few factors, including the operant concept of reward (or reinforcement), a time lag, and a simple mathematical decay function. One of the critical insights from research involving discounting instruments is support for the contention that discount functions are hyperbolic (Kirby & Maraković, 1996). If the data fit well with a hyperbolic curve, no conceptual leap of faith is required in order to assert support for construct validity. Furthermore, if the hyperbolic function fits the data better than does a set of plausible rival functions, there is an even stronger justification for the claim. To date, hyperbolic functions have fit individual discounting data better than other decay functions have.

Studies in delay discounting have also addressed the issue of external validity, or generalizability, quite effectively and in line with the multistudy sequence described by Shadish, Cook, and Campbell (2002). First, hyperbolic discounting has been demonstrated numerous times in different study settings and with fairly diverse samples, such as adults (mean age = 42.9; L. Epstein et al., 2003), older adults (mean age = 67.9; Green, Fry, & Myerson, 1994), children (Green et al., 1999), adults from indigenous cultures (Kirby et al., 2002), adults and children from non-Western cultures (Kawashima, 2006; Sorama, Ito, & Saeki, 2007), and college students (Kirby & Santiesteban, 2003). Second, hyperbolic discounting has been found with multiple species. Finally, additional theoretical predictions that have been made by discounting theory have consistently found empirical support, including magnitude effects and domain effects. There is not sufficient space in the present article to review this literature; however, the summaries that are presented in Chapman (1998), Green and Myerson (2004), Kirby (1997), and Rachlin (2000) illustrate the ubiquity of hyperbolic discounting.

Another validation-related issue that was addressed here concerns the evidence for convergent and discriminant validity of the discounting tasks. The present study

considered not only the two extant methods for measuring discounting, but also measures of the psychological construct *impulsivity*, the construct most commonly associated with delay discounting. Several variables that have historically been correlated with impulsivity were also included to provide evidence for a supporting nomological framework of variables, including biosocial characteristics (age and sex), grade point average, and income, all of which serve as a proxy for socioeconomic status.

Impulsivity and Delay Discounting

Because of its implications for the valuation of future outcomes, myopic decision making, and preference reversal, researchers have begun to use the individual discount parameter, *k*, as an operational definition for a latent impulsivity trait (see L. Epstein et al., 2003; Myerson, Green, Hanson, Holt, & Estle, 2003; Richards, Zhang, Mitchell, & de Wit, 1999). Among individual-difference researchers, impulsivity is loosely defined as a predilection for reactions to environmental stimuli that are emitted quickly and without consideration for the negative consequences (Aiken, 2000; Moeller et al., 2001; Whiteside & Lynam, 2003). If impulsivity were framed in behavioral economic terms, it might be said that individuals possess a disposition to make either economically rational or irrational intertemporal decisions (Ainslie, 1992; Daniel, 1997; Strathman, Gleicher, Boninger, & Edwards, 1994). In clinical contexts, impulsivity is associated with various behaviors that supposedly reflect shortsightedness, and it is used as a diagnostic criterion for more than a dozen separate conditions in the *DSM-IV*, such as ADHD, borderline personality disorder, eating disorders (Crean, de Wit, & Richards, 2000; Lacey & Evans, 1986), pathological gambling (Holt, Green, & Myerson, 2003), and various types of substance abuse (Madden, Petry, Badger, & Bickel, 1997; Reynolds et al., 2003).

Delay discounting is now treated, in part, as an issue of individual difference, which allows for the possibility that it has some predictive or diagnostic power. The fact that discounting procedures evolved conceptually and methodologically from animal research using concurrent chain schedules (see Ainslie, 1992; Rachlin, 2000; Richards, Mitchell, de Wit, & Seiden, 1997) makes the seeming shift toward the study of individual differences ironic. Given the dominant status of trait-based theory in the impulsivity literature, the increased interest in delay discounting as an individual difference is not surprising (Crean et al., 2000; Petry, Bickel, & Arnett, 1998). Although the appropriateness of this characterization is debatable (Logue & Anderson, 2001), it is gaining traction (Myerson et al., 2003).

The present study was intended to provide some basic psychometric data regarding these two (binary choice and FITB) discounting procedures in both paper-and-pencil and computer-based formats. Both procedures yield hyperbolic discount curves, but it is not known whether the curves that are obtained with each procedure are identical. The construct validity of these discounting procedures has also been assessed through correlations with two measures of impulsivity; previous studies have used only one impulsivity measure, and because impulsivity measures are not invariant in their relations with one another, concurrent measurement of the construct is important in this context.

METHOD

Participants

Undergraduate students (31 female and 23 male) at a large north-eastern state university, with a mean age of 20.37 years ($SD = 2.99$), participated for course credit. Fifty percent of the participants self-identified as Caucasian, 24% self-identified as African American, 2% self-identified as Hispanic, 2% self-identified as Asian, 10% indicated "Other" as their ethnicity, and the remainder did not respond to the question. The mean grade point average was 2.9/4.0 ($SD = .46$).

Apparatus and Materials

Preexperimental materials were presented in a questionnaire that featured items concerning basic demographic behavioral and financial information. The questionnaire also contained the EPQ-impulsiveness (EPQ-I) (24 items; $\alpha = .82$ to $.85$; Eysenck & Eysenck, 1978) and sensation-seeking scale-disinhibition (SSS-D) (10 items; $\alpha = .74$ to $.78$; Zuckerman, 1994) scales. Items from the latter scale are taken from form *V* of the SSS, which are thought to possess greater internal consistency than several long-form alternatives do (Zuckerman, 1994). Response options for both scales are dichotomous. These scales were selected on the basis of meta-analytic work on correlates of discounting, which showed that these two inventories had the strongest associations with k from among a group of commonly used impulsivity scales (Smith, 2004).

Delay-Discounting Tasks

Participants completed the discounting tasks in one of two modes of administration. For each mode, the participant completed both a traditional binary-choice discounting task and an FITB task. Each reward-magnitude condition in the binary-choice task involved 19 binary monetary choices. In the small-reward condition, the LL referent was \$1,000, and the SS alternatives were \$1, \$5, \$10, \$25, \$50, \$100, \$200, \$300, \$400, \$500, \$600, \$700, \$800, \$900, \$950, \$975, \$990, \$995, or \$999, in sequence. In the \$10,000-reward condition,

the LL referent and the set of SS values used in the small-reward condition were simply multiplied by 10 (e.g., the final choice in the sequence was \$10,000 vs. \$9,990). Six delay periods were used with each hypothetical reward (1 week, 1 month, 6 months, 1 year, 3 years, or 5 years). To complete this task, each participant made 228 discrete binary choices to produce one final k estimate for each of the two rewards (\$1,000 and \$10,000). The FITB procedure also featured both small- and large-reward conditions. Instead of 228 binary choices, the respondent answered 6 FITB questions in each reward-magnitude condition, and these were used to produce one final k estimate for each condition.

Mode of Administration: Paper-and-Pencil Tasks

Binary choice. In the binary-choice task, participants were shown a series of hypothetical monetary choices that were presented on separate, enumerated sheets of paper in a large binder by a trained research assistant. Each page displayed two large boxes, labeled "A" and "B." Alternative B was the LL option (either \$1,000 or \$10,000) and was fixed for each delay period (e.g., \$1,000 available in 6 months). Alternative A was the SS option, and it was always available immediately (e.g., \$10 available immediately). Participants were also given a separate form on which to record their responses, and they indicated their preference by circling either "A" or "B" on the numbered response sheet. Once the participant had indicated a choice, the research assistant presented the next choice pair. In order to minimize fatigue, research assistants were trained to switch the choice pairs while the participants recorded their responses.

Fill in the blank. The FITB portion was presented on a single one-sided 8.5×11 -in. sheet that contained 12 discounting scenarios and a blank line next to each scenario for the participant's response. Here, the participant simply wrote in a hypothetical lump sum of money that would be as valuable as the LL referent if it were delivered immediately. Although the research assistant was on hand to answer questions, this task was essentially self-administered.

Mode of Administration: Computer-Based Tasks

Participants performing the tasks on a computer were presented with the identical set of decision scenarios (for both binary or FITB tasks) on standard Pentium-class computers via a program written in VisualBASIC 6.0. Verbal instructions in the computer condition were nearly identical to those in the paper-and-pencil condition, and any textual changes reflected the distinct physical contingencies of working in one medium or the other (e.g., "click" vs. "write").

Binary-choice task. In the binary-choice task, the monitor displayed two large boxes, labeled "A" and "B," which showed the SS and LL alternatives, respectively. Participants indicated their preference with a mouse click on either box, which recorded their response and advanced them to the next pair of hypothetical rewards.

Fill in the blank. In the FITB task, the verbal descriptions of the 12 decision scenarios were also identical to those in the paper-and-pencil version. Participants typed in their desired SS value and clicked a button to record their response.

Procedure

Participants were randomly assigned to either the paper-and-pencil or the computer-based mode of administration. The order of presentation of tasks within mode of administration was randomized. Upon arriving at a central laboratory, participants were seated at isolated desks outside the lab entrance, where they completed a consent form and the preexperimental survey before being escorted to one of several isolated laboratory rooms. Participants in the paper-and-pencil condition were seated at desks with an instructions form and were told to read the instructions carefully. After answering any questions, the research assistant administered either the binary-choice or the FITB task. The second task was administered after the participant had been given the opportunity for a brief rest. Following completion of the second task, the participant was debriefed by a research assistant.

In the computer condition, participants were seated in front of a workstation in one of several computer laboratories. Participants in this condition were told that all necessary instructions would be presented on the computer monitor. After reading an instruction screen, the participant advanced to the experiment by clicking a command button. The series of questions in both the binary-choice and the FITB conditions in the computerized task was identical to the series used in the paper-and-pencil mode. At the end of each computerized task, the program automatically offered the participant a rest period and then presented them with the next task. When both tasks were completed, the program alerted the participant that the experiment was complete, and the research assistant proceeded with the debriefing.

Design

This study conforms to a 2 × 2 × 2 mixed design with two within-subjects factors and one between-subjects factor. The first within-subjects factor was the reward magnitude (\$1,000 or \$10,000), which is common in most discounting studies. The second within-subjects factor was the order of presentation of the discounting tasks (binary or FITB). The between-subjects factor was the mode of administration (paper and pencil vs. computer based).¹

Dependent Variables and Data Analysis

AUC and *k* values are reported for all analyses.

RESULTS

The data interpretation here has been guided more by effect sizes than by *p* values; however, in keeping with convention in psychology, the present study had a sample size sufficient to detect medium effects at power = .80 (Cohen, 1992).

Distribution

Table 2 shows descriptive statistics, skew, and kurtosis for nontransformed *k* values and AUC values across task type, mode of administration, and reward magnitude. As expected, the distributions of *k* estimates that were derived from both modes of administration violated the assumption of normality, showing fairly extreme positive skew, and kurtosis also departed significantly from zero. As such, all *k* values were log transformed. At $\alpha = .01$ (Tabachnik & Fidell, 2001), neither skew nor kurtosis for any of the four transformed distributions departed significantly from normality. All *k*-based analyses presented hereafter are based on the log-transformed values. AUC distributions did not depart significantly from normality.

Missing values were addressed in one of two ways. If a participant's values were missing for an entire level within a condition (e.g., all FITB values for the \$1,000 reward

were missing), the cell mean *k* and AUC values were substituted for that person. This occurred in two instances that resulted from program errors in recording responses to an external drive, and once in the paper-and-pencil FITB condition. In several cases with random missing values, nonlinear regression was performed on each individual participant's observed indifference values, where the model equation was the hyperbolic decay function shown in Equation 1. The regression analysis provided a predicted indifference value for each delay period. These best-fit values were then substituted for the missing indifference points.

Reliability

Prior to the performance of reliability tests, the AUC and *k*-estimate data were transformed to *z*-score equivalents. The alternate-form reliability coefficient comparing the AUC values that were produced by the two task types was $r = .332$ ($p = .000$). This same alternate-form coefficient was $r = .748$ ($p = .000$) for the *k* estimates. The correlation in AUC by task type within the small-reward condition was statistically significant ($r = .270, p = .024$), as was the correlation between task types in the large-reward condition ($r = .393, p = .002$). The correlation in *k* estimate by task type within the small-reward condition was statistically significant ($r = .869, p = .000$), as it was within the large-reward condition ($r = .626, p = .000$).

Validity

Discounting form: Construct validity. Tables 3A–3C show curve-fit summaries for reward magnitude nested within mode of administration and task type. Comparisons between exponential discounting and hyperbolic discounting supported hyperbolic discounting. Residual-least-squares analysis illustrated that a hyperbolic decay function of the form that was suggested by Mazur (1987) fit the data better than did an exponential decay function in all but two cases. These findings replicated the extensive evidence for the superiority of hyperbolic decay over exponential decay in scenarios that involved human time discounting (e.g., Kirby, 1997).

Reward magnitude effects: Construct validity. Figure 2 shows plots of mean points of indifference for each hypothetical reward as a function of delay. The plots illustrate that, for both the \$1,000 and \$10,000 rewards, the binary-choice task resulted in consistently steeper discounting. The experimental condition effects were also

Table 2
Descriptive Statistics for Nontransformed *k* Values and AUC

Criterion	Condition	<i>M</i>	<i>SD</i>	Skew	<i>z</i>	Kurtosis	<i>z</i>
<i>k</i>	FITB \$1,000	2.22	4.50	3.33	10.23*	10.95	17.14*
	FITB \$10,000	0.68	1.04	3.41	10.49*	13.17	10.61*
	Binary \$1,000	9.40	18.84	3.05	9.38*	10.12	15.84*
	Binary \$10,000	3.05	7.75	3.60	11.08*	23.83	37.29*
AUC	FITB \$1,000	0.694	0.24	-0.57	1.76	-0.409	0.64
	FITB \$10,000	0.748	0.22	-0.60	1.84	-0.760	1.19
	Binary \$1,000	0.552	0.27	-0.26	0.78	-0.952	1.49
	Binary \$10,000	0.635	0.23	-0.44	1.35	-0.410	0.64

Note—Skew and kurtosis were tested against a conservative error level of $p < .01$. $N = 54$.

Table 3A
R² Values for Exponential and Hyperbolic Curve Fits Across Conditions:
Divided by Mode of Administration and Task Type

	Paper and Pencil				Electronic			
	Binary		FITB		Binary		FITB	
	Small	Large	Small	Large	Small	Large	Small	Large
R ² (Hyperbolic)	.927	.997	.953	.983	.993	.994	.994	.990
R ² (Exponential)	.909	.998	.923	.974	.993	.992	.991	.985

Table 3B
R² Values for Exponential and Hyperbolic Curve Fits
Across Conditions: Collapsed Across Task Type

	Paper and Pencil		Electronic	
	Small	Large	Small	Large
R ² (Hyperbolic)	.988	.984	.979	.993
R ² (Exponential)	.977	.978	.972	.986

Table 3C
R² Values for Exponential and Hyperbolic Curve Fits
Across Conditions: Collapsed Across Mode of Administration

	Binary		FITB	
	Small	Large	Small	Large
R ² (Hyperbolic)	.981	.991	.990	.984
R ² (Exponential)	.970	.987	.983	.975

tested using a three-factor mixed model ANOVA, the results of which are shown in Tables 4 and 5.

AUC. All main and interaction effects for the AUC ANOVA were tested against a 1% error level ($\alpha = .01$), and the results are shown in Table 4. The Levene statistic, which tested for the assumption of homogeneity of variance, was nonsignificant ($p = .375$). The intraclass correlation coefficient for the within-subjects factors was $r = .249$, indicating reasonable efficiency for the within-subjects model.² The main effect of the between-subjects factor, mode of administration, was nonsignificant [$F(1,52) =$

$0.043, p = .837, r = .029$], indicating that there were no differences in discounting between the paper-and-pencil and electronic modes of administration; a power analysis (Cohen, 1992) showed that over 500 participants would be needed for this “effect” to be statistically significant. The main effects for both within-subjects factors were statistically significant and produced moderate effect sizes. For the type of discounting task, the effect size ($r = .452$) was based on $F(1,52) = 13.364, p = .001$. The main effect for reward magnitude was nearly identical, in terms of both statistical significance [$F(1,52) = 13.430, p = .010$] and

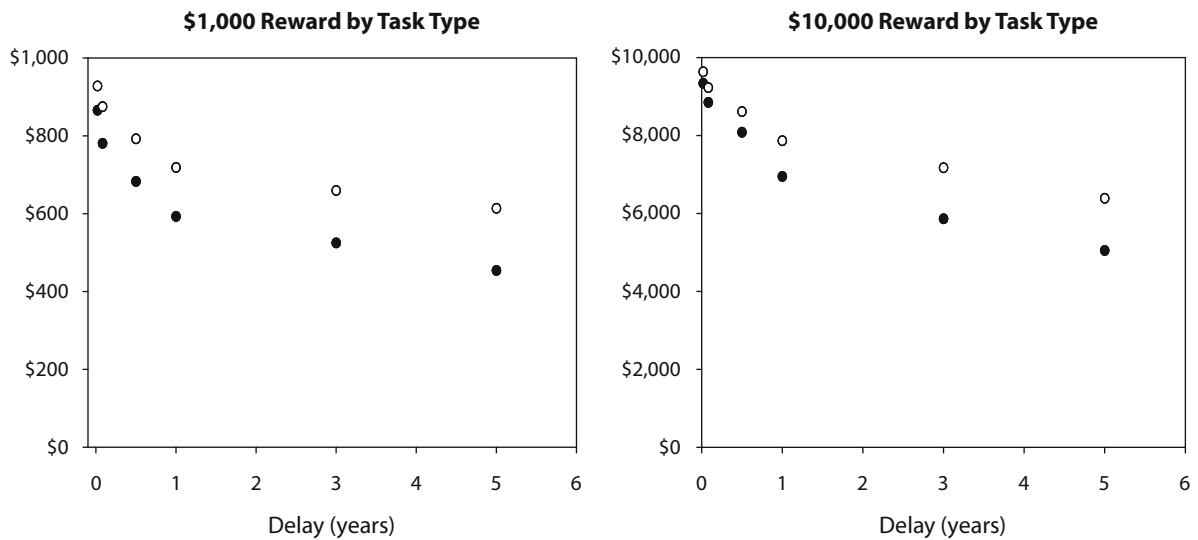


Figure 2. Observed mean indifference points by task type and reward magnitude. Open circles represent data from the FITB task, and closed circles represent data from the binary-choice task.

Table 4
ANOVA for Area Under the Curve

Source	SS	MS	F(1,52)	p	r
Between Subjects					
Mode of administration (M)	0.006	0.006	0.043	0.837	.029
Subjects within mode	7.033	0.135			
Within Subjects					
Discounting task (DT)	0.882	0.882	13.364	0.001	.452
DT × M	0.175	0.175	2.650	0.110	
DT × subjects within mode	3.406	0.066			
Reward magnitude (RM)	0.255	0.255	13.420	0.001	.453
RM × M	0.091	0.091	4.789	0.033	
RM × subjects within mode	1.012	0.019			
DT × RM	0.011	0.011	0.785	0.380	
DT × RM × M	0.000	0.000	0.000	1.000	
DT × RM × subjects within mode	0.708	0.014			

Table 5
ANOVA for *k*-Estimate Data

Source	SS	MS	F(1,52)	p	r
Between Subjects					
Mode of administration (M)	0.033	0.033	0.049	.826	.031
Subjects within mode	35.091	0.675			
Within Subjects					
Discounting task (DT)	0.883	0.883	1.798	.186	.060
DT × M	0.216	0.216	0.440	.510	
DT × subjects within mode	25.546	0.491	0		
Reward magnitude (RM)	17.277	17.277	51.573		.706
RM × M	0.028	0.028	0.083	.774	
RM × subjects within mode	17.399	0.335			
DT × RM	1.641	1.641	2.744	.104	
DT × RM × M	1.262	1.262	2.110	.152	
DT × RM × subjects within mode	31.093	0.598			

effect size ($r = .453$). None of the two-way interactions was statistically significant, although the interaction of reward magnitude with mode of administration approached the critical error level ($p = .033$). The three-way interaction was also not significant.

***k* estimates.** Experimental effects were then tested using *k* estimates, and the results of this ANOVA are shown in Table 5. For this analysis, the Levene statistic was not significant ($p = .097$). The intraclass *r* for the *k* ANOVA was lower than that for the AUC analysis [$r(I) = .066$]. None of the two- or three-way interactions was statistically significant. For the *k*-estimate criterion, the only statistically significant main effect was for reward magnitude, which was large [$F(1,52) = 51.573, p = .000, r = .706$].

Bland–Altman analysis. The leftmost panels of Figure 3 show Bland–Altman plots, with AUC as the discounting criterion. For the small-reward condition, the mean discrepancy between the AUC estimates produced by the two types of discounting tasks was .142; the 2-*SD* interval was $-.575$ to $.164$. This result mirrors that of L. Epstein et al. (2003), who showed that the *k* estimates produced by two alternative discounting tasks tended to fluctuate substantially about the mean difference. The mean difference in AUC estimates in the large-reward condition was $.113$. The 2-*SD* interval for the large-reward condition was $-.361$ to $.134$. It is clear that the binary and FITB tasks produced, in some cases, fairly dramatic differences in estimates of discounting that were based on hypotheti-

cal rewards; however, these differences were relatively stable across reward magnitude (cf. L. Epstein et al., 2003). The Bland–Altman procedure was also applied to the *k*-estimate data, and the rightmost panels of Figure 3 show the findings from this analysis. For small rewards, the mean difference score between task types was 0.43; for the large reward, it was 0.44. The 2-*SD* interval was 0.026 to 0.849 for the small-reward condition and -0.143 to 1.015 for the large-reward condition. In this case, the estimation of *k* became more disparate across tasks as the size of the rewards increased.

Evidence for Nomologic Support

Convergent validity. Table 6 shows correlations between discounting criteria and impulsivity instruments by task types, collapsed across reward magnitudes. Although all correlations are in the expected direction, the magnitude of the relationships between each of the traditional psychometric indices of impulsivity and discounting varied mostly as a function of the discounting criterion, and less as a function of task type. In the more easily interpreted scenario, the EPQ–I is significantly correlated (approximately $r = .30$) with log-transformed *k* values. Conversely, these correlations are much weaker ($r = -.01$ to $-.10$) and nonsignificant when AUC is the discounting criterion. Whereas the *k*-based correlations provide reasonable evidence for convergent validity, the AUC-based correlations do not. The SSS–D correlations

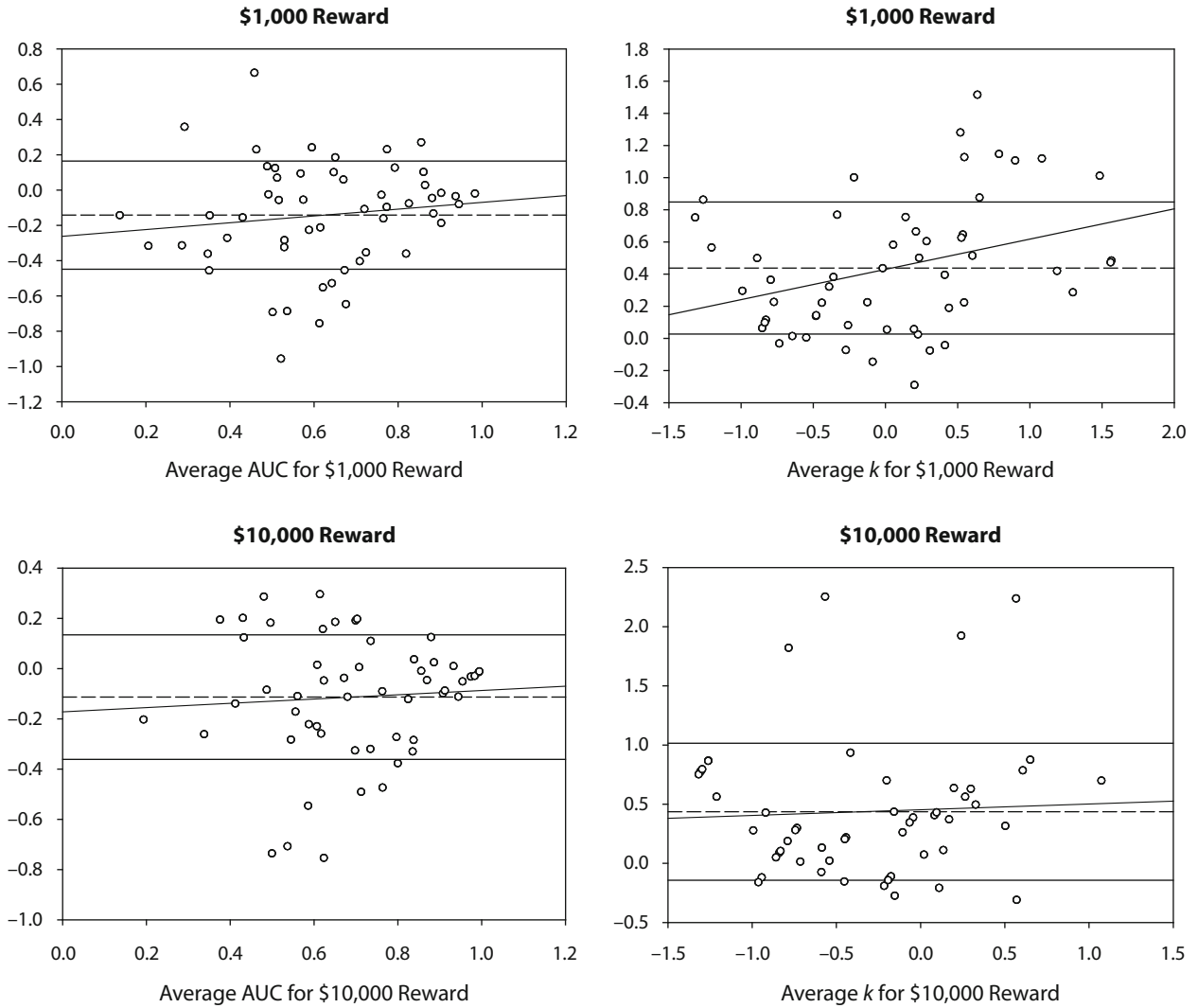


Figure 3. Bland–Altman plots for between-task AUC estimates at \$1,000 (top left) and \$10,000 (bottom left), and between-task *k* estimates at \$1,000 (top right) and \$10,000 (bottom right).

are weak (approximately $r = .10$) when *k* is the criterion. When AUC is the criterion, binary-choice discounting is not related to the SSS–D, and the FITB discounting correlation is moderate and statistically significant ($r = .28$). To place these findings in context, the correlation between the EPQ–I and the SSS–D was $r = .548$. Averaging across reward-magnitude conditions, the correlation between the

discounting criteria, *k* and AUC, was $r = -.310$. This correlation was consistently higher in the binary-choice conditions ($r = -.366$; $r = -.379$) than it was in the FITB conditions ($r = -.224$; $r = -.269$).

Concurrent validity. Table 7 shows correlations between different possible correlates of observed discounting and discounting indices by type of task, including correlations with several individual difference variables that have been shown to relate to impulsivity. Sex showed a fairly consistent but small correlation with discounting criteria. Males discounted less steeply than females did; the mean correlation between sex and AUC was $r = .175$, and the mean correlation between sex and *k* was $r = .089$. Age showed a fairly consistent but small correlation with *k* across conditions, with a mean correlation of $r = -.164$, indicating that older participants discounted less steeply than younger ones did; however, the relationship diminishes substantially when AUC is the criterion and the mean is only $r = .022$. GPA showed a fairly consistent but small correla-

Table 6
Convergent Validity Correlations

Scale	<i>k</i>		AUC	
	Binary	FITB	Binary	FITB
EPQ–I	.301*	.298*	-.018	-.103
SSS–D	.103	.129	.046	-.284*

Note—Reported values are collapsed across reward magnitudes. All values are based on *z* transformations, with the exception of gender, for which the reported point-biserial *r* is based on *z* log values; SSS–D, sensation-seeking scale (disinhibition); EPQ–I, Eysenck personality inventory (impulsivity). *N*(total) = 54. * $p < .05$.

Table 7
Concurrent Validity Correlations

Scale	<i>k</i>		AUC	
	Binary	FITB	Binary	FITB
Sex	-.157	-.026	.152	.205
Age	-.158	-.169	.075	-.031
Income	.068	.210	-.044	-.043
GPA	-.182	-.068	.298*	.064

Note—Reported values are collapsed across reward magnitudes. All values are based on *z* transformations with the exception of gender, where the reported point-biserial *r* is based on *z* log values. GPA, grade point average. *N*(total) = 54, except for income correlations, in which 13 participants (24%) did not provide estimates. **p* < .05.

tion with discounting criteria, but only for the binary-choice condition (*r* = -.182 for *k*; *r* = .298 for AUC). Finally, the relationship between income and discounting criteria was erratic, with no consistent pattern emerging.

DISCUSSION

Procedural Findings

Mode differences. Discounting was not significantly or practically different between the computerized and paper-and-pencil modes, indicating that mode of administration should not be a substantive concern for ongoing discounting research. Given the extremely small effect sizes for mode (approximately *r* = .03), these null results are not due to an underpowered study. Furthermore, there was no evidence of interaction effects, which suggests that the relative consistency of the modes held for both task types and for both reward magnitudes.

Task differences. Whether or not different discounting tasks produced divergent discounting metrics depends on which criterion is used to make that judgment. In the present study, AUC-based results clearly differed across the two task types, whereas *k*-based results did not. Clearly, some caution is warranted in selecting one task type versus the other. Some practical considerations should also be noted. Participants found the FITB tasks to be cognitively demanding, indicating during debriefing that the scenario that was presented in the instructions was complicated. The FITB procedure may encourage them to insert undesired contextual variables into their own analyses of each decision scenario. Debriefing interviews suggested that this occurred with at least some of the participants. Participants also suggested that no such problems arose with the interpretation of the binary-choice task. Another possible account of the observed between-task differences is that the instructions created a systematic bias toward shallower discounting. For example, the phrase “would be just as good as” may have created an endowment effect in which the monetary referent (in this case, the LL reward) was regarded as money that was already in the participant’s possession.

In studies that have used the FITB task, it has been presupposed that the resulting discount curve would be the same as that in the binary-choice task (Chapman, 1996; Schoenfelder & Hantula, 2003); indeed, these studies have referenced the hyperbolic discounting function in Equation 1 and later fit the data to this curve. Because the resulting data fit well with the hypothesized hyper-

bolic function, there appears to have been an assumed equivalence between the two procedures. This assumption appears to be unsupported, however. From the standpoint of validity, these data cannot establish conclusively which of the two procedures results in the measurement of “true” discounting, but because the hyperbolic discounting model was first developed using a binary-choice task, with both animals and humans (e.g., Ainslie, 1974; Rachlin & Green, 1972; Rachlin, Raineri, & Cross, 1991), the burden of proof is on the newer FITB task.

Psychometric Findings

Reliability. Alternate-form coefficients were moderate in the case of AUC (*r* = .332) and strong in the case of *k* (*r* = .748). When *k* was the index, the data met psychometric standards for basic research but remained inadequate for diagnostic purposes (Nunnally & Bernstein, 1994). Also, *k* appeared to be stable across mode of administration and task type, differing only on reward magnitude (e.g., Schoenfelder & Hantula, 2003). In contrast, psychometric standards for alternate-form reliability were not met by the AUC data. It is interesting to note that the previous research that demonstrated the test–retest reliability of discounting tasks (e.g., Johnson & Bickel, 2002; Simpson & Vuchinich, 2000) employed *k* as the sole criterion.

Construct validity. The results provide good psychometric support for the hyperbolic discounting construct. Discount curves were described better by a hyperbolic decay function than by an exponential function, consistent across all conditions, and further solidified the evidence that supports the essential shape of the discounting function (e.g., Green, Myerson, & McFadden, 1997; Kirby & Maraković, 1996; Madden et al., 1999). Furthermore, the \$1,000 reward was discounted more than the \$10,000 reward was, consistent with past magnitude effects (Chapman & Winquist, 1998; Kirby, 1997; Vuchinich & Simpson, 1998).

Convergent validity. If we consider *k* exclusively, the statistically significant but moderate positive correlation with the EPQ–I and the lack of correlation with the SSS–D in Table 6 appear to have generated reasonable evidence for convergent and divergent validity according to standard psychometric conventions (e.g., Nunnally & Bernstein, 1994; Shadish et al., 2002). On the basis of the observed correlations, one might argue that intertemporal exchange is a central conceptual component of the EPQ–I but only a peripheral component of the SSS–D. This would indicate that discounting is tightly associated with impulsivity but retains important differences from sensation seeking; however, the AUC-based data disturb this clear picture. Only one of the impulsivity–AUC correlations is statistically significant—a negative relationship between discounting and the SSS–D using the FITB task—whereas the remainder are nearly nil.

If anything, the preponderance of the data suggests that impulsivity and delay discounting are, at best, randomly associated, as Reynolds, Ortengren, Richards, and de Wit (2006) have also reported. There are a number of possible explanations for this incongruence. The most obvious explanation is that the two impulsivity scales themselves do not represent the same construct. Despite the extensive ar-

guments for multiple dimensions of impulsivity and related constructs in the personality literature, impulsivity has been consistently defined with respect to intertemporal choice (Miller, Flory, Lynam, & Leukefeld, 2003; Whiteside & Lynam, 2003). To see this in the present case, consider the items shown in Table 8, which are sampled from the EPQ-I and SSS-D scales. Intertemporal choice is a component (if not the entirety) of the construct definitions in both cases, and it is particularly central to the specification of representative behaviors. Even if the correspondence between the items is not always completely clear, the construct definitions are remarkably similar (Joireman, Anderson, & Strathman, 2003). The construct clearly sets proximal rewards (e.g., excitement or current pleasure) against future rewards that are at risk (e.g., social approval or financial security) in such a way that the former may need to be sacrificed for the latter, just as we would expect in an intertemporal exchange. Whether one construes the basis for impulsivity in terms of a desire to seek exciting stimuli in exchange for taking on greater risk (EPQ-I), or as a tendency to *think* only about short-term consequences (SSS-D), the essential problem remains one of reward valuation and intertemporal choice. In the present study, then, it is not surprising to find a moderate correlation ($r = .54$) between the two measures. Although a portion of this correlation could stem from common method variance, the datum still supports the assertion of construct correspondence and minimizes the plausibility of an argument that the lack of congruence in the impulsivity-discounting correlations stems from the orthogonality of the two impulsivity scales.

The two scales ought to be closely associated with one another *and* with discounting. Given that the SSS-D has an intertemporal-choice component, another explanation for these findings is that delay discounting, as measured by an HCM task, is simply not reliably associated with or representative of a latent impulsivity trait. This

would be a reasonable conclusion if both the k and the AUC data did not show the expected relations to the impulsivity scales; however, although some past research and the present study have found moderate correlations between k and EPQ-I, there was no correlation between AUC and impulsivity measures with the binary-choice test, and there was a statistically significant negative relationship between the SSS-D and AUC using FITB. Given that most k -impulsivity relations have been derived using the binary-choice task, and considering the quantitative limitations of k , it may be that the relation between k and impulsivity is attributable to measurement artifact.

Dependent Measures Differences

The consistent differences between k and AUC constitute one of the most salient findings in this research. The Bland-Altman analyses of task difference illustrated this pattern; the between-task differences in k estimates increased as a function of reward magnitude in such a way that the alternate form correlation was weaker with large rewards, a result that replicates that of L. Epstein et al. (2003). When AUC was used as the discounting criterion, however, the between-task differences were quite stable from the \$1,000 to the \$10,000 condition. Even the interpretation of the main effects hinged, at least in part, on the discounting index. The magnitude of the two within-subjects factors (task type and reward magnitude) varied dramatically across the two indices. Whereas the AUC data clearly showed that task difference produced a robust effect size, the equivalent statistic for the k -estimate data was small and nonsignificant.

CONCLUSIONS

The present study provides some evidence that the HMC is psychometrically justified. Such measures appear to

Table 8
Content From Two Measures of Latent Impulsivity

Scale	Item Samples	Construct Definition
EPQ-Impulsivity (Eysenck & Eysenck, 1978)	Do you often buy things on impulse? Do you generally do and say things without stopping to think? Are you an impulsive person? Do you usually think carefully before doing anything? Do you often get involved in things you later wish you could get out of? Do you get so "carried away" by new and exciting ideas that you never think of possible snags?	A tendency to act without thinking or without regard to consequences, and to engage in risky behavior.
Sensation-seeking scale-Disinhibition (Zuckerman, 1994)	Do you get so "carried away" by new and exciting ideas that you never think of possible snags? I would never want to try jumping out of a plane, with or without a parachute. I prefer friends who are excitingly unpredictable. I like to have new and exciting experiences and sensations even if they are a little frightening, unconventional, or illegal.	A trait defined by the seeking of varied, novel, complex, and intense sensations and experiences and by the willingness to take physical, social, legal, and financial risks for the sake of such experiences.

enjoy consistent supporting evidence for various types of validity and reliability. From a psychometric perspective, k is clearly supported over AUC. Irrespective of which metric we choose, however, the findings do not support a conceptual overlap between impulsivity and delay discounting. The preliminary nomologic evidence here was inconsistent across measures, since the strength of the relationship between discounting and various supporting correlates varied dramatically as a function of the discounting criterion. The results provide little support for the assertion that discounting and impulsivity represent a unitary underlying trait, since the statistical evidence is ultimately far too erratic, in terms of both convergent validity and supporting relationships, to allow for such a conclusion.

In terms of measurement, several practical findings emerged. First, researchers should exercise caution with respect to criterion selection when measuring delay discounting. Although it has been argued that AUC has superior statistical and conceptual properties as a dependent variable for studying discounting (Myerson et al., 2001), AUC produces results that differ dramatically from those of the customary criterion, k . In turn, researchers should keep in mind that the discounting task that they use could yield more or less perceived discounting depending on the criterion with which it is paired. On the basis of the present research, however, a “best practice” delay-discounting protocol would be one in which either a paper-and-pencil or a computer-based binary-choice procedure is used. Unfortunately, this research did not conclusively demonstrate the superiority of either discounting metric. Both AUC and k measures should be reported in future studies in order to provide researchers with sufficient data to revisit this issue. If researchers do pursue delay discounting for the purposes of individual-difference research, k is the preferred criterion.

AUTHOR NOTE

We thank Carolyn Ahlstrom for her advice concerning mode comparison effects and Samuel Perry for his invaluable assistance in implementing the research. Correspondence should be addressed to D. A. Hantula, Department of Psychology, Weiss Hall (265-67), Temple University, Philadelphia, PA 19122 (e-mail: hantula@temple.edu).

REFERENCES

AIKEN, L. R. (2000). *Personality: Theories, assessment, research, and applications*. Springfield, IL: Charles C. Thomas.

AINSLIE, G. W. (1974). Impulse control in pigeons. *Journal of the Experimental Analysis of Behavior*, **21**, 485-489.

AINSLIE, G. [W.] (1992). *Picoeconomics: The strategic interaction of successive motivational states within the person*. New York: Cambridge University Press.

AINSLIE, G. [W.], & MONTEROSSO, J. R. (2003). Building blocks of self-control: Increased tolerance for delay with bundled rewards. *Journal of the Experimental Analysis of Behavior*, **79**, 37-48.

AQUILINO, W. S. (1994). Interview mode effects in surveys of drug and alcohol use: A field experiment. *Public Opinion Quarterly*, **58**, 210-240.

AUDRAIN-MCGOVERN, J., RODRIGUEZ, D., TERCYAK, K. P., EPSTEIN, L. H., GOLDMAN, P., & WILEYTO, E. P. (2004). Applying a behavioral economic framework to understanding adolescent smoking. *Psychology of Addictive Behaviors*, **18**, 64-73.

BAKER, F., JOHNSON, M. W., & BICKEL, W. K. (2003). Delay discounting in current and never-before cigarette smokers: Similarities and differences across commodity, sign, and magnitude. *Journal of Abnormal Psychology*, **112**, 382-392.

BICKEL, W. K., MILLER, M. L., YI, R., KOWAL, B. P., LINDQUIST, D. M., & PITCOCK, J. A. (2007). Behavioral and neuroeconomics of drug addiction: Competing neural systems and temporal discounting processes. *Drug & Alcohol Dependence*, **90**, S85-S91.

CHAPMAN, G. B. (1996). Temporal discounting and utility for health and money. *Journal of Experimental Psychology*, **22**, 771-791.

CHAPMAN, G. B. (1998). Sooner or later: The psychology of intertemporal choice. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 38, pp. 83-113). San Diego, CA: Academic Press.

CHAPMAN, G. B., BREWER, N. T., COUPS, E. J., BROWNLEE, S., LEVENTHAL, H., & LEVENTHAL, E. A. (2001). Value for future and preventive health behavior. *Journal of Experimental Psychology: Applied*, **7**, 235-250.

CHAPMAN, G. B., & WINQUIST, J. R. (1998). The magnitude effect: Temporal discount rates and restaurant tips. *Psychonomic Bulletin & Review*, **5**, 119-123.

CHURCH, A. H. (2001). Is there a method to our madness? The impact of data collection methodology on organizational survey results. *Personnel Psychology*, **54**, 937-969.

COFFEY, S. F., GUDLESKI, G. D., SALADIN, M. E., & BRADY, K. T. (2003). Impulsivity and rapid discounting of delayed hypothetical rewards in cocaine-dependent individuals. *Experimental & Clinical Psychopharmacology*, **11**, 18-25.

COHEN, J. (1992). A power primer. *Psychological Bulletin*, **112**, 155-159.

CREAN, J. P., DE WIT, H., & RICHARDS, J. B. (2000). Reward discounting as a measure of impulsive behavior in a psychiatric outpatient population. *Experimental & Clinical Psychopharmacology*, **8**, 155-162.

CRITCHFIELD, T., & KOLLINS, S. (2001). Temporal discounting: Basic research and the analysis of socially important behavior. *Journal of Applied Behavior Analysis*, **34**, 101-122.

DANIEL, T. (1997). Delay of consumption and saving behavior: Some preliminary empirical outcomes. In G. Antonides, W. F. van Raaij, & S. Maital (Eds.), *Advances in economic psychology* (pp. 171-188). New York: Wiley.

DAVISON, M., & MCCARTHY, D. (1988). *The matching law: A research review*. Hillsdale, NJ: Erlbaum.

DIXON, M., JACOBS, E., & SANDERS, S. (2006). Contextual control of delay discounting by pathological gamblers. *Journal of Applied Behavior Analysis*, **39**, 413-422.

DU, W., GREEN, L., & MYERSON, J. (2002). Cross-cultural comparisons of discounting delayed and probabilistic rewards. *Psychological Record*, **52**, 479-492.

EPSTEIN, J. F., BARKER, P. R., & KROUTIL, L. A. (2001). Mode effects in self-reported mental health data. *Public Opinion Quarterly*, **65**, 529-549.

EPSTEIN, L., RICHARDS, J., SAAD, F., PALUCH, R., ROEMMICH, J., & LERMAN, C. (2003). Comparison between two measures of delay discounting in smokers. *Experimental & Clinical Psychopharmacology*, **11**, 131-138.

EYSENCK, S. B. G., & EYSENCK, H. J. (1978). Impulsiveness and venturesomeness: Their position in a dimensional system of personality description. *Psychological Reports*, **43**, 1247-1255.

FINEGAN, J. E., & ALLEN, N. J. (1994). Computerized and written questionnaires: Are they equivalent? *Computers in Human Behavior*, **10**, 483-496.

FORZANO, L. B., SZUBA, M., & FIGURILLI, J. M. (2003). Self-control and impulsiveness in children: Effects of visual food cues. *Psychological Record*, **53**, 161-175.

FOULADI, R. T., MCCARTHY, C. J., & MOLLER, N. P. (2002). Paper-and-pencil or online? Evaluating mode effects on measures of emotional functioning and attachment. *Assessment*, **9**, 204-215.

GREEN, L., FRY, A. F., & MYERSON, J. (1994). Discounting of delayed rewards: A life-span comparison. *Psychological Science*, **5**, 33-36.

GREEN, L., & MYERSON, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological Bulletin*, **130**, 769-792.

GREEN, L., MYERSON, J., & MCFADDEN, E. (1997). Rate of temporal discounting decreases with amount of reward. *Memory & Cognition*, **25**, 715-723.

GREEN, L., MYERSON, J., & OSTASZEWSKI, P. (1999). Discounting of

- delayed rewards across the life span: Age differences in individual discounting functions. *Behavioural Processes*, **46**, 89-96.
- GRIBBLE, J. N., MILLER, H. G., ROGERS, S. M., & TURNER, C. F. (1999). Interview mode and measurement of sexual behaviors: Methodological issues. *Journal of Sex Research*, **36**, 16-24.
- HANTULA, D., & BRYANT, K. (2005). Delay discounting determines delivery fees in an e-commerce simulation: A behavioral economic perspective. *Psychology & Marketing*, **22**, 153.
- HOLT, D. D., GREEN, L., & MYERSON, J. (2003). Is discounting impulsive? Evidence from temporal and probability discounting in gambling and nongambling college students. *Behavioural Processes*, **64**, 355-367.
- JOHNSON, M. W., & BICKEL, W. K. (2002). Within-subject comparison of real and hypothetical monetary rewards in delay discounting. *Journal of the Experimental Analysis of Behavior*, **77**, 129-146.
- JOIREMAN, J., ANDERSON, J., & STRATHMAN, A. (2003). The aggression paradox: Understanding links among aggression, sensation seeking, and the consideration of future consequences. *Journal of Personality & Social Psychology*, **84**, 1287-1302.
- KAWASHIMA, K. (2006). The effects of inflation and interest rates on delay discounting in human behavior. *Psychological Record*, **56**, 551-568.
- KIRBY, K. N. (1997). Bidding on the future: Evidence against normative discounting of delayed rewards. *Journal of Experimental Psychology: General*, **126**, 54-70.
- KIRBY, K. N., GODDY, R., REYES-GARCÍA, V., BYRON, E., APAZA, L., LEONARD, W., ET AL. (2002). Correlates of delay-discount rates: Evidence from Tsimane' Amerindians of the Bolivian rain forest. *Journal of Economic Psychology*, **23**, 291-316.
- KIRBY, K. N., & MARAKOVIĆ, N. N. (1996). Delay-discounting probabilistic rewards: Rates decrease as amounts increase. *Psychonomic Bulletin & Review*, **3**, 100-104.
- KIRBY, K. N., & SANTIESTEBAN, M. (2003). Concave utility, transaction costs, and risk in measuring discounting of delayed rewards. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 66-79.
- KOWAL, B. P., YI, R., ERISMAN, A., & BICKEL, W. K. (2007). A comparison of two algorithms in computerized temporal discounting procedures. *Behavioural Processes*, **75**, 231-236.
- LACEY, J. H., & EVANS, C. D. (1986). The impulsivist: A multi-impulsive personality disorder. *British Journal of Addiction*, **81**, 641-649.
- LAUTENSCHLAGER, G. J., & FLAHERTY, V. L. (1990). Computer administration of questions: More desirable, or more social desirability? *Journal of Applied Psychology*, **75**, 310-314.
- LOGUE, A. W., & ANDERSON, Y. D. (2001). Higher-education administrators: When the future does not make a difference. *Psychological Science*, **12**, 276-281.
- MADDEN, G. J., BICKEL, W. K., & JACOBS, E. A. (1999). Discounting of delayed rewards in opioid-dependent outpatients: Exponential or hyperbolic discounting functions? *Experimental & Clinical Psychopharmacology*, **7**, 284-293.
- MADDEN, G. J., EWAN, E. E., & LAGORIO, C. H. (2007). Toward an animal model of gambling: Delay discounting and the allure of unpredictable outcomes. *Journal of Gambling Studies*, **23**, 63-83.
- MADDEN, G. J., PETRY, N. M., BADGER, G. J., & BICKEL, W. K. (1997). Impulsive and self-control choices in opioid-dependent patients and non-drug-using control patients: Drug and monetary rewards. *Experimental & Clinical Psychopharmacology*, **5**, 256-262.
- MATTHEWS, J. N. S., ALTMAN, D. G., CAMPBELL, M. J., & ROYSTON, P. (1990). Analysis of serial measurements in medical research. *British Medical Journal*, **300**, 230-235.
- MAZUR, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M. L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *Quantitative analysis of behavior: Vol. 5. The effect of delay and intervening events on reinforcement value* (pp. 55-73). Hillsdale, NJ: Erlbaum.
- MEAD, A. D., & DRASGOW, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, **114**, 449-458.
- MILLER, J., FLORY, K., LYNAM, D., & LEUKEFELD, C. (2003). A test of the four-factor model of impulsivity-related traits. *Personality & Individual Differences*, **34**, 1403-1418.
- MITCHELL, S. H. (1999). Measures of impulsivity in cigarette smokers and nonsmokers. *Psychopharmacology*, **146**, 455-464.
- MOELLER, F. G., DOUGHERTY, D. M., BARRATT, E. S., SCHMITZ, J. M., SWANN, A. C., & GRABOWSKI, J. (2001). The impact of impulsivity on cocaine use and retention in treatment. *Journal of Substance Abuse Treatment*, **21**, 193-198.
- MYERSON, J., GREEN, L., HANSON, J. S., HOLT, D. D., & ESTLE, S. J. (2003). Discounting delayed and probabilistic rewards: Processes and traits. *Journal of Economic Psychology*, **24**, 619-636.
- MYERSON, J., GREEN, L., & WARUSAWITHARANA, M. (2001). Area under the curve as a measure of discounting. *Journal of the Experimental Analysis of Behavior*, **76**, 235-243.
- NUNNALLY, J. C., & BERNSTEIN, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- OSTASZEWSKI, P. (1997). Temperament and the discounting of delayed and probabilistic rewards: Conjoining European and American psychological traditions. *European Psychologist*, **2**, 35-43.
- OSTASZEWSKI, P., GREEN, L., & MYERSON, J. (1998). Effects of inflation on the subjective value of delayed and probabilistic rewards. *Psychonomic Bulletin & Review*, **5**, 324-333.
- PETRY, N. M., BICKEL, W. K., & ARNETT, M. (1998). Shortened time horizons and insensitivity to future consequences in opioid-dependent individuals. *Addiction*, **93**, 729-738.
- PETTIT, F. A. (2002). A comparison of World-Wide Web and paper-and-pencil personality questionnaires. *Behavior Research Methods, Instruments, & Computers*, **34**, 50-54.
- PRUESSNER, J. C., KIRSCHBAUM, C., MEINLSCHMID, G., & HELLHAMMER, D. K. (2003). Two formulas for computation of area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology*, **28**, 916-931.
- RACHLIN, H. (2000). *The science of self-control*. Cambridge, MA: Harvard University Press.
- RACHLIN, H., & GREEN, L. (1972). Commitment, choice and self-control. *Journal of the Experimental Analysis of Behavior*, **17**, 15-22.
- RACHLIN, H., RAINERI, A., & CROSS, D. (1991). Subjective probability and delay. *Journal of the Experimental Analysis of Behavior*, **55**, 233-244.
- RAINERI, A., & RACHLIN, H. (1993). The effect of temporal constraints on the value of money and other commodities. *Journal of Behavioral Decision Making*, **6**, 77-94.
- REYNOLDS, B., KARRAKER, K., HORN, K., & RICHARDS, J. B. (2003). Delay and probability discounting as related to different stages of adolescent smoking and non-smoking. *Behavioural Processes*, **64**, 333-344.
- REYNOLDS, B., ORTENGREN, A., RICHARDS, J., & DE WIT, H. (2006). Dimensions of impulsive behavior: Personality and behavioral measures. *Personality & Individual Differences*, **40**, 305-315.
- RICHARDS, J. B., MITCHELL, S. H., DE WIT, H., & SEIDEN, L. S. (1997). Determination of discount functions in rats with an adjusting-amount procedure. *Journal of the Experimental Analysis of Behavior*, **67**, 353-366.
- RICHARDS, J. B., ZHANG, L., MITCHELL, S. H., & DE WIT, H. (1999). Delay or probability discounting in a model of impulsive behavior: Effect of alcohol. *Journal of the Experimental Analysis of Behavior*, **71**, 121-143.
- RICHMAN, W. L., KIESLER, S., WEISBAND, S., & DRASGOW, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology*, **84**, 754-775.
- ROSENTHAL, R., & ROSNOW, R. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.
- SAUNDERS, B., & FOGARTY, G. (2001). Time discounting in relation to career preferences. *Journal of Vocational Behavior*, **58**, 118-126.
- SCHOENFELDER, T. E., & HANTULA, D. A. (2003). A job with a future? Delay discounting, magnitude effects, and domain independence of utility for career decisions. *Journal of Vocational Behavior*, **62**, 43-55.
- SHADISH, W. R., COOK, T. D., & CAMPBELL, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York: Houghton Mifflin.
- SIMPSON, C. A., & VUCHINICH, R. E. (2000). Reliability of a measure of temporal discounting. *Psychological Record*, **50**, 3-14.
- SMITH, C. L. (2004). *Risk sensitivity in intertemporal choice: A synthesis of foraging and behavioral economic theory*. Unpublished doctoral dissertation, Temple University, Philadelphia.

- SMITH, C. L., & HANTULA, D. A. (2003). Pricing effects on foraging in a simulated Internet shopping mall. *Journal of Economic Psychology*, *24*, 653-674.
- SORAMA, M., ITO, M., & SAEKI, D. (2007). Self-control and delay discounting in preschool children. *Japanese Journal of Behavior Analysis*, *20*, 101-108.
- STANTON, J. M. (1998). An empirical assessment of data collection using the Internet. *Personnel Psychology*, *51*, 709-725.
- STRATHMAN, A., GLEICHER, F., BONINGER, D. S., & EDWARDS, C. S. (1994). The consideration of future consequences: Weighing immediate and distant outcomes of behavior. *Journal of Personality & Social Psychology*, *66*, 742-752.
- TABACHNICK, B. G., & FIDELL, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn & Bacon.
- TURNER, C. F., KU, L., ROGERS, S. M., LINDBERG, L. D., PLEACK, J. H., SONENSTEIN, F. L. (1998). Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science*, *280*, 867-873.
- TURNER, C. F., KU, L., SONENSTEIN, F. L., & PLECK, J. H. (1996). Impact of audio-CASI on bias in reporting of male-male sexual contacts. In R. Warnecke (Ed.), *Health survey research methods: Conference proceedings* (pp. 171-176). Hyattsville, MD: National Center for Health Statistics.
- VUCHINICH, R. E., & SIMPSON, C. A. (1998). Hyperbolic temporal discounting in social drinkers and problem drinkers. *Journal of Experimental & Clinical Psychopharmacology*, *6*, 292-305.
- WHITESIDE, S. P., & LYNAM, D. R. (2003). Understanding the role of impulsivity and externalizing psychopathology in alcohol abuse: Application of the UPPS impulsive behavior scale. *Experimental & Clinical Psychopharmacology*, *3*, 210-217.
- WILKERSON, J. M., & NAGAO, D. H. (2002). Socially desirable responding in computerized questionnaires: When questionnaire purpose matters more than the mode. *Journal of Applied Social Psychology*, *32*, 544-559.
- WILSON, F. R., GENCO, K. T., & YAGER, G. G. (1985). Assessing the equivalence of paper-and-pencil vs. computerized tests: Demonstration of a promising methodology. *Computers in Human Behavior*, *1*, 265-275.
- ZUCKERMAN, M. (1994). *Behavioral expressions and biosocial bases of sensation seeking*. New York: Cambridge University Press.

NOTES

1. From a purely statistical perspective, it may be advantageous to design the study so that both task and mode are within-subjects variables. Task was chosen as a within-subjects variable and mode as a between-subjects variable because the tasks comparison was more conceptually and methodologically interesting, and because we could not run a fully within-subjects study because of fatigue effects and possible response-set artifacts.
2. The aggregated variance term was the simple mean of the factor \times subject mean squares (see Rosenthal & Rosnow, 1991).

(Manuscript received September 18, 2007;
revision accepted for publication April 16, 2008.)