

Published in final edited form as:

*Behav Processes*. 2015 February ; 111: 55–59. doi:10.1016/j.beproc.2014.11.008.

## Long-term test–retest reliability of delayed reward discounting in adolescents

Andrey P. Anokhin\*, Simon Golosheykin, and Richard C. Mulligan

Washington University School of Medicine, St. Louis, MO, USA

### Abstract

Delay discounting (DD), a decline in subjective value of a reward with increasing temporal delay in receipt of that reward, is an established behavioral indicator of impulsivity. Preference for smaller-immediate over larger-delayed rewards, has been implicated in the basic neurobehavioral mechanisms of risk for addictive disorders and related externalizing psychopathology. Establishing long-term stability of DD in adolescence is a necessary step towards its validation as an intermediate phenotype, or marker of risk, in neurobiological and genetic studies. Previous studies have demonstrated moderate to high test–retest reliability of DD, however, these studies utilized adult samples and examined relatively short retest intervals. Due to continuing development of brain and behavior, stability of temporal discounting behavior in adolescence may differ from that in adulthood. Here, two cohorts of adolescents aged 16 ( $n = 126$ ) and 18 ( $n = 111$ ) were administered a computerized test of DD and re-tested two years later. DD rate showed a modest but significant decrease with age, suggesting a reduction in overall impulsivity from middle to late adolescence. Significant test–retest correlations were observed in both cohorts (.67 and .76, respectively,  $p < .001$ ) indicating longitudinal stability of individual differences in decision-making behavior during middle and late adolescence.

### Keywords

Decision making; Delay discounting; Impulsivity; Reward

## 1. Introduction

Impulsive choice is a distinct component of a broader, multifaceted “impulsivity” construct. It is commonly operationalized through delay discounting (DD) paradigms. Delay discounting refers to decrease in the subjective value of a reward with increasing temporal delay in receipt of that reward (Reynolds, 2006; Reynolds et al., 2006). Steeper DD reflects a tendency to choose smaller but immediate rewards over larger but delayed rewards. Human and animal studies have shown that DD is implicated in the basic biobehavioral mechanisms that underlie addictive behaviors and other externalizing psychopathology (reviewed in Bickel et al., 2013; Dalley et al., 2011; Mackillop, 2013; Reynolds, 2006).

Evidence that has linked risk for addiction to substances of abuse with the propensity for discounting delayed rewards suggests that DD may be an intermediate phenotype (endophenotype) for a range of disorders characterized by high levels of impulsivity, including substance use disorders (SUD) as well as attention deficit hyper-activity disorder (ADHD) and conduct disorder (CD). A focus on the relatively homogenous component processes comprising liability to psychiatric disorder may be more fruitful than examining the complex diagnostic phenotypes themselves and might facilitate identification of the neurobiological and genetic underpinnings of addiction and psychopathology.

An important prerequisite for such an intermediate phenotype is its intra-individual stability. Previous research has shown that D measures represent a stable, trait-like characteristic. Across several studies, estimates of the test–retest reliability of DD measures ranged from .55 to .90 (Baker et al., 2003; Beck and Triplett, 2009; Johnson et al., 2007; Kirby, 2009; Ohmura et al., 2006; Simpson and Vuchinich, 2000; Smits et al., 2013; Weafer et al., 2013). However several aspects of these previous studies warrant further investigation.

First, previous studies were based on data collected from adult participants. Recently, there has been increasing interest in DD as an experimental measure of impulsivity during adolescence, a period of development marked by increased risk for impulsive behaviors and substance abuse (Bava and Tapert, 2010; Crews et al 2007). However, due to continuing brain development which primarily includes the regions that are critically involved in decision making (Casey et al., 2005; Paus, 2005), this period is characterized by significant cognitive and behavioral changes. Consequently, the stability of DD behavior in adolescence may differ from that in adulthood. Therefore, data obtained in adult samples cannot be directly generalized to the adolescent population, and it is important to determine the long-term stability of DD during this period of major developmental change.

Second, most previous studies used short retest intervals (typically, a few weeks), with the notable exception of one study (Kirby, 2009) that used a retest interval of one year and showed significant test–retest stability of DD assessed using a 27-item questionnaire. Third, most laboratory studies of DD have been based on small samples ( $n = 15–33$ ) making it difficult to precisely estimate test–retest reliability. For example, with a sample size of  $n = 30$ , the 95% confidence interval of a correlation of  $r = 0.6$  would range between 0.31 and 0.79 (i.e., from “low” to “high”). One notable exception is a recent study by Weafer et al. (2013) that included a large sample of 128 young adults and found a high test–retest correlation for DD ( $r = 0.89$ ), however, mean retest interval was only about 9 days.

One of our recent studies demonstrated significant longitudinal stability and heritability of inter-temporal choice in adolescents using a real-money, single-trial delayed gratification procedure (Anokhin et al., 2011). However, little is known about developmental test–retest stability of more commonly used indices of DD that are based on varying amounts of hypothetical money and differing delays in receipt of those rewards. The purpose of the present study was to assess longitudinal stability of individual differences of DD measures in middle to late adolescence using a well-powered, population-representative sample and a version of the DD paradigm that can probe how manipulation of reward amount and temporal delays to receipt of those rewards can affect subjective value of those rewards and

decision-making. An additional aim was to determine whether DD undergoes systematic changes during middle and late adolescence.

## 2. Material and methods

### 2.1. Participants

Two longitudinal cohorts of adolescents participated in the study: 16-years-olds ( $n = 126$ , 65 females,  $M$  age  $\pm$  SD:  $16.6 \pm .26$ ) and 18-years-olds ( $n = 111$ , 59 females,  $M$  age  $\pm$  SD:  $18.7 \pm .27$ ). The sample was 84% Caucasian, 12% Black, and 4% were other minorities. Participants were administered a computerized DD test (described below) twice with an average interval between the two test administrations of approximately 2 years (age 16 cohort:  $24.7 \pm 2.2$  months; age 18 cohort:  $23.5 \pm 2.2$  months). Thus, DD data were collected at ages 16 and 18 for the younger cohort and at ages 18 and 20 for the older cohort, however, no data were available yet to compare DD at ages 16 and 20. All individuals included in the present analysis participated in a larger ongoing twin study of brain, cognition, and behavior. They were originally ascertained randomly through use of a state birth records database to ensure that the sample was maximally representative of the general population. Therefore, the present sample matched well the general population with respect to variables that potentially could bias the results, such as general intelligence (as assessed using Raven's Standard Progressive Matrices test) and socioeconomic status (parents' report). Exclusion criteria were minimal but included inability to understand task instructions, uncorrectable vision or hearing impairments, or current illicit drug or alcohol intoxication (verified by breathalyzer and urine drug test). Participants with reporting recent (one week) drug use or current intoxication or showing positive results on the urine drug test were excluded. Additionally, participants showing positive results on the alcohol breathalyzer test were excluded. Tobacco smokers were given an opportunity to smoke when they arrived for the laboratory session to minimize potential nicotine withdrawal effects, but not within the last 60 min before the start of assessments. All experiments on human subjects were conducted in accordance with the declaration of Helsinki. The study was approved by Washington University Institutional Review Board and written informed consent was obtained from all participants.

### 2.2. Discounting task

We used a computerized delay discounting task described previously (Mitchell, 1999; Mitchell and Wilson, 2010). In this task, participants were presented with a series of hypothetical choices wherein participants chose between a variable hypothetical amount of money available immediately and a delayed "standard" amount of \$100 presented at one of six possible temporal delays. Questions and response options were presented on a computer screen, and participants used the computer mouse to make their responses. On each of the 138 trials of this task, participants were presented with a question: "At this moment, what would you prefer?" Two choice options were displayed underneath this question (e.g. "\$100 in 90 days", "\$30 now"). One option was always a standard amount of \$100 available after one of six delays: 0, 7, 30, 90, 180, or 365 days. The other option was an alternative amount of money available immediately. The test questions were formed by the combination of 6 standard amounts (\$100 available at one of 6 delays) and 23 alternative amounts (ranging

from \$5 to %105 and available immediately, i.e. at 0 delay), which resulted in a total of 138 questions. One question that would require participants to choose between identical items (\$100 now or \$100 in 0 days) was omitted, thus the final set included 137 choice questions. For each question, a pair of immediate and delayed amounts was selected at random without replacement. The order in which the immediate and delayed amount was presented (first or second in the pair) was varied randomly. Participants made their choices by indicating the preferred option with a computer mouse. As response time was not limited, the duration of the task was variable but it was typically completed within 10 min. Further details of the procedure can be found elsewhere (Mitchell, 1999; Mitchell and Wilson, 2010).

### 2.3. Discounting measures

For each of the 6 delays of the standard amount (\$100), an indifference (switch) point was determined. The indifference point was defined as being midway between the smallest value of the immediate alternative that was accepted and the largest value of the immediate alternative that was rejected (Mitchell and Wilson, 2010). That is, the indifference or switch point is the value at which the participant acted as if they were indifferent to the choice between a particular amount that was available immediately and the delayed standard amount. This amount may be viewed as the “subjective value” of the delayed standard amount for that particular individual when receipt of the standard amount is delayed by a specific period of time. Indifference points can be inferred from an individual’s pattern of choices, i.e. switching from the immediate to the delayed reward and vice versa as a function of change in the amount of immediate reward. For example, if the participant’s choice switched to a delayed \$100 reward available in 180 days when immediate reward option dropped to less than \$65 or, conversely, the choice switched to immediate reward when it exceeded \$65, then, for a given individual and a given delay, \$100 to be received in 180 days is subjectively worth only \$65. Next, we built an empirical discounting curve by plotting indifference values against the corresponding delay values (Fig. 1) and computed area under the curve (AUC) as a quantitative measure of DD. Smaller AUC values indicate a steeper function and thus greater degree of temporal discounting, i.e. preference of smaller immediate rewards over larger but delayed rewards. The advantage of the AUC measure over parametric measures that can be obtained by fitting certain regression models is that it does not make any assumptions about the form of the discounting function (Myerson et al., 2001).

### 2.4. Statistical analyses

To test for significant age-related changes in the intervals from 16 to 18 years and from 18 to 20 years we used a paired *t*-test of differences between two measurements of AUC in each cohort. To examine whether age-related changes might depend on the delay we conducted a repeated-measures analysis of variance (ANOVA) of individual indifference points with factors “delay” (6 levels) and age (2 levels) in each of the cohorts. To test for gender differences with respect to DD, we applied an independent samples *t*-test of AUC values. Finally, to estimate test–retest reliability, we computed Pearson product-moment correlation between the first and the second measurement in each of the two age cohorts. Pearson correlation was chosen over intraclass correlation because only two within-subject measurements were available, furthermore, they were ordered (test and retest) and cannot be

apriori assumed to be equivalent and interchangeable. However, for comparison purpose we also computed the intraclass correlation coefficient (ICC).

Previous literature has shown that some individuals may demonstrate atypical response patterns in DD tasks suggesting misunderstanding of instructions, insufficient attention, or low motivation (e.g. Johnson & Bickel, 2008). Therefore, we recomputed test–retest correlations after excluding individuals that showed atypical choice behavior. According to the task design some items involved a choice between the standard amount of \$100 and a smaller alternative amount (i.e. \$20) at zero delay. Since choosing a smaller amount over a larger amount when both are available immediately is erratic, these questions can be considered as “control” questions indicating possible inattention and non-reflective task performance. A substantial deviation of the indifference point at 0 delay from \$100 would indicate a high frequency of erratic choices and thus questionable data. Therefore, we excluded individuals who made two or more atypical choices of this kind (Exclusion 1). Next, we excluded individuals with “nonsystematic” response patterns using a model-free algorithm proposed by Johnson and Bickel (2007), with criteria modified to fit the shorter time span used in the present study. Data were categorized as nonsystematic if any indifference point was greater than the preceding point by at least 20% (suggesting a significantly non-monotonous function) and/or if the last indifference point was not less than the first indifference point by at least 5% (suggesting that delay had no effect on reward value). The combination of the Exclusion 1 criterion and the above criteria for nonsystematic data resulted in a more stringent set of exclusion criteria (Exclusion 2). To examine whether and how the exclusion of questionable data affected test–retest reliability, we computed test–retest correlations under three conditions: no exclusions (all data used), Exclusion 1, and Exclusion 2.

### 3. Results

Consistent with previous research, empirical discounting curves (Fig. 1) showed the expected decrease of subjective value of reward with increasing time to its receipt. The degree of DD as indicated by the AUC measure decreased significantly from age 16 to age 18 (two-tailed paired  $t$ -test:  $t = -3.33$ ,  $df = 115$ ,  $p = .001$ ), although the effect size of this change was quite modest ( $d = .25$ ). However, no significant change was observed from age 18 to age 20 (two-tailed paired  $t$ -test:  $t = -1.52$ ,  $df = 103$ ,  $p = .13$ ). There were no significant gender differences in the AUC measure of DD at any of the ages studied (Fig. 2). Mean AUC values are presented in Table 1.

To examine whether the effects of age depend on the delay to the larger reward and whether these effects are modulated by gender, we have conducted a mixed-design ANOVA in each cohort with indifference point (subjective value of delayed reward) as dependent variable, delay and age as within-subject factors and gender as between-subject factor. In the younger cohort (age 16 at the first assessment), this analysis revealed a highly significant main effect of delay on the subjective value ( $F_{5,121} = 65.3$ ,  $p < 0.001$ ) indicating a strong discounting effect, a significant main effect of age ( $F_{1,125} = 6.68$ ,  $p = 0.011$ ), and significant delay by age interaction ( $F_{15,121} = 2.63$ ,  $p = 0.027$ ) indicating that age differences are present at longer (>90 days) but not shorter delays. However, no main or interaction effects of gender

were observed. Similar analysis in the older cohort (at ages 18 and 20) also showed a highly significant main effect of delay on the subjective value ( $F_{5,106} = 54.3, p < 0.001$ ), whereas neither the main effect of age, nor delay by age interaction were significant. However, a significant age by gender interaction emerged in this cohort ( $F_{5,102} = 10.7, p < 0.01$ ), indicating that subjective values slightly increased in females but decreased in males from age 18 to age 20.

Most importantly, DD showed highly significant test–retest correlations in both age cohorts indicating high long-term stability (Table 2). The exclusion of atypical test performers did not affect these results in a substantial way. Intraclass test–retest correlations were very close to Pearson correlations (up to the second decimal point), except correlations between ages 18 and 20 computed in the entire sample and under Exclusion 2 condition, where ICC were slightly larger than Pearson correlations (entire sample: 0.788 and 0.769, respectively; Exclusion 2: 0.780 and 0.764, respectively).

#### 4. Discussion

The present study sought to estimate long-term stability of individual differences in delay discounting (DD) during adolescence. The findings indicate moderate to high test–retest reliability of DD over two 2-year intervals, 16–18 and 18–20 years of age. To the best of our knowledge, the present study is the first to demonstrate stability of DD behavior over two different two year periods. These results are broadly consistent with previous studies showing moderate to high test–retest reliability of DD measures obtained using different procedures such as questionnaires, real-money delay gratification tests, and laboratory paradigms that adjust amount of reward and delay to its receipt. However, several important aspects of the present study make it distinct from previous research and significantly extend the current knowledge about long-term stability of individual differences in temporal discounting behavior.

First, the present study extends evidence obtained on adult samples by demonstrating individual stability of DD in adolescence, a period characterized by significant developmental changes in brain and behavior that could potentially result in reduced stability of individually-specific patterns of choice behavior. Nevertheless, our data show that despite the presence of systematic age-related changes from age 16 to 18, individual differences showed considerable stability over this period. Another important feature of the present study is that the sample is representative of the general population. Many previous studies have utilized samples of university students that may be biased toward lower discounting rates and show restricted range of variance. It is not unreasonable to expect that individuals such as college students who make a life choice to pursue a remote goal (larger but delayed reward) and forego smaller but sooner rewards would show lower propensity to discount delayed reward. Finally, by using relatively large samples, our study provides a more accurate and confident estimate of the test–retest reliability associated with DD than in most previous studies.

Another important finding is a significant decrease in DD with age between 16 and 18 years, suggesting that age-related decline in the propensity to impulsive choice demonstrated in our



previous study of younger adolescents (ages 12–14) using a different DD paradigm (Anokhin et al., 2011) continues into late adolescence. However, the lack of significant change from age 18 to age 20 in the present study suggests that most of these age-related changes occur before age 18. A significant age by gender interaction suggests that the direction of age-related changes may be different in males and females. Taken together, these findings suggest a rather complex age-related dynamics of DD, warranting a more thorough examination of longitudinal changes in DD over larger time intervals. The ability of the present study to address developmental changes is limited because currently available data permit to examine age-related changes in the same individuals over a two-year period only.

It is important to note that, despite a systematic age-related shift in the absolute value of DD, individual differences, i.e. the relative ranking of individuals within the sample remains relatively stable. This long-term stability of individual differences is important because it makes DD a suitable measure for prospective longitudinal designs. For example, DD can be used as potential prospective predictor of problem behaviors such as substance abuse or pathological gambling. Conversely, DD can be used for the investigation of the effects of certain environmental insults on cognitive function and behavior, such as heavy alcohol exposure in underage drinkers.

Finally, a significant delay by age interaction suggests that age-related decrease in DD primarily occurs due to selective increase of the subjective value of rewards at larger delays, suggesting an expansion of the “time horizon” for decision making. A possible mechanism underlying this age-related increase in subjective valuation of delayed rewards could be improved ability to represent the future. It is reasonable to suggest cognitive processes underlying choice behavior involve a “competition” between mental and neural representation of alternative options. A stronger representation of delayed outcomes would result in their greater influence on decision making. The ability to represent distant future increases with age and is likely related to the maturation of the prefrontal brain regions that are known to support prognosis and planning.

Future developmental studies combining behavioral and imaging measures should determine whether developmental dynamics in choice behavior can be directly linked to maturation of the underlying neural substrates. Behavioral neuroscience research using animal models (Dalley et al., 2008) and neuroimaging studies in humans have implicated a number of brain regions in intertemporal choice behavior, including regions that play a key role in reward processing such as ventral striatum and regions that are involved in planning and prospective thought including areas of the prefrontal and posterior cingulate cortex (Carter et al., 2010; Luo et al., 2012; McClure et al., 2004; Peters and Buchel, 2011). Future research using neuroimaging techniques should examine developmental changes and stability of the structure and function of specific neuroanatomical and neurochemical substrates underlying stable individual differences in intertemporal choice. Given the evidence that regions involved in cognitive control and inhibition of impulsive choices show protracted maturation in contrast to the regions supporting reward processing (Casey et al., 2005; Paus, 2005), we expect that age-related shifts in DD during the transition from

adolescence to young adulthood will be primarily related to developmental changes in the prefrontal cortex, rather than early maturing subcortical reward-related structures.

Certain limitations of the present study are worth noting. First, the present study employed a procedure involving hypothetical money. This could have potentially affected the subjective value of the rewards and biased the results. However, many previous studies have shown high correlations between values associated with hypothetical choices and real rewards (Johnson and Bickel, 2002; Lagorio and Madden, 2005; Madden et al., 2003; Madden et al., 2004). Second, the present results should be generalized to other DD paradigms with caution because evidence for cross-test correlations is scarce. In particular, there is evidence that in DD paradigms that adjust amount of reward and delay to its receipt such as the one used here, the amount of the standard reward is inversely related to the degree of discounting at the group level. Whether this might affect test–retest reliability of DD needs to be clarified in future studies. Finally, it is important to note that the current study has limited ability for drawing definitive conclusions about developmental changes in DD, in particular, its power to detect modest changes from age 18 to 20 may be limited.

In summary, the present results suggest that a standard laboratory choice procedure used in DD research provides a reliable measures of impulsivity that can be used in genetic and clinical studies of adolescent behavior.

## Acknowledgements

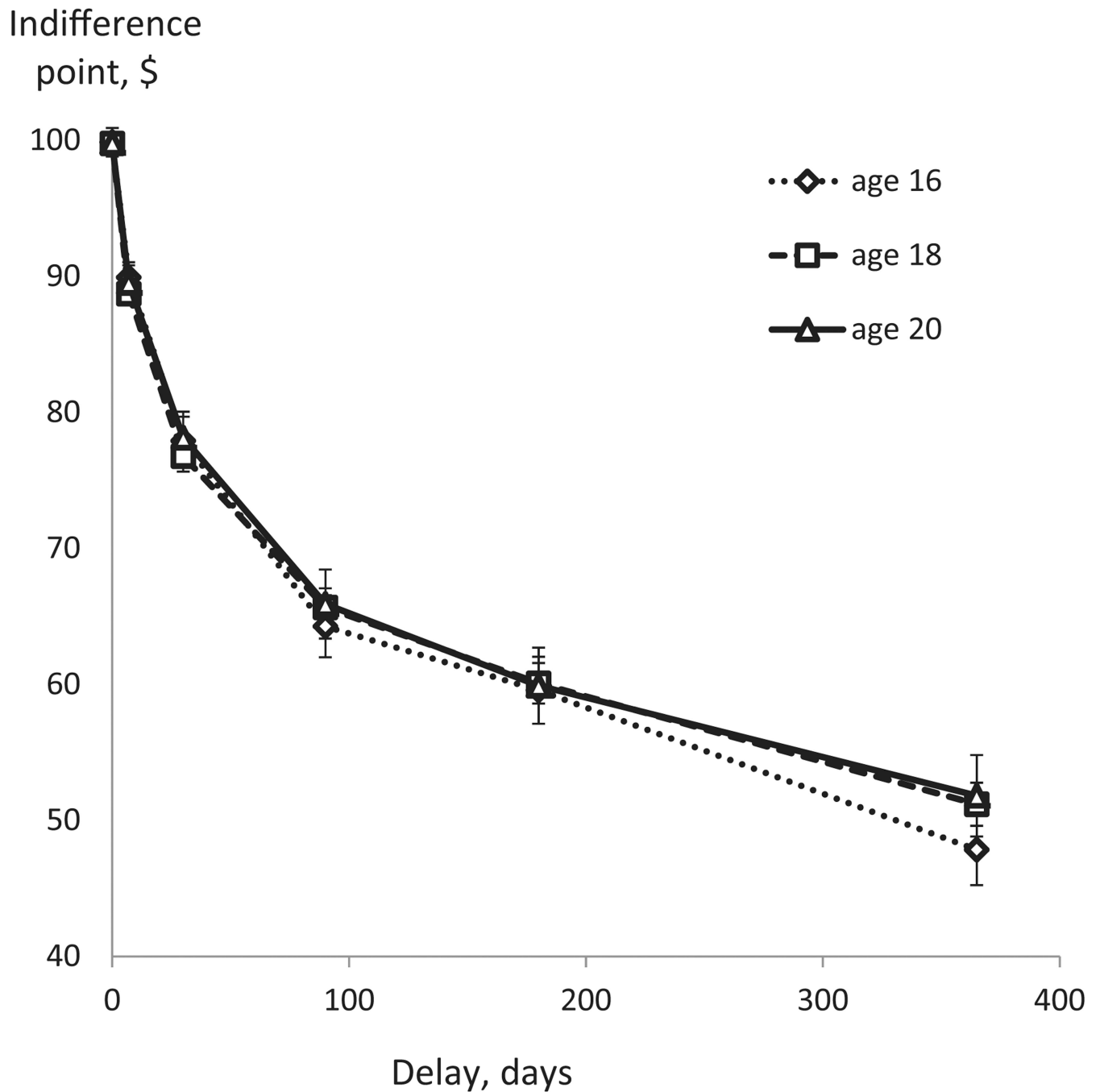
This work was supported by grants DA018899 and DA027096 from the National Institutes of Health (NIH) to A.A. The authors are grateful to Dr. Suzanne Mitchell for her advice on implementation of a computerized test of delay discounting. The authors also acknowledge the generous giving of time and effort by the study participants.

## References

- Anokhin AP, Golosheykin S, Grant JD, Heath AC. Heritability of delay discounting in adolescence: a longitudinal twin study. *Behav. Genet.* 2011; 41:175–183. [PubMed: 20700643]
- Baker F, Johnson MW, Bickel WK. Delay discounting in current and never-before cigarette smokers: similarities and differences across commodity, sign, and magnitude. *J. Abnorm. Psychol.* 2003; 112:382–392. [PubMed: 12943017]
- Bava S, Tapert SF. Adolescent brain development and the risk for alcohol and other drug problems. *Neuropsychol. Rev.* 2010; 20:398–413. [PubMed: 20953990]
- Beck RC, Triplett MF. Test-retest reliability of a group-administered paper-pencil measure of delay discounting. *Exp. Clin. Psychopharmacol.* 2009; 17:345–355. [PubMed: 19803634]
- Bickel WK, Koffarnus MN, Moody L, Wilson AG. The behavioral-and neuro-economic process of temporal discounting: a candidate behavioral marker of addiction. *Neuropharmacology.* 2013; 76:518–527. [PubMed: 23806805]
- Carter RM, Meyer JR, Huettel SA. Functional neuroimaging of intertemporal choice models: a review. *J. Neurosci. Psychol. Econ.* 2010; 3:27–45.
- Casey BJ, Tottenham N, Liston C, Durston S. Imaging the developing brain: what have we learned about cognitive development? *Trends Cogn. Sci.* 2005; 9:104–110. [PubMed: 15737818]
- Crews F, He J, Hodge C. Adolescent cortical development: a critical period of vulnerability for addiction. *Pharmacol. Biochem. Behav.* 2007; 86:189–199. [PubMed: 17222895]
- Dalley JW, Everitt BJ, Robbins TW. Impulsivity, compulsivity, and top-down cognitive control. *Neuron.* 2011; 69:680–694. [PubMed: 21338879]

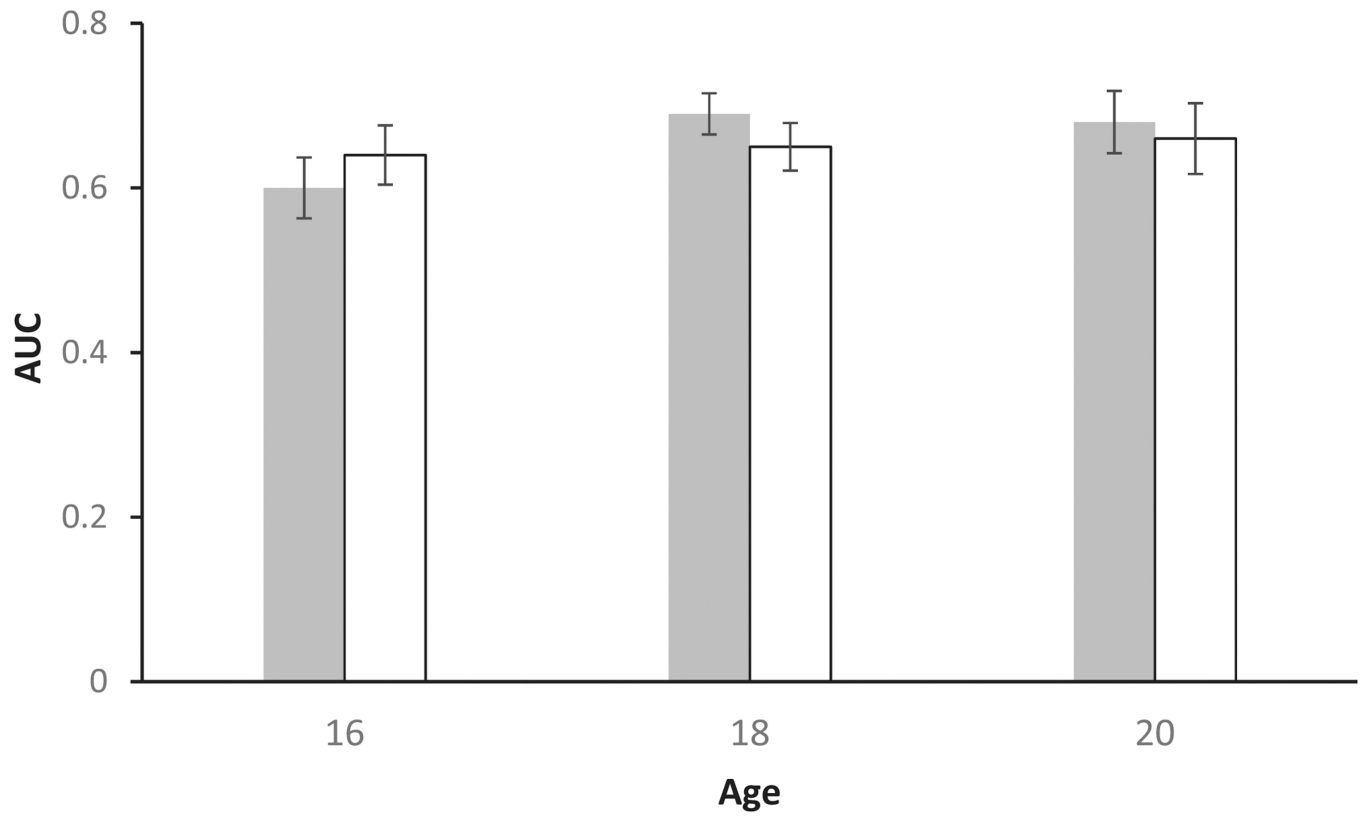


- Dalley JW, Mar AC, Economidou D, Robbins TW. Neurobehavioral mechanisms of impulsivity: fronto-striatal systems and functional neurochemistry. *Pharmacol. Biochem. Behav.* 2008; 90:250–260. [PubMed: 18272211]
- Johnson MW, Bickel WK. Within-subject comparison of real and hypothetical money rewards in delay discounting. *J. Exp. Anal. Behav.* 2002; 77:129–146. [PubMed: 11936247]
- Johnson MW, Bickel WK, Baker F. Moderate drug use and delay discounting: a comparison of heavy, light, and never smokers. *Exp. Clin. Psychopharmacol.* 2007; 15:187–194. [PubMed: 17469942]
- Kirby KN. One-year temporal stability of delay-discount rates. *Psychon. Bull. Rev.* 2009; 16:457–462. [PubMed: 19451368]
- Lagorio CH, Madden GJ. Delay discounting of real and hypothetical rewards III: steady-state assessments, forced-choice trials, and all real rewards. *Behav. Processes.* 2005; 69:173–187. [PubMed: 15845306]
- Luo S, Ainslie G, Pollini D, Giragosian L, Monterosso JR. Moderators of the association between brain activation and farsighted choice. *NeuroImage.* 2012; 59:1469–1477. [PubMed: 21856429]
- Mackillop J. Integrating behavioral economics and behavioral genetics: delayed reward discounting as an endophenotype for addictive disorders. *J. Exp. Anal. Behav.* 2013; 99:14–31. [PubMed: 23344986]
- Madden GJ, Begotka AM, Raiff BR, Kastern LL. Delay discounting of real and hypothetical rewards. *Exp. Clin. Psychopharmacol.* 2003; 11:139–145. [PubMed: 12755458]
- Madden GJ, Raiff BR, Lagorio CH, Begotka AM, Mueller AM, Hehli DJ, Wegener AA. Delay discounting of potentially real and hypothetical rewards: II Between- and within-subject comparisons. *Exp. Clin. Psychopharmacol.* 2004; 12:251–261. [PubMed: 15571442]
- McClure SM, Laibson DI, Loewenstein G, Cohen JD. Separate neural systems value immediate and delayed monetary rewards. *Science.* 2004; 306:503–507. [PubMed: 15486304]
- Mitchell SH. Measures of impulsivity in cigarette smokers and non-smokers. *Psychopharmacology (Berl).* 1999; 146:455–464. [PubMed: 10550496]
- Mitchell SH, Wilson VB. The subjective value of delayed and probabilistic outcomes: outcome size matters for gains but not for losses. *Behav. Processes.* 2010; 83:36–40. [PubMed: 19766702]
- Myerson J, Green L, Warusawitharana M. Area under the curve as a measure of discounting. *J. Exp. Anal. Behav.* 2001; 76:235–243. [PubMed: 11599641]
- Ohmura Y, Takahashi T, Kitamura N, Wehr P. Three-month stability of delay and probability discounting measures. *Exp. Clin. Psychopharmacol.* 2006; 14:318–328. [PubMed: 16893275]
- Paus T. Mapping brain maturation and cognitive development during adolescence. *Trends Cogn. Sci.* 2005; 9:60–68. [PubMed: 15668098]
- Peters J, Buchel C. The neural mechanisms of inter-temporal decision-making: understanding variability. *Trends Cogn. Sci.* 2011; 15:227–239. [PubMed: 21497544]
- Reynolds B. A review of delay-discounting research with humans: relations to drug use and gambling. *Behav. Pharmacol.* 2006; 17:651–667. [PubMed: 17110792]
- Reynolds B, Ortengren A, Richards JB, de Wit H. Dimensions of impulsive behavior: Personality and behavioral measures. *Personality and Individual Differences.* 2006; 40:305–315.
- Simpson CA, Vuchinich RE. Reliability of a measure of temporal discounting. *Psychol. Rec.* 2000; 50:3–16.
- Smits RR, Stein JS, Johnson PS, Odum AL, Madden GJ. Test-retest reliability and construct validity of the experiential discounting task. *Exp. Clin. Psychopharmacol.* 2013; 21:155–163. [PubMed: 23421359]
- Weafer J, Baggott MJ, de Wit H. Test-retest reliability of behavioral measures of impulsive choice impulsive action, and inattention. *Exp. Clin. Psychopharmacol.* 2013; 21:475–481. [PubMed: 24099351]



**Figure 1.**

Discounting of \$100 amount as a function of delay to its receipt. Average data at each age (16, 18, 20) are presented. Horizontal axis: delay in days; line markers indicate delays of 0, 7, 30, 90, 180, and 365 days. Vertical axis: Indifference points, or subjective value of the \$100 amount at each of the delays.



**Figure 2.** Means ( $\pm$ s.e.) of area under the curve (AUC) by age and gender. Note that smaller AUC values indicate steeper discounting and vice versa.

**Table 1**

Test–retest reliability of individual differences in delay discounting.

Test–retest interval (ages in years)	Sub-sample analyzed	<i>r</i>	<i>n</i>	95% CI
16–18	All	.65	126	.54–.74
	Exclusion 1	.67	116	.55–.76
	Exclusion 2	.72	83	.60–.81
18–20	All	.77	111	.68–.84
	Exclusion 1	.76	104	.67–.83
	Exclusion 2	.76	78	.65–.84

Notes: All test–retest correlations in the Table are highly significant ( $p < .001$ ); Exclusion 1: Individuals with erratic responses are excluded (e.g. preferring a smaller amount over a larger amount at zero delay); Exclusion 2: Exclusion 1 plus non-systematic response pattern excluded according to modified criteria from Johnson and Bickel (2008) including non-monotonous decline of the indifference value with increasing delay or insensitivity to delay (see text for details).