Journal of Economic Psychology 46 (2015) 89–97

Contents lists available at ScienceDirect

Journal of Economic Psychology

journal homepage: www.elsevier.com/locate/joep

A simulated financial savings task for studying consumption and retirement decision making

Derek J. Koehler *, Jesse Langstaff, Wu-Qi Liu

University of Waterloo, Canada

ARTICLE INFO

Article history:
Received 30 August 2013
Received in revised form 17 November 2014
Accepted 22 December 2014
Available online 30 December 2014

JEL classification:
D91
J26

PsycINFO Classification:
2340
3900

Keywords:
Financial decision making
Life-cycle model of consumption
Intertemporal choice
Time preference
Risk attitudes
Retirement savings

ABSTRACT

We describe a simulated financial decision making task that requires the participant to make decisions, over the course of a life cycle, regarding how much of their income to consume immediately and how much to save for a later retirement phase in which no income will be generated. The savings task was developed to be readily understood and performed online by members of a diverse participant population. A preliminary study ($N = 165$) involving such a population was conducted in which performance on the savings task was observed as length of retirement phase and the presence of a consumption-smoothing goal were experimentally manipulated. Results suggested that most participants understood the task and responded sensibly; for example, they saved a greater portion of current income for later consumption when faced with a long, compared to a short, retirement phase. Responsiveness of saving levels during the task to retirement length was found to be correlated with an independent measure of participants' financial risk attitudes. Consumption smoothing during the task was found to be correlated with a measure of individual differences among participants in temporal delay discounting. Compared to experimental savings tasks developed in previous research, the task described here may offer some practical advantages in requiring less extensive participant instructions, providing a user-friendly graphical interface, being readily performed online, and potentially being more accessible as a consequence to participants with limited education or financial sophistication.

1. Introduction

Deciding how much money to save, whether to buffer against unexpected expenses or for later consumption during retirement, can be difficult. Economists and psychologists, among other social scientists, are interested in how, and how well, people make such decisions. One methodological challenge is that real-life financial decisions can be investigated observationally but not, generally speaking, experimentally, because key variables such as income, expenses, and length of working and retirement periods are not readily subject to experimental manipulation. Experiments investigating decisions in simulated financial environments, then, may serve as an informative complement to observational research investigating corresponding decisions in real financial environments.

* Corresponding author at: Department of Psychology, University of Waterloo, 200 University Ave. W., Waterloo, Ontario N2L 3G1, Canada. Tel.: +1 (519) 888 4567x35013.

E-mail address: dkoehler@uwaterloo.ca (D.J. Koehler).

http://dx.doi.org/10.1016/j.joep.2014.12.004

By now there exists a small body of such experiments (e.g., Ballinger, Palumbo, & Wilcox, 2003; Brown, Chua, & Camerer, 2009; Carbone & Hey, 2004; Fehr & Zych, 1998; Hey & Dardanoni, 1988). Researchers have noted that life-cycle models of consumption and savings in economics are typically optimized using computationally intensive methods, naturally raising the question of how closely human performance approximates the optimal benchmark (Benartzi & Thaler, 2007; Brown et al., 2009; Winter, Schlafmann, & Rodepeter, 2012). Indeed, a number of experiments on savings decisions have revealed suboptimal performance, in some cases arising from systematic departures in how people make savings decisions relative to the optimized savings model, such as overconsumption and undersaving (Ballinger et al., 2003; Brown et al., 2009; Carbone & Hey, 2004; Fehr & Zych, 1998; Hey & Dardanoni, 1988). These results suggest that savings experiments can provide insights into corresponding behavioral shortcomings in real-life financial decisions.

A major complication in identifying suboptimal financial decision making, whether in the field or in the lab, is heterogeneity in the utility derived from consumption and savings due, for instance, to individual differences in risk attitude and time preference. Lab experimenters have developed a convenient methodological solution in the form of induced utility, in which simulated consumption during the savings task yields utility "points" (or experimental currency, which is later converted into a payment) according to a function that is fixed for all participants. In theory, participant preferences are experimentally equated via the method of induced utility, and performance can then be evaluated against a common optimal benchmark.

The goal of researchers conducting savings experiments, to investigate the role of various key variables in the life-cycle model (e.g., income shocks, habit formation, longevity risk, etc.), coupled with the use of the induced utility methodology, has led to the construction of experimental savings tasks that are, on the one hand, quite sophisticated but, on the other, potentially confusing and difficult to explain to participants. Instructions can run to many pages (or require extensive one-on-one instruction from the experimenter), involving tables and the use of calculators or spreadsheets for relevant utility calculations (e.g., Ballinger, Hudson, Karkoviata, & Wilcox, 2011; Brown et al., 2009; Hey & Dardanoni, 1988). Due to their complexity, furthermore, these tasks can require lengthy experimental sessions (e.g., 45–60 min of training followed by a main task taking 75–90 min in Fehr & Zych, 1998; and approximately 90 min for training and task completion in Brown et al., 2009). In some settings, such as one-on-one in-lab sessions involving students from selective universities, use of a task with intensive computational and time demands is not necessarily a barrier to collection of instructive, meaningful data. In other settings, however, such demands may be prohibitive.

The goal of the present research was to develop a simplified savings task that requires less time, detailed instructions, and participant education to complete, and that as a consequence can be readily administered to large numbers of demographically diverse participants in an online setting (e.g., Amazon Mechanical Turk). One step we took to simplify our savings task, relative to those used in previous savings experiments, was to reduce some forms of uncertainty (e.g., income shocks), to eliminate certain complicating factors (e.g., habit formation), and to largely remove the need for observational learning (e.g., to infer relevant probability distributions). Another, possibly more radical, step we took was to abandon the use of an induced utility function. Instead, participants were either left to rely entirely on their own exogenous preferences, facets of which would presumably be revealed by their choices during the savings task, or at most were given only the broad, qualitative goal of smoothing their simulated lifetime consumption. Because we did not try to eliminate, using an induced utility function, the impact on task performance of heterogeneity among individuals in risk attitudes and time preference, we instead attempted to measure these constructs with scales administered following the savings task. Indeed, one of our research questions was whether individual differences in risk attitude and time preference would correlate with performance measures derived from the savings task.

Below, we describe the savings task we developed, and then report the results of a preliminary study in which we examined how participants completing the task responded to changes in the anticipated length of the "retirement" phase for which they were saving. Of particular interest was the extent to which participants attempted to smooth their discretionary income consumption over the simulated life cycle. To this end, we compared the spontaneous behavior of one group completing the task to another group that was explicitly encouraged to engage in consumption smoothing. The study was largely exploratory, with the main goal to examine whether participants could understand the savings task and perform it in a sensible manner that yielded meaningful data.

2. Savings task

Participants played several rounds ("lives") of a savings task, referred to as a "Game of Life," that would require them to make decisions about how much to spend and save over simulated life-cycles of work and retirement. The complete set of task instructions is included in Appendix A of the supplementary materials.[1]

Each round (simulated life) consisted of 24 periods, divided into work and retirement phases. Income was received during periods in the work phase but not in the retirement phase. Participants were informed, at the outset of the round, how many periods there were to be (i.e., there was no uncertainty regarding how long the simulated life would last), how much income each work period would bring, and at which period they would retire and no longer receive any income. This information was presented graphically in a grid (see Fig. 1 for a screenshot). Income varied from 10 to 32 units per period, and increased

[1] A working version of the savings task can be accessed at http://arts.uwaterloo.ca/~dkoehler/savingsgame. Please contact the authors for a copy of the program code, which was written in PHP and JavaScript.

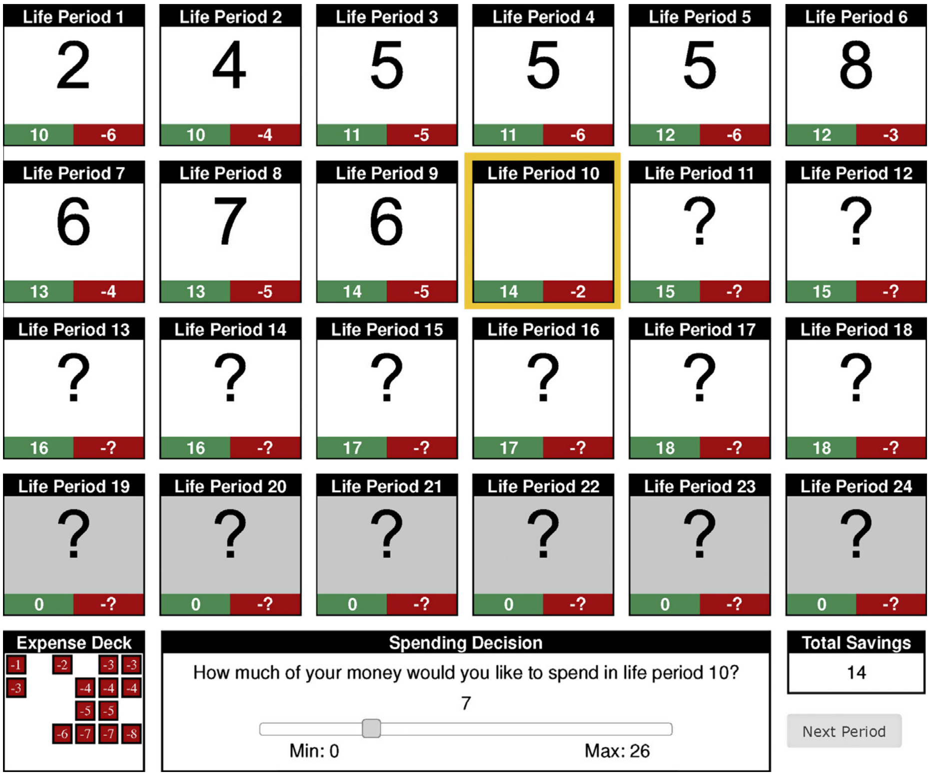

Fig. 1. Screenshot showing an example round of the savings task in progress. Current period is highlighted in yellow. Income for each period is shown on a green background. Expenses are shown on a red background, with uncertain expenses in future periods labeled with a question mark. Discretionary spending (consumption) in past rounds is shown in black. Retirement periods have a gray background and indicate income as 0. Expense deck in the lower-left corner shows remaining expense levels to be encountered in future periods in the round. Total savings up to the current period are indicated in the lower-right corner. On each round, the participant indicated how much they wished to spend (after meeting the period's expenses) using the slider at the bottom center, which indicated at its upper end the maximum amount available to be spent. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

incrementally until retirement. The instructions encouraged participants to think of each unit as corresponding to approximately 10 thousand dollars, and each period as corresponding to approximately 3 years. Total income per round was 252 units.

Each period within a round carried associated expenses that had to be paid in that period. The magnitude of expenses was a random variable whose value was revealed at the beginning of each period, ranging from 1 to 8 units per period. Expenses for each period were drawn at random, without replacement, from an "expense deck" that was depicted and updated in a corner of the screen. The expense deck, then, allowed participants to calculate the total expenses they would incur over the round (108 units), as well as to see at any point in the round an updated list of the expenses they were still to incur in future periods. The only uncertainty regarded the specific period in which each remaining expense value from the deck would be encountered.

In each period, participants made a discretionary spending (consumption) decision, using a slider to indicate the amount they wished to spend, after meeting their expenses for the period. The amount available for spending, which was indicated at the upper end of the slider, was calculated as savings to date plus current period income less current period expenses; participants could not spend more than they had earned or saved. Any money not spent in the current period was retained in a no-interest savings account, the balance of which was continually displayed.

The grid displaying the participant's progress through the round highlighted the current period for which a spending decision was to be made. The grid indicated income, expenses, and spending for each previously-completed period, as well as income to be earned in future periods. Retirement periods, listed as having zero income, were highlighted in gray to emphasize the point in the round at which the participant would no longer earn an income but would still have to meet expenses. As noted above, participants could not borrow to increase consumption (i.e., spend more than they had available from previous savings and current income); they could, however, go into debt if their expenses exceeded what was available from past savings and current-period earnings. Spending (consumption) was necessarily constrained to be zero in such circumstances.

After being presented with the instructions, participants were given a multiple-choice comprehension quiz regarding various aspects of the game and its graphical display. The quiz showed a screenshot of a round in progress and asked participants to identify the current period number, income for that period, expenses for that period, savings to date, total amount

Table 1
Income per period in the short and long retirement conditions. Retirement periods (with zero income) are listed in bold.

Period	Income (units)	
	Short retirement	Long retirement
1	10	10
2	10	12
3	11	14
4	11	16
5	12	18
6	12	20
7	13	22
8	13	24
9	14	26
10	14	28
11	15	30
12	15	32
13	16	**0**
14	16	**0**
15	17	**0**
16	17	**0**
17	18	**0**
18	18	**0**
19	**0**	**0**
20	**0**	**0**
21	**0**	**0**
22	**0**	**0**
23	**0**	**0**
24	**0**	**0**

available for spending (shown on the slider), how much had been spent in a particular previous period, how many times a particular expense level was still to be encountered in future periods (from the expense deck), and in which period retirement would begin. The answer to each question was to be selected from four possible options. Participants could click on a "hint" next to any question for which they needed help. Participants could not move on to play the game itself until they answered all 8 comprehension questions correctly.

Participants then played 8 rounds of the game. The key factor that differed between rounds was the length of the retirement phase. In rounds with a short retirement phase, the final 6 periods of the 24 were retirement periods in which no income was earned; in rounds with a long retirement phase, the final 12 periods earned no income. Total income earned over the working periods was equated across short and long retirement conditions (see Table 1). In addition, expense variance differed between rounds, such that total expenses over the round were equated but period-to-period variance in expenses was either relatively high or relatively low.[2] Crossing the two factors (retirement length and expense variance) yields four unique round types. Each participant played through the four round types once in a randomized order, and then played those same four rounds again in the same order in which they were encountered initially.

Two primary dependent measures were obtained from each round. *Retirement savings* was defined as the amount held in savings at the point of retirement. *Spending variability* was measured by the standard deviation of discretionary spending over the 24 periods in the round. (We used the standard deviation rather than the variance as it provided a less heavily skewed measure of spending variability.) Every participant had a score on these two dependent measures for each of the four unique round types, averaged over the two repetitions of each.[3] In addition, web page submission times were logged to provide an estimate of time spent completing the savings task, as well as time spent reading the instructions.

Participants were randomly assigned either to a goal-present condition or a goal-absent condition, which differed only in one paragraph of the task instructions. In the goal-absent condition, participants were instructed to play the game in a manner that would lead to a "happy and satisfying life." Participants in the goal-present condition received identical instructions, with one addition: to set a goal of achieving a smooth level of consumption over periods in each round (life-cycle). The goal was justified by noting that, "research has shown that people who were able to maintain a fixed level of consumption or spending across the lifetime, smoothing over unpredictable expenses and the loss of income in retirement, tend to be happiest and most satisfied with their lives."

[2] Specifically, expenses per period followed either a uniform (high variance condition) or triangular (low variance condition) distribution over the interval from 1 to 8 units. This factor proved to have little impact on savings task performance and so receives minimal discussion here.

[3] Data were also analyzed without averaging across the two repeated rounds of a particular type, with block (first and second instance of a round type) as an additional within-subject factor. The only significant effects of this factor, which did not qualify the results reported for the averaged data in the main text, was for the savings measure: Savings were higher on average in the second block than the first, but only on long-retirement rounds ($M = 130$ in the first block and $M = 140$ in the second). No other main effects or interactions with the block variable were statistically significant, so this factor is not considered further here.

3. Other measures

After completing the savings task, participants completed standardized measures of risk attitude and time preference. A few relevant demographic variables were also measured.

3.1. Risk attitude

Risk attitudes were measured using the 30-item revised Domain Specific Risk Attitude (DOSPERT) scale, which measures risk attitude across a wide range of domains (Blais & Weber, 2006; Weber, Blais, & Betz, 2002).[4] An example item is "betting a day's income at the horse races." Respondents were asked to rate how likely they were to engage in the activity or behavior, on a 7-point scale ranging from 1 (extremely unlikely) to 7 (extremely likely).[5] The ratings were summed across all items to obtain a risk-taking score, with higher scores indicating a stronger preference for risk. In addition to the domain general risk-taking score, we also obtained a score specific to the financial domain by summing ratings from the subset of six DOSPERT items pertaining specifically to financial decision making. In our sample, reliability of the DOSPERT was reasonably high both for the full scale ($\alpha = .86$) and for the financial domain subscore ($\alpha = .74$).

3.2. Time preference

Participants completed 2 measures assessing their impulsivity and time preference for money. The first, domain-general measure was the Barratt Impulsiveness Scale (BIS-11; Patton, Stanford, & Barratt, 1995). The scale consisted of 30 items and ratings were made using a 4-point scale ranging from 1 (Rarely/Never) to 4 (Almost Always/Always). An example of an item would be "I do things without thinking." Ratings were summed over items to obtain an impulsivity score. In our sample, reliability of items on the BIS-11 was high ($\alpha = .88$).

The second time preference measure was obtained using the Kirby and Marakovic (1996) delay-discounting scale. Participants were presented with 27 decisions in which they chose between an immediate reward and a delayed reward. An example would be "$11 today or $30 in 7 days". Kirby and colleagues offer a scoring method that estimates the respondent's discounting rate, but we found the resulting measure to be problematic due to skewness and to inconsistencies in choices across items. For our purposes, which did not require estimation of a model-based discount rate, we used a simpler model-free measure of how many times the smaller-sooner reward was chosen over the larger-later one across the 27 decisions (following Ersner-Hershfield, Garton, Ballard, Samanez-Larkin, & Knutson, 2009). Reliability of responses to the Kirby scale, in our sample, was high ($\alpha = .91$).

3.3. Demographics

A final set of questions asked for the participant's age, gender, highest completed education level, annual personal income, and amounts of assets and debts in different categories such as bank accounts, investments, pension plan, loans, and credit card balance. Participants reported their age, education, and income by selecting the appropriate bracket from a list. Many participants did not complete the items asking about specific assets and debts, however, possibly because they were perceived as overly intrusive or – more likely given that most people were willing to disclose their income bracket – because they did not have the numbers readily available to report. In the analyses below, we look only at the age, gender, education, and income variables.

4. Participants

Participants were recruited through Amazon Mechanical Turk (AMT) and paid $2 for their participation. Research suggests that AMT has certain advantages over the traditional laboratory setting and psychology student participant pool (Horton, Rand, & Zeckhauser, 2011; Mason & Suri, 2012; Paolacci, Chandler, & Ipeirotis, 2010; Suri & Watts, 2011).

Of the 165 participants, data from 16 participants were excluded, leaving a final sample of 149. A participant met the criterion for exclusion if he or she left one or more pages of the online questionnaire entirely uncompleted. Among the 149 remaining participants, missing responses on the DOSPERT, Kirby, and BIS-11 measures constituted less than 0.1% of the total number of responses (11 out of 12,963). We reconstructed these missing data points using an expectation–maximization algorithm. There was no missing data from the savings task itself. For analyses where values for relevant demographic variables were not provided, observations with missing data were omitted from the analysis. Out of the final sample of 149 participants, 76 were female and 71 were male (2 participants did not disclose their gender). Based on interpolation of the selected bracket ranges, mean age was 29 years, mean income was $31,920 (among the 138 of the 149 participants willing to disclose it), and mean education level was a Bachelor's degree.

[4] An additional single-item measure of risk attitude, from Eckel and Grossman (2002) and Eckel and Grossman (2008), was also collected but did not correlate with any task performance variables, possibly due to its lower reliability compared to the DOSPERT, and so is not discussed further here.

[5] The DOSPERT includes a second set of ratings for each of the target activities, in which participants judge the riskiness of the activity rather than whether they personally would engage in that activity, but we did not collect those ratings in this study.

5. Results

As noted, we focused on two dependent measures from the savings task: retirement savings and variability in discretionary spending (consumption) over periods. With respect to the first dependent measure, we examine whether participants were able to save sufficiently for retirement, and in particular whether they made sufficient adjustments in saving in response to the manipulation of retirement length. With respect to the second dependent measure, we examine whether participants smoothed their consumption over periods, and whether assigning an explicit goal of smoothing consumption influenced observed variability in spending across periods.

5.1. Retirement savings

Collapsed across all rounds, mean savings at retirement was 117 units. One way of assessing the reasonableness of this level of savings is to consider that retirement constituted, on average, 37.5% of the periods in a given round (25% in short-retirement rounds and 50% in long-retirement rounds). Therefore one benchmark for average savings rates across conditions (e.g., if the individual is aspiring to maintain spending at pre-retirement levels) is 37.5% of the 252 units of income received per round, or 94.5 units. Approximately 85% of the participants had average retirement savings (across the 8 rounds) of at least 94.5 units. It is worth noting that, in every round, income increased across periods up until retirement. One possible reason for the observed level of savings being higher, on average, than what is needed for replacement of average income across the round, then, is that participants were hoping to maintain an increasing consumption profile that reflected their rising income across periods.

A mixed-model analysis of variance (ANOVA) conducted on the retirement savings measure, with retirement length (short versus long) and expense variability (low versus high) as within-subject variables and goal (absent versus present) as a between-subject factor, revealed a main effect of retirement length, $F(1,147) = 379$, $\eta_p^2 = 0.72$, $p < .001$, and no other statistically significant main effects or interactions. Participants saved more when facing a long retirement ($M = 135$ units) than they did when facing a short retirement ($M = 98$ units). In anticipation of a long retirement phase constituting 50% of the periods in the round, participants saved approximately half (54%) of their total income, on average. In anticipation of a short retirement phase constituting 25% of the round, participants saved substantially more than a quarter (39%) of their total income.

The difference in mean retirement savings for long versus short retirement rounds was calculated for each participant. All but 8 of the 149 participants had greater mean savings for long than for short retirement rounds. One participant was a clear outlier on the difference measure, with a negative difference in savings between long and short retirement rounds of 69 units, which placed her more than 4.5 standard deviations from the mean difference in savings across participants. With the outlier removed (for this and subsequent analyses in this section), the mean difference in savings for long and short retirement rounds was 37 units (SD = 21).

We examined correlations between the difference variable, representing increased saving for long versus short retirement rounds, and our measures of risk attitude and time preference, along with the demographic variables. A greater increase in savings in response to a longer retirement period in the savings task was negatively correlated with the financial risk-taking measure from the DOSPERT ($r = -.18$, $p = .031$). In other words, relatively more financially risk-averse participants tended to increase their savings level more in response to a lengthening retirement period than did less risk-averse participants. (Correlations were in the same direction but not statistically significant for the domain-general DOSPERT.) Higher-income participants also tended to increase their savings more when faced with a longer retirement period than did lower-income participants, $r = .16$, $p = .064$; the correlation with income increases to $r = .18$ ($p = .032$) when the single participant in the highest income bracket is removed from the analysis. Although the correlations are modest in magnitude, it is a promising result that individual variability in savings behavior in response to changes in retirement length in the savings task is predictable from participant characteristics such as risk attitude and income.

Although individual differences in time preference did not correlate with the increase in savings made in anticipation of a longer retirement, there was a modest though not statistically significant correlation between delay discounting (as measured by the Kirby scale) and average levels of savings at retirement ($r = -.14$, $p = .096$), in the expected direction that higher discount rates were associated with lower savings. This finding is consistent with prior research in non-experimental settings (Chabris, Laibson, Morris, Schuldt, & Taubinsky, 2008; Diamond & Köszegi, 2003; Finke & Huston, 2013), again suggesting that the savings task developed here does reproduce meaningful aspects of real-life financial behavior. There was also a marginally significant tendency for women ($M = 120$ units, SD = 22 units) to save more, on average, than men ($M = 113$ units, SD = 25 units), $t(145) = 1.58$, $p = .117$.

5.2. Discretionary spending variability (consumption smoothing)

We measured, for each round completed by a participant, the standard deviation of discretionary spending over the 24 periods in the round. If the participant makes no attempt to smooth consumption, and instead spends whatever is available in the current period, their expected spending standard deviation would be approximately 10 units. In the case of perfect smoothing of consumption, by contrast, the expected spending standard deviation is 0.

We found that 5 of the 149 participants had mean spending standard deviations greater than 10 units. Data from those 5 participants are excluded in the analyses reported below. Among the remaining participants, the mean spending standard deviation was 3.53 units (SD = 1.91 units). This value is well below what would be expected if participants made no attempt at smoothing their consumption.

The spending standard deviation measure was subjected to a mixed-model ANOVA with retirement length and expense variance as within-subject variables and goal as a between-subject factor. Retirement length was found to have a significant main effect on spending variability, $F(1,142) = 9.71$, $\eta_p^2 = 0.064$, $p = .002$, with greater mean spending standard deviations in the long ($M = 3.77$ units) than in the short ($M = 3.30$ units) retirement rounds. This result is likely attributable to the greater variability in income over periods in the long-retirement rounds than in short-retirement rounds. Even so, it is notable that the mean spending standard deviations in the two conditions fell much closer to one another than would be expected if participants had made no attempt at smoothing and instead simply spent whatever was available in a given period. Under such a strategy, spending standard deviations would be expected to be approximately 11.8 and 6.8 in the long- and short-retirement conditions, respectively. Compared to these benchmarks, the typical participant appears to have made an effort to smooth consumption over periods to offset variability in income and expenses.

There was no main effect on spending variability of the goal manipulation, nor did it interact with any other factor. The results discussed above imply that participants generally made some attempt to smooth consumption; the lack of an influence of the goal manipulation suggests that to the extent that they engaged in such smoothing, they tended to do so spontaneously, whether the instructions encouraged it or not. It remains an open question, of course, whether instructions or other manipulations that more effectively promote the goal of consumption smoothing might have a greater impact on measured variability in discretionary spending.

Spending variability was correlated with amount saved for retirement in the savings task: participants with higher mean spending standard deviations tended to have lower mean savings just prior to retirement, $r = -.29$, $p < .001$. Among participants falling above the median spending standard deviation, mean retirement savings was 110 units, versus 124 units for those falling below the median on spending variability. Although we are unable to say anything about the causal connection between these two variables, successful consumption smoothing was associated with greater accumulated savings for retirement in the savings task.

Participants' mean spending standard deviations were positively correlated with the tendency to choose the smaller-sooner option on the Kirby delay-discounting scale, $r = .21$, $p = .012$. In other words, more patient participants (as measured by the Kirby scale) were more effective in smoothing their consumption (discretionary spending) across periods in the savings task. Spending variability in the savings task did not correlate significantly with the impulsivity measure (BIS-11 scores) or with the risk attitude measures. Spending variability was negatively correlated with both age ($r = -.22$, $p = .009$) and education ($r = -.19$, $p = .023$). Older and more educated participants tended to smooth their consumption more effectively in the savings task than did younger and less educated participants. Because age and education themselves were correlated in the sample, $r = .24$, we attempted to examine their unique contributions in a regression of spending standard deviation on the two simultaneously-entered demographic variables. In this analysis, age remained a significant predictor, $t(141) = 2.20$, $p = .03$, while education was a marginally significant predictor, $t(141) = 1.75$, $p = .08$.

5.3. Completion times

Time spent by a participant completing the savings task was estimated from logged web page submission date stamps, as was time spent on the instructions. We report median results as response time measures are typically skewed. Median time to complete all 8 rounds of the savings task was 18 min and 41 s, which produces an average 140 s per round. Median time to read the instructions page was 1 min 49 s, plus 2 min and 35 s to complete the comprehension quiz. Median time to complete the entire task including instructions was 23 min and 23 s. These results indicated that the savings task, as we had hoped, could be completed by a typical AMT worker in well under 30 min.

We also examined correlations between performance on the savings task and time spent completing it. Savings (on average, or the difference between that for short and long retirement rounds) was uncorrelated with time spent completing the task (or reading the instructions). We did find, however, a significant negative correlation between spending standard deviation and time spent completing the savings task, $r = -.21$, $p = .012$. Participants who spent more time completing the task were more effective in smoothing their consumption across periods.

6. Discussion

The main purpose of the present research was to develop a simple experimental savings task that could be readily understood and performed online by members of a diverse participant population. Several aspects of the results indicate that the large majority of the study's participants, selected from the AMT crowdsourcing marketplace, gave meaningful responses to the savings task that suggested they understood the instructions, took the task seriously, and responded sensibly to variations in relevant task factors. Participants accumulated generally adequate levels of savings for impending retirement in the task, for instance, and virtually all of them increased their savings when faced with a long compared to a short retirement phase during which they would earn no income. Most participants also engaged, apparently spontaneously, in an attempt to smooth their consumption (i.e., discretionary spending) over periods in the simulated life cycle.

Performance on the savings task also correlated with relevant individual difference measures of risk attitude and time preference, as well as with some demographic variables. Specifically, compared to those who scored higher, people who scored lower on a measure of financial risk-taking were more sensitive to anticipated retirement length, more sharply increasing their savings when faced with a long rather than a short retirement. Participants with lower delay-discounting scores exhibited greater consumption smoothing compared to those with higher delay-discounting scores. Consumption smoothing also increased with participant age and education level. It is perhaps worth noting that, of the individual difference measures we collected of risk attitude and time preference, only those measures that were specific to the domain of financial decision making (the DOSPERT financial risk-taking subscale, and the Kirby delay-discounting scale) emerged as significant predictors of savings task performance; the domain general measures (e.g., DOSPERT overall score; BIS-11 impulsivity scores) did not. One possible interpretation of this observation is that the savings task is effective in reproducing and capturing cognitions that participants bring specifically to the domain of financial decision making. As a general caveat, it should be acknowledged that correlations between individual difference measures and savings task performance in the present study were quite modest in magnitude, though that does not mean that they are uninformative (for discussion on this point, see Chabris et al., 2008).

We readily acknowledge that the simplifying features of our task and measures necessarily limit the generalizability of our results to more naturalistic (and complicated) settings in which financial decisions are made. For instance, income over the life cycle was entirely predictable in our task, yet many people save primarily as a precaution against income shocks rather than for retirement (Carroll, 1997; Carroll & Samwick, 1997). There are many potential variations on the savings task and its implementation that were beyond the scope of our immediate objectives but could be fruitful avenues for future research. Features such as income shocks, longevity risk, and uncertain returns on savings, all could be readily incorporated in the task framework developed here. In the present study we also focused on only two dependent measures derived from savings task performance; many other potentially informative measures, particularly those looking more closely at period-level responses (e.g., trends in discretionary spending over periods, or lagged responses to earlier-period expense or income shocks), might also be investigated in future research.

Another remaining question is whether savings-task performance would be affected in a systematic way by the introduction of monetary incentives, as opposed to the flat-rate payment provided to participants in the present study. Implementation of incentives would not be completely straightforward as it would require specification of the function mapping consumption over periods onto points that could be redeemed as payment for performance. It could be possible to finesse this problem to some extent by rewarding effective consumption smoothing, though some kind of loss function would still need to be implemented and effectively communicated to participants. Alternatively, we see some value in leaving the utility function unspecified and allowing participants to be guided by their own preferences, which could be validated in future research against independent measures of those same postulated preferences.

Economic games, such as the savings task developed here, are typically limited in their ability to impose the kinds of self-regulatory challenges that often lead to suboptimal financial decision making in everyday life. In our savings task, for example, there was no "temptation" associated with immediate consumption (i.e., present-biased preferences; O'Donoghue and Rabin, 1999), which could help to explain why we did not observe any systematic bias toward undersaving. Brown et al. (2009) similarly found little evidence of undersaving in a standard economic game, but did observe undersaving in a study involving viscerally tempting consumption (of juice rewards by thirsty participants). As noted by Brown et al., the absence of temptation in many economics games could be construed as an advantage, in that it provides a test of whether bounded rationality alone can produce anomalous economic behavior such as undersaving, or whether instead visceral factors triggering present-biased preferences are a necessary contributor.

Acknowledgement

This research was funded by a Grant to the first author from the Natural Sciences and Engineering Research Council of Canada.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.joep.2014.12.004.

References

Ballinger, T. P., Hudson, E., Karkoviata, L., & Wilcox, N. T. (2011). Saving behavior and cognitive abilities. *Experimental Economics, 14*, 349–374.
Ballinger, T. P., Palumbo, M. G., & Wilcox, N. T. (2003). Precautionary saving and social learning across generations: An experiment. *Economic Journal, 113*, 920–947.
Benartzi, S., & Thaler, R. (2007). Heuristics and biases in retirement savings behavior. *Journal of Economic Perspectives, 21*, 81–104.
Blais, A., & Weber, E. U. (2006). A Domain-Specific Risk-Taking (DOSPERT) scale for adult populations. *Judgment and Decision Making, 1*, 33–47.

Brown, A. L., Chua, Z. E., & Camerer, C. F. (2009). Learning and visceral temptation in dynamic savings experiments. *Quarterly Journal of Economics, 124*, 197–231.
Carbone, E., & Hey, J. D. (2004). The effect of unemployment on consumption: An experimental analysis. *Economic Journal, 114*, 660–683.
Carroll, C. D. (1997). Buffer-stock saving and the life cycle/permanent income hypothesis. *Quarterly Journal of Economics, 112*, 1–55.
Carroll, C. D., & Samwick, A. A. (1997). The nature of precautionary wealth. *Journal of Monetary Economics, 40*, 41–71.
Chabris, C. F., Laibson, D. I., Morris, C. I., Schuldt, J. P., & Taubinsky, D. (2008). Individual laboratory-measured discount rates predict field behavior. *Journal of Risk and Uncertainty, 37*, 237–269.
Diamond, P., & Köszegi, B. (2003). Quasi-hyperbolic discounting and retirement. *Journal of Public Economics, 87*, 1839–1872.
Eckel, C. C., & Grossman, P. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior, 23*, 281–295.
Eckel, C. C., & Grossman, P. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior and Organization, 68*, 1–17.
Ersner-Hershfield, H., Garton, M. T., Ballard, K., Samanez-Larkin, G. R., & Knutson, B. (2009). Don't stop thinking about tomorrow: Individual differences in future self-continuity account for saving. *Judgment and Decision Making, 4*, 280.
Fehr, E., & Zych, P. K. (1998). Do addicts behave rationally? *Scandinavian Journal of Economics, 100*, 643–662.
Finke, M. S., & Huston, S. J. (2013). Time preference and the importance of saving for retirement. *Journal of Economic Behavior and Organization, 89*, 23–34.
Hey, J. D., & Dardanoni, V. (1988). Optimal consumption under uncertainty: An experimental investigation. *Economic Journal, 98*, 105–116.
Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics, 14*, 399–425.
Kirby, K. N., & Maraković, N. N. (1996). Delay-discounting probabilistic rewards: Rates decrease as amounts increase. *Psychonomic Bulletin and Review, 3*, 100–104.
Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods, 44*, 1–23.
O'Donoghue, T., & Rabin, M. (1999). Doing it now or later. *American Economic Review, 89*, 103–124.
Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgement and Decision Making, 5*, 411–419.
Patton, J. M., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt Impulsiveness Scale. *Journal of Clinical Psychology, 51*, 768–774.
Suri, S., & Watts, D. J. (2011). Cooperation and contagion in web-based networked public goods experiments. *PLoS ONE, 6*, e16836.
Weber, E. U., Blais, A., & Betz, N. E. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making, 15*, 263–290.
Winter, J. K., Schlafmann, K., & Rodepeter, R. (2012). Rules of thumb in life-cycle saving decisions. *The Economic Journal, 122*, 479–501.